

Risk fixers and sweet spotters: A study of the different approaches to using visual sensitivity analysis in an investment scenario

T. Torsney-Weir¹, S. Afroozeh¹, M. Sedlmair², and T. Möller^{1,3}

¹ Faculty of Computer Science, University of Vienna, Vienna, Austria

² Computer Science, Jacobs University, Bremen, Germany

³ Data Science @ University of Vienna, Vienna, Austria

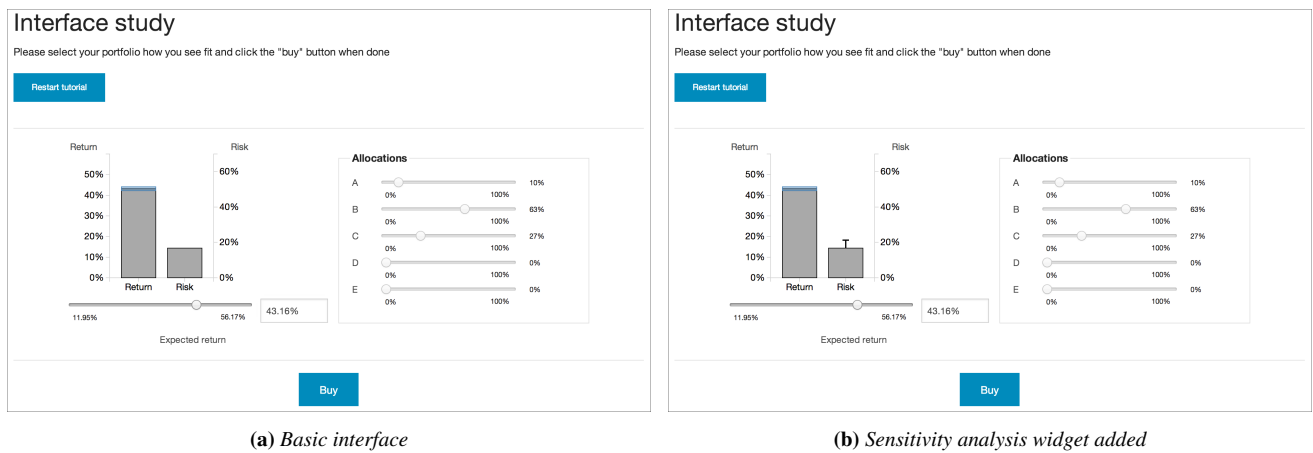


Figure 1: The two interfaces presented to participants in the user study. (a) is interface without sensitivity which is divided into two sections. On the left of the screen are the bar charts showing the expected return (left) and expected risk (right). The expected return bar is interactive. The user can adjust the expected return by dragging the top of the corresponding bar in the plot. The right section of the interface shows the optimal investment allocations computed by the Markowitz [Mar52] model. When the expected return changes, the system automatically recomputes the optimal investment choices and displays the results. (b) is the interface with the additional sensitivity feature. We encode the gradient of the expected risk with respect to the expected return as the length of a small “whisker” glyph on top of the expected risk bar. This represents the increase in expected risk given a small increase in expected return.

Abstract

We present an empirical study that illustrates how individual users’ decision making preferences and biases influence visualization design choices. Twenty-three participants, in a lab study, were shown two interactive financial portfolio optimization interfaces which allowed them to adjust the return for the portfolio and view how the risk changes. One interface showed the sensitivity of the risk to changes in the return and one did not have this feature. Our study highlights two classes of users. One which preferred the interface with the sensitivity feature and one group that does not prefer the sensitivity feature. We named these two groups the “risk fixers” and the “sweet spotters” due to the analysis method they used. The “risk fixers” selected a level of risk which they were comfortable with while the “sweet spotters” tried to find a point right before the risk increased greatly. Our study shows that exposing the sensitivity of investment parameters will impact the investment decision process and increase confidence for these “sweet spotters.” We also discuss the implications for design.

CCS Concepts

•Human-centered computing → Empirical studies in visualization; Visual analytics;

1. Motivation

Visualization designers are faced with a plethora of choices when designing an analysis tool. Guidelines about which visual encodings and interactions to use and when to use them allow for faster and more effective design decisions. Perceptual studies [HB10, War04, CM84, BHR17] and task taxonomies [AES05, BM13, Shn96] help us understand, at a low-level, what visual encodings are most effective and what tasks users will perform. However, Teovanović et al. [TKS15] find that cognitive biases and eventual decision making performance varies on an individual level. This implies that two different users will use different low-level tasks and visual encodings to perform the same high-level task. If we could discern different classes of user behavior for the same problem then we could develop guidelines for which visual encodings and interactions better address the decision making preferences of these different user groups. In this study we develop a protocol and present evidence of these different user classes. In addition we show that adding additional visualization components may only help certain user groups.

We perform the investigation in the context of a user study that measures the effect of the addition of sensitivity analysis to a visualization tool for portfolio optimization. The specific question we seek to answer is: “Does a proper visual encoding of a sensitivity analysis measure convey sensitivity and lead people to make more informed decisions?” Furthermore, what are the cognitive consequences of adding these measures? Sensitivity analysis [STCR04] is one method designed to facilitate this analysis. On the one hand, these measures elicit additional information about the input/output relationship of simulations. On the other hand, they add some complexity to the analysis. Instead of computing numerical measures, it may be simpler to give the users interactive control over their simulation and let them infer the input/output relationships that way.

Our hypothesis was that encoding the gradient of expected risk versus expected return would be effective for all participants. After our study we found that this was not effective for all participants. However, we do identify a subset of participants (12/23), which we named the “sweet spotters,” for which we saw a clear effect of increased confidence in their investment decisions. These users focused on the non-linear relationship of the risk/return trade-off and sought a point right before the risk increased greatly and tried to invest there. The “risk fixers,” which did not focus on the non-linear relationship of the risk/return curve, did not find the sensitivity analysis feature as helpful.

2. Study design

We chose an investment task because it is a domain where non-specialists regularly interact with complex models. In addition, they need to understand these models to work productively with them. Selecting an investment portfolio is a very open-ended decision. There is no notion of a “correct” answer and a faster answer may not be better [APM*11]. There are many optimal solutions to the problem which lie along the “efficient frontier” or “Pareto front.” Each participant could have their own notion about what an ideal risk/return trade-off is. For this study we wanted to maintain ecological validity as much as possible by not forcing a participant to

make a “correct” investment. Therefore, we measured participants’ confidence in their decision as a metric. This has been used as a quality metric in other open-ended tasks such as visual encodings of uncertainty [CG14] and decision support systems [MG00, Yi08]. The questions used in our study are adapted from those used in Yi’s PhD thesis [Yi08] and are included in the supplementary material.

We employed a within-subject A-B-A study design presenting the two interfaces shown in Figure 1. The only difference between the two interfaces is the addition of a sensitivity indicator encoded as a small whisker on top of the risk bar for the B-test (shown in Figure 1b) in order to mitigate external factors. Whiskers have been shown to be undesirable for showing error distributions [CG14]. However, we are using the whisker to indicate the change in bar height as the return value increases (i.e. the gradient) rather than a distribution. The expectation is that if the B-test has any real effect then we would see a change in the confidence results for the sensitivity interface and then a return to the baseline confidence when given the A-test interface again.

We conducted a lab study administered in the offices of our research group consisting of 23 participants recruited from major universities. There were 8 females and 15 males with age ranges from 20 to 40 years old with an average age of 27.7. Few participants had prior investment experience. Participants were not compensated for their participation in the study nor given course credit. Before starting the study the participants were given an overview of the interface. We identified the sensitivity whisker to each participant before they used the B-test interface. After each participant finished they selected via a form which of the two interfaces they preferred overall as well as a semi-structured interview about the study. Prior work on behavioral decision making shows that experts tend to use their expertise to identify a solution rather than comparing different solutions against each other [LKOS01]. Thus, we decided to use investing amateurs.

Before analyzing the results, we split the participants into two groups based on their investment strategy as reported during their interview. Personality factors have been shown to have an influence on visual layout [CCH*14] and user strategy [OYC15].

After analyzing the results, as a preliminary determination of what is driving this decision making process, we coded the interviews using grounded theory [Cha06] methods. Two coders analyzed the qualitative data of 22 interviews (one participant asked not to be recorded) and assigned relevant codes to various phrases spoken by the participants.

3. Results

We divided the participants based on their response to a question about how they chose their final portfolio. We found that our initial hypothesis that encoding sensitivity analysis in the form of local change in expected risk given a small change in expected return is not universally effective for all participants (Figure 2a). When examined separately, however, the “sweet spotters” (12/23 participants), did indicate an increase in confidence and preference for the interface with the sensitivity feature. The “risk fixers” did not indicate such a response (Figure 2b). The “sweet spotters,” focused on the non-linear relationship of the risk/return trade-off and sought

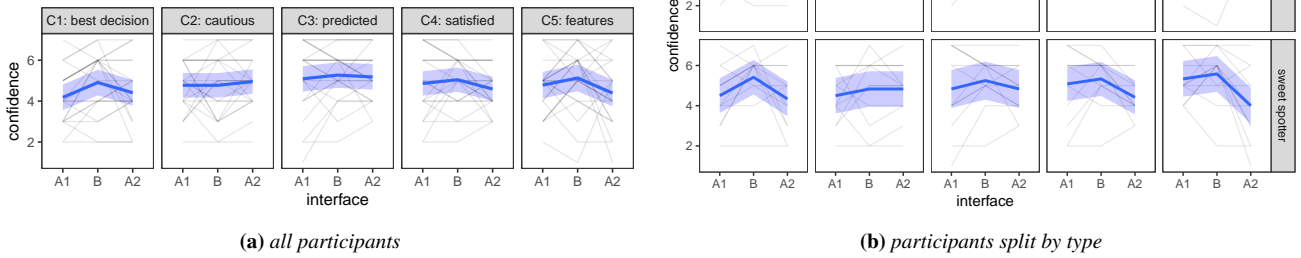


Figure 2: (a) The blue line shows the fitted response to the participants confidence questions. The confidence interval is shown as a light blue band. While there is some response to interface B, it is not a very strong one. (b) Fitted responses to the different confidence questions as the participants progressed through the study. We split the participants into groups of “risk fixers” and “sweet spotters.” Note the lack of change in confidence for the risk fixers but the increased confidence from the “sweet spotters”.

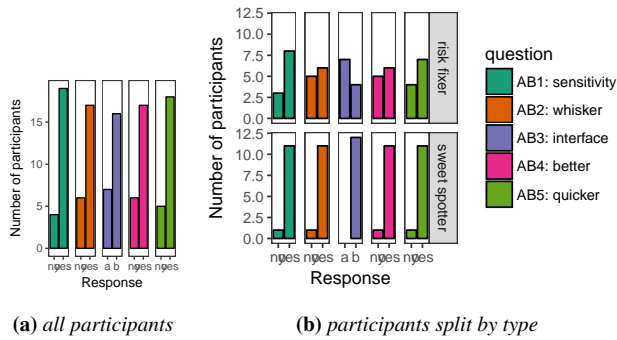


Figure 3: Results of the interface preference question. The columns marked “no” refer to the interface without the sensitivity analysis widget and the columns marked “yes” refer to the interface with the sensitivity analysis widget. When viewed in aggregate (a) the users seemed to prefer the interface with the sensitivity analysis widget but upon further inspection (b) this is due to the overwhelming preference for this interface by the “sweet spotters.”

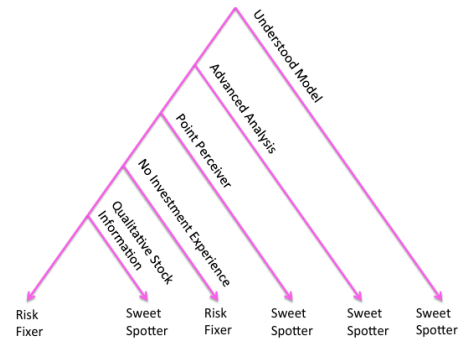


Figure 4: The decision tree we built using the coded interview data. We can see that the “sweet spotters” seemed to focus on the quantitative/modelling side of the system while the risk fixers did not.

a point right before the risk increased greatly and tried to invest there. The “risk fixers” did not focus on the non-linear relationship and instead had a pre-selected notion of what an acceptable risk level is and adjusted the interface to that goal. For our analysis we rely on plots of data and confidence bands rather than statistical significance testing. For an excellent discussion of the advantages of confidence intervals see *Understanding the new statistics* [Cum12].

3.1. Interface preferences

In Figure 2 we show the results of the confidence questionnaire (question detail is in supplementary material). The increased confidence and fall to the baseline, especially with respect to questions C1, C4, and C5, is consistent with the expected behavior if the additional sensitivity element is effective. In fact, the sweet spotters’ confidence dropped even further than the baseline between inter-

face A1 and A2 after they were shown the sensitivity widget. We believe this preference is due to the difference in how the users accomplished the portfolio selection task. The “sweet spotters” noticed that the sensitivity widget would jump right before reaching the point where the risk/return changed greatly. The “risk fixers” only cared about finding a particular risk value.

We also split up the results of our A/B interface questions by whether the participants are “risk fixers” or “sweet spotters.” The histograms of these results are shown in Figure 3. Here we can see the clear difference between the preferences of the two groups. The “risk fixers” have no clear preference for the with- versus without-sensitivity interfaces. However, the overwhelming majority of “sweet spotters” preferred the interface with the sensitivity widget. We believe, and indeed this some participants mentioned this in the interviews, that the widget helped them identify the “sweet spot” in the risk/return curve that they were searching for.

3.2. Coding

To further elucidate the reasons these participants chose a particular strategy we coded the semi-structured interviews using grounding theory methods [Cha06]. Two coders independently listened to recordings of the interviews and produced code lists. The coders then met and discussed any differences in order to resolve the codes. The inter-coder reliability, as measured by Krippendorff's alpha [Kri13], is 0.67.

We used these codes and user class to build a decision tree designed to classify the users based on the interview codings, shown in Figure 4. We used this tree to understand which codes are most predictive of the user type. We found that the most common characteristics for "risk fixers" was a lack of confidence in their choices and a focus on irrelevant visual information (the widget). "Sweet spotters" were very number-focused and explored the model.

4. Discussion

Our original hypothesis that adding a local sensitivity feature in the form of the gradient of expected risk to expected return would increase the confidence about investment decisions turned out to be wrong as stated. We did, however, find that feature helpful for the "sweet spotters" group. This makes sense as these participants were trying to find the point right before the risk changed drastically. We had not anticipated that there would be these two types of users, one that tried to optimize the gradient of the risk/return curve and one that just picked a (maximum) risk value. This separation only came about as a result of careful consideration of participants responses to our post-test interview.

We can use our decision tree to identify the key factors for categorizing users. The "risk fixers" seem to exhibit many of the investor biases that are laid out by Sahi et al. [SAD13]. For example, the "risk fixers" tended to be *adverse to losses* and *played safe* with the level of risk. This may be the reason they did not identify the heuristic employed by the "sweet spotters." However, in the Sahi et al. study few participants employed financial models to make investment decisions. In addition to requiring the participants to use financial models, our study measured participant's actual performance versus their self-reported investment methodology.

4.1. Proper user characterization

In general our results echo the importance of understanding user characteristics in the visualization design process and early iterative design [LD11, SMM12]. Namely, that one needs to firmly understand users' cognitive biases and decision making preferences before designing an interface. Speaking with potential users and understanding their needs is vital to producing a well-designed tool. Iteration on design is also vital. Our initial assumption, augmented with feedback from a financial industry expert, was that users would pick a risk value which they would be comfortable with. This was also our mode of thinking when we wrote up the investment scenario. It was only once participants started using the interface that they realized the benefit of the sensitivity feature. This was true for both experienced and novice participants in investing. This is a strong case for iterative design. This also has implications for activity-centered design [Nor05].

4.2. Adaptive user interfaces

Our results also have applications for adaptive user interfaces. Toker et al. [TCCH12], evaluate the effectiveness of taking into account user characteristics in visualization displays. Hudlicka and Billingsley propose a framework to adapt an interface to address user biases [HB99]. We have identified that there are two ways that participants went about performing the portfolio optimization task. These two types had different interface requirements, one found the sensitivity analysis feature we added useful while the other did not. With a proper automatic identification of the type of user one could design a user interface that would automatically add additional analysis features in order to support those users.

4.3. Unnecessary features

We did not find a great difference between the with- and without-sensitivity interfaces for the "risk fixer" group. This finding indicates that you may be able to add small features like our use of the sensitivity widget and not negatively impact people. But you can help others do their analysis better. In our study we only added a small glyph which, according to the 50/50 split in Figure 3b, did not negatively affect the "risk fixers." Other visual elements may be distracting, though, especially in light of what we discussed in Section 4.1. The sensitivity widget was a relatively small addition and this makes it easy to ignore. Users may find larger changes more distracting. This is an exciting opportunity for future work.

4.4. Critical reflection and future work

This study was performed on a limited subject pool of 23 participants. We intend to extend this study on a larger scale using something like mechanical turk along with additional personality tests like numeracy [ZFU*07] and maximizing versus satisficing [SWM*02]. Currently, we identify user behavior based on manually coding the interviews with participants. It would be impractical to do this for a few hundred participants and mechanical turk users may not write sufficient detail about how they went about their analysis to produce a reliable classification. We will develop a more reliable method for detecting the user behavior perhaps based on mouse data, similar to what was done in Brown et al. [BOZ*14].

One unanticipated issue is that a number of participants mentioned in the interview that they had never invested before and that affected their confidence. Users have trouble understanding a model that they have never used before, even this simple-seeming portfolio model. It would be interesting to also run a test with expert investors and see if we get similar results. We could also employ a much more complex investment model in order to discern if we still see these two types of users exist.

5. Conclusion

Given a single task, namely select a portfolio to invest in, participants had two very different ways of accomplishing this task. We then evaluated the participant interviews to identify user factors that contribute to these behaviors. By identifying the different types of analyses that different users want to perform, we can build visualization systems that address shortcomings in their methods of analysis while supporting their strengths.

References

- [AES05] AMAR R., EAGAN J., STASKO J. T.: Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization (InfoVis)* (Oct. 2005), IEEE Computer Society, pp. 111–117. doi:10.1109/INFVIS.2005.1532136. 2
- [APM*11] ANDERSON E. W., POTTER K. C., MATZEN L. E., SHEPHERD J. F., PRESTON G. A., SILVA C. T.: A user study of visualization effectiveness using eeg and cognitive load. *Computer Graphics Forum* 30, 3 (2011), 791–800. URL: <http://dx.doi.org/10.1111/j.1467-8659.2011.01928.x>, doi:10.1111/j.1467-8659.2011.01928.x. 2
- [BHR17] BAE J., HELLDIN T., RIVEIRO M.: Understanding indirect causal relationships in node-link graphs. *Computer Graphics Forum* 36, 3 (July 2017), 411–421. doi:10.1111/cgf.13198. 2
- [BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (Dec. 2013), 2376–2385. doi:10.1109/tvcg.2013.124. 2
- [BOZ*14] BROWN E. T., OTTLEY A., ZHAO H., LIN Q., SOUVENIR R., ENDERT A., CHANG R.: Finding Waldo: Learning about users from their interactions. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec. 2014), 1663–1672. doi:10.1109/tvcg.2014.2346575. 4
- [CCH*14] CONATI C., CARENINI G., HOQUE E., STEICHEN B., TOKER D.: Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. *Computer Graphics Forum* 33, 3 (June 2014), 371–380. doi:10.1111/cgf.12393. 2
- [CG14] CORRELL M., GLEICHER M.: Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec. 2014), 2142–2151. doi:10.1109/tvcg.2014.2346298. 2
- [Cha06] CHARMAZ K.: *Constructing grounded theory: A practical guide through qualitative analysis*. Introducing Qualitative Methods. Sage Publications Ltd, 2006. 2, 4
- [CM84] 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 (1984), 531–554. doi:10.2307/2288400. 2
- [Cum12] CUMMING G.: *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge, 2012. 3
- [FZFU*07] FAGERLIN A., ZIKMUND-FISHER B. J., UBEL P. A., JANKOVIC A., DERRY H. A., SMITH D. M.: Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making* 27, 5 (Sept./Oct. 2007), 672–680. doi:10.1177/0272989X07304449. 4
- [HB99] HUDLICKA E., BILLINGSLEY J.: Affect-adaptive user interface. In *Human-Computer Interaction: Ergonomics and User Interfaces, Proceedings of HCI International '99 (the 8th International Conference on Human-Computer Interaction), Munich, Germany, August 22-26, 1999, Volume 1* (Jan. 1999), pp. 681–685. 4
- [HB10] HEER J., BOSTOCK M.: Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010), ACM, pp. 203–212. doi:10.1145/1753326.1753357. 2
- [Kri13] KRIPPENDORFF K.: *Content Analysis: An Introduction to Its Methodology*. SAGE Publishing, 2013. 4
- [LD11] LLOYD D., DYKES J.: Human-centered approaches in geovisualization design: Investigating multiple methods through a long-term case study. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 2498–2507. doi:10.1109/TVCG.2011.209. 4
- [LKOS01] LIPSHITZ R., KLEIN G., ORASANU J., SALAS E.: Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making* 14, 5 (2001), 331–352. URL: <http://dx.doi.org/10.1002/bdm.381>, doi:10.1002/bdm.381. 2
- [Mar52] MARKOWITZ H. M.: Portfolio selection. *The Journal of Finance* 7, 1 (Mar. 1952), 77–91. doi:10.2307/2975974. 1
- [MG00] MADSEN M., GREGOR S.: Measuring human-computer trust. In *Proceedings of the 11th Australasian Conference on Information Systems* (2000), pp. 6–8. 2
- [Nor05] NORMAN D. A.: Human-centered design considered harmful. *interactions - Ambient intelligence: exploring our living environment* 12, 4 (July 2005), 14–19. URL: <http://doi.acm.org/10.1145/1070960.1070976>, doi:10.1145/1070960.1070976. 4
- [OYC15] OTTLEY A., YANG H., CHANG R.: Personality as a predictor of user strategy. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (2015), ACM Press, pp. 3251–3254. doi:10.1145/2702123.2702590. 2
- [SAD13] SAHI S. K., ARORA A. P., DHAMEJA N.: An exploratory inquiry into the psychological biases in financial investment behavior. *Journal of Behavioral Finance* 14, 2 (Apr./June 2013), 94–103. doi:10.1080/15427560.2013.790387. 4
- [Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages* (1996), pp. 336–343. doi:10.1109/VL.1996.545307. 2
- [SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (Nov./Dec. 2012), 2431–2440. doi:10.1109/TVCG.2012.213. 4
- [STCR04] SALTELLI A., TARANTOLA S., CAMPOLONGO F., RATTO M.: *Sensitivity analysis in practice: A guide to assessing scientific models*. Wiley Publishing, Apr. 2004. 2
- [SWM*02] SCHWARTZ B., WARD A., MONTEROSSO J., LYUBOMIRSKY S., WHITE K., LEHMAN D. R.: Maximizing versus satisfying: Happiness is a matter of choice. *Journal of Personality and Social Psychology* 83, 5 (Nov. 2002), 1178–1197. doi:10.1037/0022-3514.83.5.1178. 4
- [TCCH12] TOKER D., CONATI C., CARENINI G., HARATY M.: Towards adaptive information visualization: On the influence of user characteristics. In *Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization* (2012), UMAP'12, Springer-Verlag, pp. 274–285. doi:10.1007/978-3-642-31454-4_23. 4
- [TKS15] TEOVANOVIĆ P., KNEŽEVIĆ G., STANKOV L.: Individual differences in cognitive biases: Evidence against one-factor theory of rationality. *Intelligence* 50 (May 2015), 75–86. doi:10.1016/j.intell.2015.02.008. 2
- [War04] WARE C.: *Information visualization: Perception for design*, second ed. Morgan Kaufmann, San Francisco, CA, USA, Apr. 2004. 2
- [Yi08] YI J. S.: *Visualized decision making: development and application of information visualization techniques to improve decision quality of nursing home choice*. PhD thesis, School of Industrial and Systems Engineering, Georgia Institute of Technology, 2008. 2