EuroVis Workshop on Visual Analytics (2015)
E. Bertini and J. C. Roberts (Editors)

Visual Scaffolding in Integrated Spatial and Nonspatial Analysis

G.Elisabeta Marai [1]

[1]University of Illinois at Chicago, USA

Abstract

Collaborative visual analytics that feature mixtures of spatial and nonspatial data occur across disciplines, and are particularly common in bioinformatics, neuroscience and geospatial analysis. In this work we analyze, from a human-centric perspective, data collected from the design and evaluation of three successful visual analysis tools, spanning seven case studies. We focus on the importance of the users' background to the design process, and we discuss the importance of visual scaffolding to such collaborative, integrated spatial and nonspatial visual analysis tools. Scaffolding is a psychology concept which denotes the support given during a learning process. We further present evidence that spatial and nonspatial coordinated views can serve as a form of visual scaffolding for expert-level, collaborative visual analyses.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Visual analysis processes and workflows—Cognitive and perceptual aspects of visual analytics—Collaborative visual analytics

1. Introduction and Background

An increasing number of big data problems are collaborative endeavors that feature mixtures of spatial and nonspatial data. For example, a typical epidemiology study looking at the effects of aging on brain function across large groups of population may require collaboration among: 1) public-health experts who formulate hypotheses, analyze demographics, and recruit the subjects; 2) neurosurgeons and biologists to analyze the subjects' brain scans and physiological measurements; 3) statisticians to design algorithms for computing correlations among the various demographics and measurements; 4) high-performance computing specialists to run the algorithms at supercomputing centers; and 5) the visual analysis experts who help analyze and present the results. The data in the same study may span both the spatial domain — in the form of spatial fields such as brain scans — and the nonspatial domain — in the form of demographic measurements or scores registered when assessing the functional proficiency of a subject's brain. The domain experts often have complementary expertise, and collaborative spaces are essential for the visual analysis process.

Collaborative visual analytics that feature mixtures of spatial and nonspatial data occur in fact across disciplines, and are particularly common in bioinformatics, neuroscience and geospatial analysis [DH01, KKH02, Rhy03, RWK*99, MPLB07]. In these applications, two paradigms are prevalent for integrating spatial and nonspatial data: overlays and coordinated multiple views (also known as multiple linked views). Overlays are predominantly used in geospatial analysis and augmented reality, while coordinated views are prevalent in almost every type of visual analytics (including geospatial). From the existing literature, however, it is not clear under what circumstances one paradigm would be preferable to the other—both paradigms can handle complex data. Furthermore, coordinated views suffer from context-switching and real estate problems, while overlays typically do not [WBWK00].

One possible factor in the human-centered design [KSDK11, MMAM14, GDJ*13, SP06, WKvD*08, TM04, SMM12, Mun09] of these integrated visual analysis tools is the user background and their familiarity with a particular type of representation. In this work we take a deeper look at the human component of these design processes. To this end, we analyze post-hoc the data collected from the design and evaluation of three successful, published visual analysis tools, spanning seven case studies.

In our analysis, we consider the resulting designs from the perspective of instructional scaffolding [NB78], a psy-

DOI: 10.2312/eurova.20151097

chology concept which denotes the support given during the learning process in order to promote deeper learning. In this context, visual scaffolding is support that includes images and words that can be seen. Visual scaffolding is commonly used in education, for example in the instruction of English as a second language.

2. Methods

We analyze the design process, as well as the adoption and usage of three successful visual analysis systems. Each system targets a separate domain: bioinformatics, epidemiology, and astronomy. Each system integrates spatial and nonspatial data, and all systems support inter-expert collaboration. Each system has benefited from the regular participation in the design process of between five and twelve domain experts, and extended feedback from several off-site research groups. In general, the experts had complementary expertise.

The three systems were developed over a period of nine months, respectively two years and three years. Each system has been evaluated by the respective domain experts, and each system has been adopted by the domain experts and their respective labs as a research tool.

2.1. Design and Evaluation Process

We briefly describe the nature of each of the three systems used in this analysis, as well as the process used to design these systems.

Bioinformatics: The first system is a visual mining and analysis tool to help identify protein mutations across family structural models, and to help discover the effect of these mutations on protein function (Figure 1) [LWC*14]. Distributed data sources for 3D structure and nonspatial sequence information are seamlessly integrated into a common visual interface. Multiple linked views and a computational backbone allow comparison at the molecular and atomic levels, while a trend-image visual abstraction allows for the sorting and mining of large collections of sequences and of their residues.

Epidemiology: The second system is a visual framework for the integration, comparison, and exploration of correlations in spatial and nonspatial geriatric data [MMH*13]. These data are in general high-dimensional and span both the spatial, volumetric domain — through magnetic resonance imaging volumes — and the nonspatial domain, through variables such as age, gender, or walking speed. A linked-view design geared specifically at interactive visual comparison integrates spatial and abstract visual representations to enable the users to effectively generate and refine hypotheses in a large, multidimensional, and fragmented space. The software was described as their "dream tool" by users.

Astronomy: The third system is a "Scientific Google Sky" which assists the visual integration, mining and interactive navigation of large-scale astronomy observations [LCO*14]. The system seamlessly integrates spatial data — in the form of multi-modality sky images collected through telescopes —, with nonspatial data — in the form of spectral distributions. Several astronomy research groups have adopted the system, and use it to more easily and quickly identify patterns and outliers in their data. The researchers were eager to use the tool in classrooms.

Requirements: To characterize the domain data and tasks, we utilized ethnographic observation, interviews, focus groups, and critiques of existing visualization tools [Mun09]. Usability and user experience requirements were also determined: learnability, ease of installation, and ease of use. From the scaffolding perspective, we also considered the users' backgrounds and expertise.

Visual encoding: In the encoding stage we utilized participatory parallel design, guided by regular feedback, in which the data types and tasks were mapped to multiple visual and interaction encodings—as long as the encodings were appropriate for the data and tasks earlier determined. The design stage employed whiteboard sketches, as well as paper and lightweight to fully-developed interactive prototypes.

Evaluation: For each system, the domain experts have been given access to both lightweight and fully developed prototypes at regular intervals during the design and development of the system. We observed the use of the system in the context of group meetings, where we noted collaborative hypothesis generation and discussion of observations. From the scaffolding perspective, the users were provided with compelling tasks (identified during the domain characterization), as well as with templates and guides in the form of demonstrations and tutorials.

3. Results

In each case, the result of the participatory design process was a coordinated multiple views solution. Analyzing the encoding process and its series of prototypes, we note that the spatial-data experts posed initially significant resistance to nonspatial abstractions. The nonspatial experts posed resistance to the spatial representations, and in particular resisted design attempts at overlaying. The more unfamiliar the visual representation, the more resistance was posed, and the more support was necessary — sometimes in the form of toy case studies to demonstrate the benefits of that particular representation for the expert's precise domain. Overlays were never embraced. In the longer run, some of the linked visual abstractions were found to be intuitive and rather essential in better understanding the data. However, visual support from familiar visual representations (such as volume renderings and hierarchies) was essential in this training period.

All experts involved in the evaluation of these system have

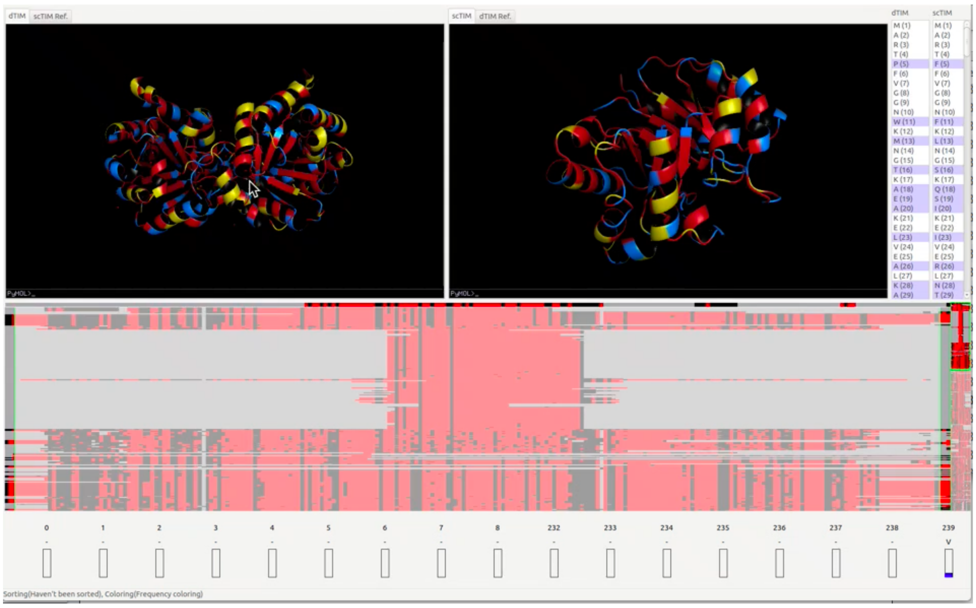

Figure 1: *A visual mining and analysis tool to help identify protein mutations across family structural models. The system integrates spatial information regarding molecular and atomic structures with a trend-image visual abstraction that allows for the sorting and mining of large collections of sequences and of their residues.*

stated that having different ways (abstract and spatial) of seeing the data was particularly helpful. They also commented that the visualization tools could be very useful for non-expert users, who are not very familiar with the domain, as using the tools would improve such users' understanding.

In terms of use, the domain experts have adopted each final prototype, and the systems are now used regularly by their teams. For the purpose of this analysis, for each system we asked the experts to complete a total of seven case studies, using each time the "think aloud" approach. The seven case studies were distributed as follows: two epidemiology cases involving 3 experts, three astronomy cases involving 5 experts, and two bioinformatics cases involving 4 experts.

For the analysis of the system usage, we divided virtually each interface into spatial and nonspatial sections. We then analyzed the case study output, with respect to the experts' background, in terms of the virtual sections the experts referred to in their analysis, and of the timing of these references. Approximately half of the experts had expertise in the spatial domain (molecular structure, brain gray matter and white matter structure, stellar object locations on sky), and half had primarily nonspatial expertise (gene sequences, brain function, spectral data). There was no significant domain expertise overlap among the experts working on a particular case — e.g., a structural biologist (whose expertise covers 3D spatial protein structure) may be superficially familiar, but not conversant, with the genomics terminology — where nonspatial information such as gene sequencing is essential. Both spatial and nonspatial representations were used for common feature and outlier detection.

Our post-hoc analysis finds that, without exception, experts who specialized in spatial structures started their analyses by focusing on the spatial sections of the interface, and then gradually worked up towards the nonspatial sections to further strengthen or refute their observations (Figure 2 left). In contrast, without exception, experts who did not specialize in spatial structures started their analyses with the nonspatial representations — as long as these representations were basic, familiar visual representations such as pixel rows, plots, trees, or networks —, and gradually worked up towards the spatial representations, to see if the spatial layout might give them additional information (Figure 2 right). In general, the multi-view approach allows experts to use any view to form a base, exploratory belief, and the other views to strengthen or refute that initial belief.

4. Discussion and Conclusion

The design experience across these systems points to the benefits of coordinated multiple views in the context of spatial and nonspatial integration, in particular when domain experts have complementary expertise. Our experience has been that a familiar representation (a "scaffold") was essential in orienting the user, and that prototyping attempts at overlaying information over that familiar representation were not successful. Linking information to that familiar

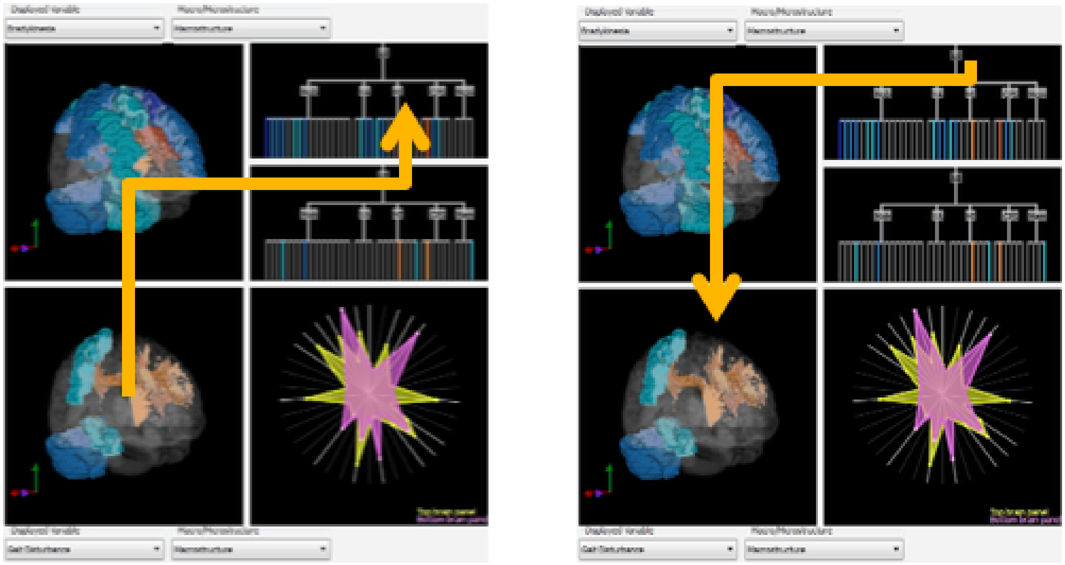

Figure 2: *Two visual analyses in a visual framework for the integration, comparison, and exploration of correlations in spatial and nonspatial geriatric data. Left: a typical analysis lead by a neurosurgeon starts with the familiar brain imaging representation, and then progresses towards the nonspatial representations. Right: a typical analysis lead by an epidemiologist starts with the familiar nonspatial representation of brain function clusters, and progresses towards the less familiar spatial representation of the brain. In general, the multi-view approach allows experts to use any view to form a base, exploratory belief, and the other views to strengthen or refute that initial belief.*

representation through linked views was, however, successful. The separation into multi-views allowed the experts to gradually build up confidence in the tool. Despite initial resistance to certain nonspatial representations, the approach made the learning-curve less steep and facilitated early adoption of a tool.

The usage analysis indicates that users with different backgrounds may employ a spatial and nonspatial visualization application differently. This may be particularly important for the design of collaborative meeting environments, in which users contribute complementary expertise. In our work the disadvantages of context-switching between the views were clearly outweighed by the user familiarity with a particular type of analysis (spatial or nonspatial). Finally, linking the views allowed the user to harness and expand their previous analysis experience.

While arguably the participatory design process influences the subsequent usage of the system, the design of collaborative visual spaces for expert problems is difficult and dangerous in the absence of regular feedback from domain experts. Use analyses are further hampered by the typically small numbers of domain experts available in any given domain of expertise. However, we note that "large scale studies" on the equivalent of rare "rocket scientists" are likely unfeasible, as are eye-tracking studies — which require expert-user commitment to the study. Furthermore, collaborative visual analyses that require complementary expertise in spatial and nonspatial data involve typically domain experts, and not novice users. Last but not least, visual scaffolding, just like any scaffolding, should be ultimately removed; it is not clear how that could be accomplished in problem-driven visualization, once the users' "dream tool" has been deployed and adopted.

While the domain experts are already motivated by compelling tasks, and visualization researchers sometimes offer templates and guides in the form of demonstrations and tutorials, further visual guidance in the form of familiar representations appears to be necessary. Linked view designs can effectively serve as "visual scaffolding" in the context of collaborative, integrated spatial and nonspatial visual analysis.

In conclusion, our analysis supports a multi-view design of visual tools that seek to integrate spatial and nonspatial information. While the information would thus be somewhat fragmented, our evidence points to the benefits of separating the views.

Acknowledgments:

This work has been supported by NSF CAREER IIS-0952720 and NSF CBET-1250171. This work would not have been possible without my students T. Luciani, A. Maries, S. Myers, D. Oliphant, J. Wenskovitch, or without our wonderful collaborators across disciplines.

References

[DH01] DOLEISCH H., HAUSER H.: Smooth brushing for focus+context visualization of simulation data in 3d. In *Journal of WSCG* (2001), pp. 147–154. 1

[GDJ*13] GOODWIN S., DYKES J., JONES S., DILLINGHAM I., DOVE G., DUFFY A., KACHKAEV A., SLINGSBY A., WOOD J.: Creative user-centered visualization design for energy analysts and modelers. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2516–2525. doi:http://doi.ieeecomputersociety.org/10.1109/TVCG.2013.145. 1

[KKH02] KNISS J., KINDLMANN G., HANSEN C.: Multidimensional transfer functions for interactive volume rendering. *IEEE Transactions on Visualization and Computer Graphics 8*, 3 (July 2002), 270–285. URL: http://dx.doi.org/10.1109/TVCG.2002.1021579, doi:10.1109/TVCG.2002.1021579. 1

[KSDK11] KOH L. C., SLINGSBY A., DYKES J., KAM T. S.: Developing and applying a user-centered model for the design and implementation of information visualization tools. In *Information Visualisation (IV), 2011 15th International Conference on* (2011), pp. 90 – 95. © 2011 IEEE.. URL: http://openaccess.city.ac.uk/555/. 1

[LCO*14] LUCIANI T., CHERINKA B., OLIPHANT D., MYERS S., WOOD-VASEY W., LABRINIDIS A., MARAI G.: Large-scale overlays and trends: Visually mining, panning and zooming the observable universe. *IEEE Transactions on Visualization and Computer Graphics* (Jul 2014), 1–14. 2

[LWC*14] LUCIANI T., WENSKOVITCH J., CHEN K., KOES D., TRAVERS T., MARAI G.: Fixingtim: Interactive exploration of sequence and structural data to identify functional mutations in protein families. *BMC Proceedings 2014, 8(Suppl 2):S3* (Aug 2014), 1–9. 2

[MMAM14] MCKENNA S., MAZUR D., AGUTTER J., MEYER M.: Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* (2014). URL: http://www.sci.utah.edu/publications/Mck2014a/McKenna_TVCG2014.pdf. 1

[MMH*13] MARIES A., MAYS N., HUNT M. O., WONG K. F., LAYTON W., BOUDREAU R., ROSANO C., MARAI G. E.: Grace: A visual comparison framework for integrated spatial and non-spatial geriatric data. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2916–2925. doi:http://doi.ieeecomputersociety.org/10.1109/TVCG.2013.161. 2

[MPLB07] MOORE E. B., POLIAKOV A. V., LINCOLN P., BRINKLEY J. F.: Mindseer: a portable and extensible tool for visualization of structural and functional neuroimaging data. *BMC Bioinformatics 8* (2007). 1

[Mun09] MUNZNER T.: A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (Nov. 2009), 921–928. URL: http://dx.doi.org/10.1109/TVCG.2009.111, doi:10.1109/TVCG.2009.111. 1, 2

[NB78] NINIO A., BRUNER J.: The achievement and antecedents of labelling. *Journal of Child Language 5*, 01 (1978), 1–15. 1

[Rhy03] RHYNE T.: Does the difference between information and scientific visualization really matter? *IEEE Computer Graphics and Applications 23*, 3 (2003), 6Ð8. 1

[RWK*99] RUBEL O., WEBER G. H., KERANEN S. V. E., FOWLKES C. C., HENDRIKS C. L. L., SIMIRENKO L., SHAH N. Y., EISEN M. B., BIGGIN M. D., HAGEN H., SUDAR D., MALIK J., KNOWLES D. W., HAMANN B.: Pointcloudxplore: Visual analysis of 3d gene expression data using physical views and parallel coordinates. In *Proceedings of the Eurographics/IEEE-VGTC Symposium on Visualization* (1999). 1

[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 (2012), 2431–2440. doi:http://doi.ieeecomputersociety.org/10.1109/TVCG.2012.213. 1

[SP06] SHNEIDERMAN B., PLAISANT C.: Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization* (New York, NY, USA, 2006), BELIV '06, ACM, pp. 1–7. URL: http://doi.acm.org/10.1145/1168149.1168158, doi:10.1145/1168149.1168158. 1

[TM04] TORY M., MÖLLER T.: Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics 10*, 1 (Jan. 2004), 72–84. URL: http://dx.doi.org/10.1109/TVCG.2004.1260759, doi:10.1109/TVCG.2004.1260759. 1

[WBWK00] WANG BALDONADO M. Q., WOODRUFF A., KUCHINSKY A.: Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (New York, NY, USA, 2000), AVI '00, ACM, pp. 110–119. URL: http://doi.acm.org/10.1145/345513.345271, doi:10.1145/345513.345271. 1

[WKvD*08] WASSINK I., KULYK O., VAN DIJK D. E., VAN DER VEER G., VAN DER VET D. P.: Applying a user-centred approach to interactive visualization design. In *Trends in Interactive Visualization*, Zudilova-Seinstra E., Adriaansen T., van Liere R., (Eds.), Advanced Information and Knowledge Processing. Springer Verlag, London, November 2008. URL: http://doc.utwente.nl/64840/. 1