EuroRV3: EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (2015)
W. Aigner, P. Rosenthal, and C. Scheidegger (Editors)

On the Reproducibility of VisRuption: A Tool for Intuitive and Efficient Visualization of Airline Disruption Data

Nicholas Hugo Müller Linda Pfeiffer Peter Ohler Paul Rosenthal

Technische Universität Chemnitz, Germany

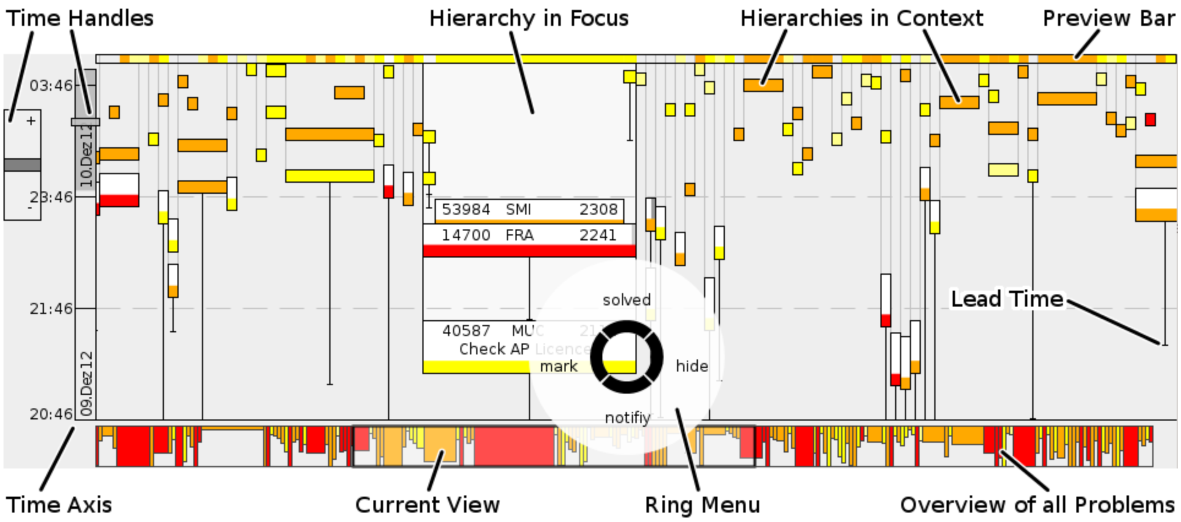

Figure 1: *Annotated screen shot of VisRuption showing a serious problem situation with 1,000 problems. The visible time span is set to seven hours and hierarchies are sorted with respect to airport.*

1. Introduction

Managing the vast amount of resources and processes of large airlines with several hundred aircraft and several thousand operated flights per day is a very complex task and makes computer-aided operation irreplaceable. Moreover, there is a multitude of disruptions which can occur every day during airline operation and can result in very expensive delays or cancellations [CTA04, MHR10, Now09]. In our paper at EuroVis 2013 [RPMO13], we have presented a design study about the tool VisRuption for providing an intuitive and efficient access to airline disruption data.

The tool and the whole design study originated from a research project in close collaboration and partially funded by Lufthansa Systems AG. This industry involvement was in the first place enabling to tackle this research question and to gain valuable input. However, especially the research and development contract resulted in some issues that we want to discuss in the following sections.

2. Design Process

The strong relation to Lufthansa Systems AG enabled us to implement a strict user-centered and iterative design process [SMM12]. During the initial phase, we could interview several domain experts from Lufthansa Systems and, in addition, from several international airlines at an invitation-only customer conference. The final set of user requirements was obtained during workshops at several real control centers, which would have been much more cumbersome if not impossible without support of the industrial partner.

As valuable as this connection was, it came with a large set of restrictions which are quite common in high-value business and security-related environments. It was not possible to record any interview with audio or video or even take pictures of the workplaces. During the analysis and design phase we had to solely rely on handwritten notes and short-term feedback from our person of contact at Lufthansa Systems.

Our state-of-the-art review of visualization methods for airline disruption data was again highly supported by the industry partners, since they typically have a detailed knowledge about their competitors [Jep]. They could help to find information about the competing designs and even allowed us to compare to a quite recent internal Master's thesis [Hös09] by providing detailed information about the design process there. However, publishing of the provided knowledge was, again, strongly regulated.

During the design and prototyping cycles we were able to utilize once more the connection of the industry partners to their customers. We constantly collected feedback at customer conferences, which saved a lot of time and expenses for traveling to the international airlines. In addition, we had a constant development relation to two airlines for continuous evaluation of designs, mock-ups, and prototypes. However, these detailed insights came again with the restriction of nondisclosure. Thereby rendering the reproducibility of the whole design process nearly impossible.

In the original paper, we have presented a detailed problem description which could be used to extract similar use cases or initiate related projects. Still, a validation is only possible by going a similar way like we did, with the help of an airline software company. The same applies for the design requirements, which can only be validated by generating the same insight into the operations control process. However, when accepting the problem definition and design requirements as valid, it is possible to build new concepts and competing visual analytics tools on basis of our paper.

Finally, we were obviously not allowed to release a working prototype and could only publish screenshots and videos of the program. Still, we tried our best to explain every design choice and its motivation together with the functionalities in detail. This should allow every graduate student to reimplement the system and reproduce the principle functionality.

3. User Study

During the development of the project it got immediately clear that an adequate and valid assessment of the quality and performance of the system would be only possible under realistic conditions. As a first consequence, we needed to conduct the study with real airline operators. Since we wanted to compare to the current version of Lufthansa Systems' tool NetLine, we ideally needed NetLine users. They are familiar with the whole problem complex, are well trained in analyzing airline disruption data, and know the features and procedures of NetLine very well. In contrast, computer science students or a random selection of participants could either only evaluate the general usability of the tool, which was far to few for our collaborators, or would need an elaborate introduction into the complicated disruption management process. Still, they would perform very differently from real power-users and would most-likely be unable to cope with all the features and complex procedures of the NetLine system. Getting in contact with the specific user group would, in general again, be very cumbersome. With the help of our collaborators this was much easier but again with the cost of some nondisclosure.

When conducting such a user study with real users and their real-world tools, it would be very hard to construct realistic synthetic data. Giving the users only toy examples would not realistically reproduce the working reality and constructing a whole flight plan of several days for a middle-sized airline would have implied a huge and disproportional effort. This is why we decided to implement an interface to the actual internal NetLine data bases, which obviously could also not be documented. This gave us the opportunity to use recorded real world data from a real airline for the user study. We explicitly excluded operators from this airline from the study to eliminate learning effects. Using real data unfortunately limits again the reproducibility since we are not able to provide a data set for testing of competing implementations.

In terms of the actual user study, we tried to publish as much of the design and results as possible due to the nondisclosure agreement with our collaborators. We carefully described the choice and structure of the user group and the procedure of the user study. Since we did not rely on specific hardware, it would be possible to reproduce the study if similar users are selected and the tools are available. Regarding the stimuli, we published the anonymized questions the users had to answer, though omitting the used data.

Regarding the results of the user study we were not allowed to publish detailed raw data. Still we described in detail which tests we used and how the used data was aggregated. Where possible, we stated detailed mean and other statistical key values. Concerning the demographic questionnaire, we were only able to present the key findings and correlations. The raw data needed to be kept close due to protection of employees and company knowledge. The results of the questionnaire to subjectively compare the two tools could be published in more detail and we would be able and are willing to also release to anonymized raw data on demand. However, since all other data is not available and the publication would give only minor additional insight, we decided to not put this data on a webpage.

4. Conclusions

We have presented all reproducibility aspects of our design study that led to the prototype VisRuption. Due to restrictions by industry collaborators it was very hard to adequately release reproducibility information. We tried to incorporate as much information as we were allowed to into the original paper, such that at least a reimplementation of a working prototype without interface to real data should be easily possible. Possibly, some specific reward of the community for good reproducibility would have made our efforts and negotiations with collaborators into this direction more fruitful.

References

[CTA04] COOK A. J., TANNER G., ANDERSON S.: *Evaluating the true cost to airlines of one minute of airborne or ground delay: final report*. Tech. rep., Eurocontrol, 2004. 1

[Hös09] HÖSEL J.: *3D Information Visualization for Disruptions Monitoring in Airline Crew Tracking*. Master's thesis, Hochschule für Technik und Wirtschaft Berlin, 2009. 2

[Jep] JEPPESEN SYSTEMS AB: Disruption Management. www.jeppesen.com/industry-solutions/aviation/commercial/-carmen-integrated-operations-control.jsp. 30.03.2015. 2

[MHR10] MAZZOCCHI M., HANSSTEIN F., RAGONA M.: The 2010 volcanic ash cloud and its financial impact on the european airline industry. *CESifo Forum 11*, 2 (2010), 92–100. 1

[Now09] NOWAK I.: A business perspective to robust airline planning. In *Proceedings of EURO* (2009). 1

[RPMO13] ROSENTHAL P., PFEIFFER L., MÜLLER N. H., OHLER P.: Visruption: Intuitive and efficient visualization of temporal airline disruption data. *Computer Graphics Forum 32*, 3 (2013), 81–90. 1

[SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans. Vis. Comput. Graphics 18*, 12 (2012), 2431–2440. 1