Authors' manuscript. The published version can be found at
http://ivi.sagepub.com/content/13/4/313
doi: 10.1177/1473871613481693

Augmenting the Educational Curriculum with the VAST Challenge: Opportunities and Pitfalls

Christian Rohrdantz, Florian Mansmann,
Chris North, Daniel A. Keim

January 28, 2013

Abstract

With its mission to move science into practice, the VAST Challenge has become an integrated part of the annual Visual Analytics Science and Technology Conference since its inception in 2006. In this article we discuss how we can transfer this objective into a classroom setting by using the VAST Challenge data sets and by encouraging student submissions to the challenge. By means of Bloom's Taxonomy of Educational Objectives for Knowledge-Based Goals, we show how the VAST Challenge enables the integration of additional learning objectives into two types of courses: A dedicated course that focuses on the contest participation and an integrated course that uses the contest data to emphasize practical course elements. The core contribution of this article is that we assess the opportunities and pitfalls that we experienced at University of Konstanz in Germany and Virginia Tech in the U.S.A. when augmenting the educational curriculum with the VAST Challenge.

1 Introduction

The VAST Challenge, since 2008, is a participation category of the IEEE Visual Analytics Science and Technology (VAST) Symposium/Conference continuing the footsteps of the VAST 2006 and 2007 contests. The challenge provides realistic data sets and visual analytics scenarios with a ground truth that is published after the submission is closed and participation is open to the public. The purposes of the challenge are manifold: to identify the most useful visual analytics practices among all handed-in solutions, allow data analysts to apply and test their newly developed methods, and provide visual analytics lecturers with realistic data for their students to work on. In this article we focus on the latter point and introduce different ways to augment the educational curriculum using the VAST Challenge. We give usage examples for different kinds of courses and challenges, describe our past experiences, point out opportunities and pitfalls, and share our lessons-learned.

The rest of this article is structured as follows: In Section 2, we associate concrete learning objectives to the different steps necessary for working on the challenge tasks and shape the course planning accordingly. In addition, we provide an overview of useful teaching resources. Next, we describe the use of the VAST Challenge either as a dedicated course (cf. Section 3) or integrated into undergraduate or graduate courses (cf. Section 4). In Section 5, we provide our lessons learned from past experiences and discuss opportunities and pitfalls of using the VAST Challenge as part of the educational curriculum. Finally, Section 6 concludes our key findings.

2 Teaching Visual Analytics using the VAST Challenge

There are many good reasons to use material from the VAST Challenge for teaching. It is otherwise difficult to get real-world datasets and if so, typically, no ground truth is available. The challenge scenarios are very motivating for students and help them to put the abstract concepts they learn in their university studies into an exciting real-world analysis context.

In this section we briefly show how the use of the VAST Challenge reflects on the learning objectives of the course. Furthermore, we discuss relevant teaching resources and how software and technology can be used in a meaningful and motivating way.

2.1 Learning Objectives

To systematically plan a course, Bloom's Taxonomy of Educational Objectives [1] can give some guidance to the nature of the learning objectives and how to assess them in the course. For our concrete course planning, we used a revised version [2] of this taxonomy that takes into account more recent research in psychology for ordering the cognitive dimensions.

In principle, this revised taxonomy consists of two dimensions, the *Knowledge Dimension* and the *Cognitive Process Dimension*. As described in more details in Table 2, the Knowledge Dimension consists of factual, conceptual procedural, and metacognitive knowledge at the highest level. Likewise, Table 3 contains an ordered list of verbs that describe simple to complex cognitive processes. At the highest level these dimensional values are *1. Remember*, *2. Understand*, *3. Apply*, *4. Analyze*, *5. Evaluate*, and *6. Create*.

Applied to our course planning, we define our course objectives as follows:

- **Objective O1** – Project Planning
  Students are supposed to plan what they want to do with respect to the employed visualization and analytics methods and how the workload will be distributed among the team members. This objective involves *Organizing (4.2)* in the cognitive dimension as well as *Factual (A)* and *Conceptual Knowledge (B)* in the knowledge dimension.
- **Objective O2** – Application of Learned Methods to Data
  Students are expected to apply the learned visual analytics methods to

<table>
<tr><th rowspan="2">The Knowledge Dimension</th><th colspan="6">The Cognitive Process Dimension</th></tr>
<tr><th>1. Remember</th><th>2. Understand</th><th>3. Apply</th><th>4. Analyze</th><th>5. Evaluate</th><th>6. Create</th></tr>
<tr><td>A. Factual Knowledge</td><td></td><td></td><td></td><td rowspan="2">O1</td><td></td><td></td></tr>
<tr><td>B. Conceptual Knowledge</td><td></td><td></td><td rowspan="2">O2</td><td></td><td rowspan="2">O4</td></tr>
<tr><td>C. Procedural Knowledge</td><td></td><td></td><td></td><td rowspan="2">O3</td></tr>
<tr><td>D. Metacognitive Knowledge</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Table 1: The Classification of Our Learning Objectives According to the Revised Taxonomy.

the contest dataset. This objective involves *Executing (3.1)* tools and methods (e.g., using standard software to display time-series data) and *Implementing* (i.e., adapting or innovating methods for application on the contest data). The focus for this objective is on *Conceptual (B)* and *Procedural Knowledge (C)*.

- **Objective O3** – Evaluation of Project Outcome
  To achieve O3, students should be able to perform the cognitive processes of *Checking (5.1)* and *Critiquing (5.2)* on their project outcome. This involves not only *Procedural (C)*, but also *Metacognitive Knowledge (D)* such as knowledge about cognitive tasks and context *(Db)*.

- **Objective 4** – Presentation of Methods and Results
  During the course students are expected to present their intermediate and final analysis results. For this they need to describe how and why they used certain visual analytics methods to transform the contest data into an interactive visualization. This objective involves *Generating (6.1)* and *Producing (6.3)* from the cognitive dimension as well as *Conceptual (B)* and *Procedural Knowledge (C)*.

Table 1 summarizes our four learning objectives by classifying them into the highest levels of the Revised Taxonomy. In contrast to a classical course that primarily focuses on *remembering*, *understanding*, and *applying* factual and conceptual knowledge, the VAST challenge not only allows students to reach higher cognitive processes such as *analyze*, *evaluate*, and *create*, but also enables them to address all knowledge types as shown above.

2.2 Teaching Resources

Visual Analytics can be described as the interplay between *Information Visualization*, *Knowledge Discovery*, and *Interaction* methods with the goal of enhancing the *cognitive process of analytical reasoning*. For the lecturer and students we therefore recommend a selection of textbooks that should help to gain an overview of each respective domain and to look up specific methods when needed.

The *Information Visualization* field can be approached from different perspectives. Tufte's books (e.g. "Envisioning Information"[3]) mostly convey a critical perspective on the use of diagrams, often with nicely illustrated historical case studies. In 1999 Card, Mackinlay, and Shneiderman published "Readings in Information Visualization: Using Vision to Think"[4], a useful collection of papers that make up the foundation of information visualization. Another perspective is to explicitly assess visualization methods with respect to human perception as done in Colin Ware's book "Information Visualization – Perception for Design"[5]. Finally, Ward, Grinstein, and Keim wrote "Interactive Data Visualization: Foundations, Techniques, and Applications"[6], a textbook for use in classroom teaching. More information can be be inferred from an international survey[7] among professors in the field.

Knowledge Discovery is also a complex topic that can be addressed in many different ways. For the sake of brevity, we limit our recommendation here to Berthold et al.'s "Guide to intelligent data analysis"[8] and Fayyad, Grinstein, and Wierse's book "Information Visualization in Data Mining and Knowledge Discovery"[9] that aims at bridging the gap between knowledge discovery and information visualization.

The *Interaction* field is also very broad and there is thus a need for first focusing on the fundamentals. In his book "Information Visualization: Design for Interaction"[10] Robert Spence describes, for example, how interaction can be designed in such a way that the interface supports exploratory tasks, which are typical for the VAST Challenge. To dig deeper into the topic, "Research Methods in Human-Computer Interaction"[11] by Lazar and Feng is a helpful starting point.

To learn about the *cognitive process of analytical reasoning*, a variety of references are helpful to understand the human dimension of visual analytics. Johnson-Laird's book "How We Reason"[12] provides a broad overview of cognitive issues in reasoning. Heuer's "Psychology of Intelligence Analysis"[13] examines cognitive biases, and "Structured Analytic Techniques for Intelligence Analysis"[14] presents numerous techniques for overcoming bias that are broadly applicable in many domains. Pirolli and Card[15] offer a useful reference model of the analytical process that can guide tool designers. Esser[16] reviews benefits and pitfalls associated with collaborative analysis and groupthink.

While there is no teaching book specifically on the topic of *Visual Analytics* available yet, the book "Illuminating the Path: The Research and Development Agenda for Visual Analytics"[17] by Thomas and Cook nicely introduces the core methodology. Keim et al.'s edited book "Mastering the Information

Age-Solving Problems with Visual Analytics"[18] furthermore investigates specific relevant visual analytics subfields, such as data management, data mining, space and time, infrastructure, perception, and cognition as well as evaluation. Note that both books are accessible on the web without any costs. Finally, Georgia Institute of Technology's "Visual Analytics Digital Library"[19] bundles many valuable resources for teaching visual analytics.

2.3 Software and Technology

Software There are several software solutions available that integrate a large number of data analysis methods, which can readily be applied for data set exploration and generating first hypotheses and findings.

The open-source solutions KNIME[20], WEKA[21], and RapidMiner[22] support the exploration of diverse data types, offer collections of established *machine learning methods*, and basic visualization methods that can be customized through comprehensive user interfaces.

More flexible and *scalable data processing* can be achieved using scripting languages such as PERL[23] or Python[24]. UNIX commands like *grep*, *cut*, *sed*, and *awk* are an alternative.

More *advanced visualizations* can be generated with the open-source software R[25] and its numerous add-on packages or with the Tableau Software[26] which offers a wide range of visualization methods that can be created through an easy-to-use user interface. Moreover, an easily customizable web interface for the creation of diverse visualizations is provided by IBM's Many Eyes[27].

Highly flexible visualizations can be created using open-source libraries like D3 Data-Driven Documents[28,29], a visualization framework written in javascript, or Prefuse[30,31] written in java. For advanced user interactions the java library Piccolo2D[32,33] is recommendable.

Further specialized tools and libraries exist for *certain data types* like GraphViz[34] for graph visualization, or Jigsaw[35,36] for visual text analytics, to name just two prominent out of the numerous examples.

Technology For the solution of some challenge tasks the use of new technology can be very helpful and at the same time motivating for the students. This includes for example the use of large high-resolution screens (see Figure 1), interlinked displays, touch tables, and gesture recognition. That way, more data can be displayed at once, or several coordinated views on the same data can be interactively explored using suitable easy-to-use interaction devices.

3 Applied Visual Analytics: Dedicating a course to the contest participation

Applied Visual Analytics is the title of a course that is organized at the University of Konstanz, Germany, and is based on the VAST Challenge. It gives

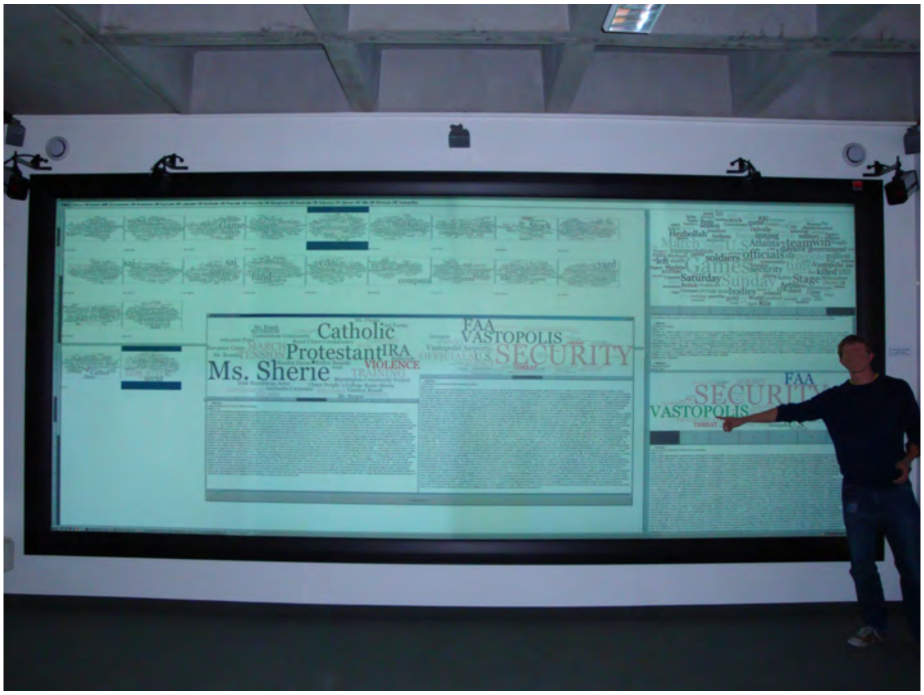

Figure 1: Comparative analysis of text clouds in Mini Challenge 3, 2011, at the Konstanz Powerwall display. Reprinted from Bertini et al.[37] (© 2011 IEEE).

advanced students the opportunity to put theoretical knowledge into practice. The students are required either to bring background knowledge in Data Mining and Information Visualization or visit the corresponding lectures in parallel. The participants are split into teams of 2 or 3 students working together on one Mini-Challenge task. For each Mini-Challenge we aim to have at least one student team, so that we can solve the Grand Challenge[1] joining all efforts.

Teaching the Analytic Process First the students are given brief introductions to standard data analysis and visualization tools as listed in Section 2.3. Next, the students are asked to explore how far they can get with any of those tools analyzing the given data. Usually, at some point the standard tools do not provide further support and students have to develop their own tools for preprocessing, automated data analysis, or visualization. They are asked to assign anything they do to a certain step of the *KDD Process*[38], *Visual Analytics Pipeline*[39], or a similar process flow. This helps them to structure, summarize, reason about, and explain their activities.

[1]Unlike in the year 2012, in previous years there was a Grand Challenge analysis task, that could only be solved when having solutions for all of the Mini Challenges

Iterative Result Refinement / Feedback Sessions The students of a team have to subdivide the tasks among themselves and they have to present their progress every week, as well as problems they face. Results and problems are then discussed with the whole class. This ensures that the other students do not only focus on their tasks, but also actively participate in the solution of the other tasks. From week to week the results become more complete and elaborate. In the end, the whole class is aware of what all teams have done and can bring pieces together for solving the Grand Challenge.

Example for Obtained Analysis Results

The Applied Visual Analytics course has been conducted throughout the last four years and several times has led to the winning of awards[40,41,37], among them Grand Challenge awards in 2009 and 2011. From the 2011 solution we would like to present the Mini Challenge 1 (MC1) solution[37] as an example for results that students have obtained within the course.

The data consists of microblogging messages from mobile devices (similar to Twitter messages) that come with a time stamp and the GPS coordinates of the locations from where messages were sent. Analyzing about 1 million messages sent over the course of three weeks and some additional metadata (a geo-map, information about weather conditions), participants shall characterize the outbreak of an epidemic spread. One of the challenges for the students is to integrate different special data types: time, geo-space, text, and meta-data. The students had to make a decision, to which visual variables to map the different data types and how to convey the inter-dependencies among them.

Figure 2 gives an impression of the tool developed to solve MC 1, showing screenshots of the user interface for three crucial days, May 18th to May 20th. In the middle of each screenshot, a map is displayed where messages are represented as red dots. At the left side, a panel for filtering and configuration is provided and weather conditions, like wind direction and strength, are shown at the top left. The bottom line shows the development of the data volume over time and allows the selection of arbitrary time intervals. Selecting both a time range and a map area the user can narrow down the analysis and request only those messages that pass the filter. From these, tag clouds can be produced on demand as shown at the right side of the display. The data development over time can also be displayed through animation pressing a play button. When monitoring the dataset using animation the three days shown in Figure 2 stick out clearly.

Only those messages are marked as red dots that report about sicknesses according to a keyword-based classification method. It can be seen that from May 18th on people from the central city area report *fever*, *flu*, and *pneumonia*, and from May 19th on people in the southwestern city area report symptoms of *diarrhea*. Searching for possible causes the students designed an algorithm to discover anomalies in the geo-spatial distribution of message content. They could identify a truck accident on a bridge located right in the middle of the two affected areas as the cause. The truck had spilled its apparently contami-

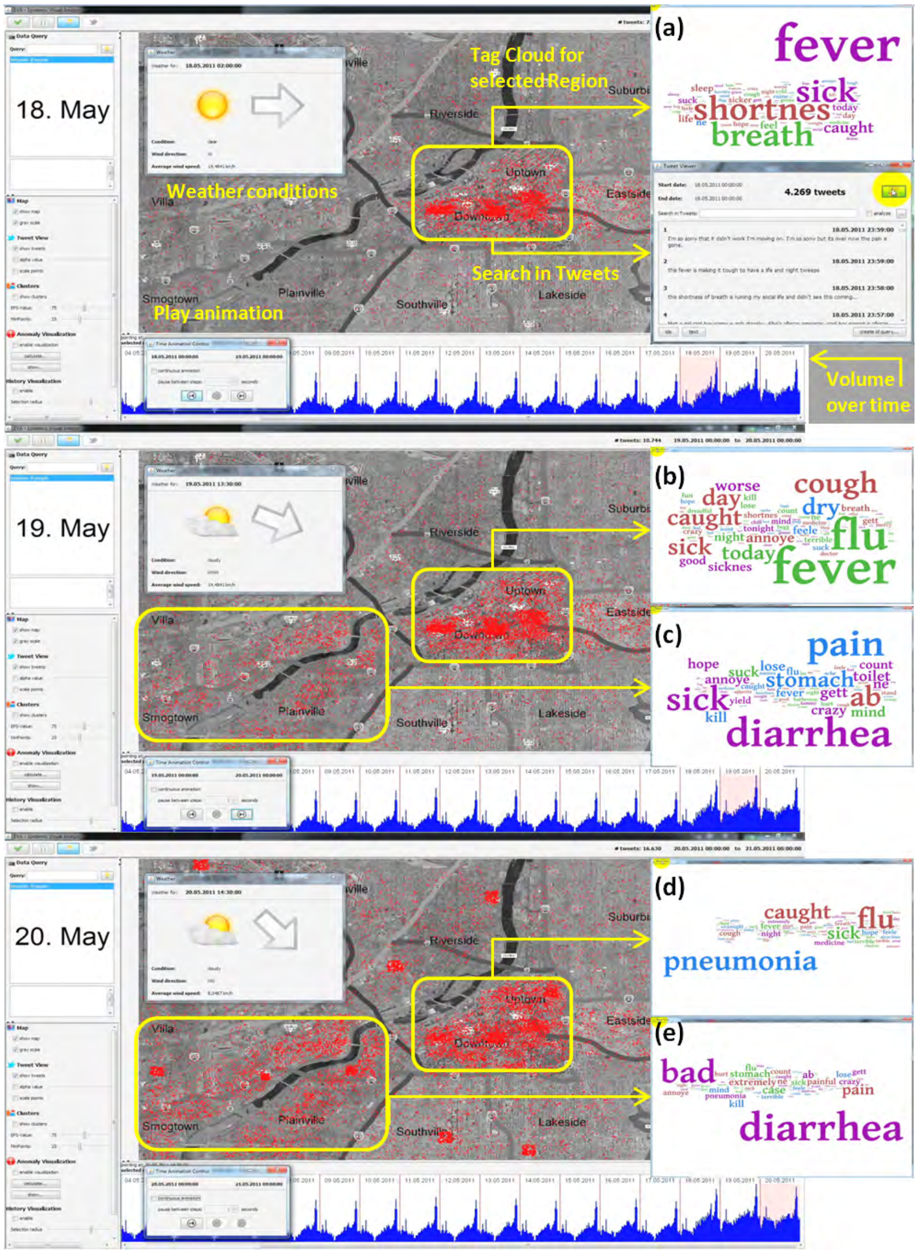

Figure 2: Tool designed for the solution of MC 1, 2011, showing geo-locations and text content of messages for three different days. Reprinted from Bertini et al.[37] (© 2011 IEEE).

nated cargo into the river and the students came up with the hypothesis that the disease spread waterborne along the river and in wind direction towards downtown. More information about the approach is provided in Bertini et al.[37].

The exciting real-world scenario got the students very much involved generating and testing hypotheses about the causes and impacts of the epidemic spread. The Grand Challenge also required entrants to write a debrief, which explains the most relevant analysis results and findings to decision makers. Concrete recommendations for actions that authorities should take had to be deduced from the obtained results.

4 Integrating the VAST Challenge into Courses

The curriculums of two existing computer science courses at Virginia Tech were augmented with the use of the VAST Challenge in the form of semester projects. *Introduction to Human-Computer Interaction* is an undergraduate course in which students learn about HCI theory, the usability engineering process, user interface design principles, and user interface development methods. *Information Visualization* is a graduate course that covers visualization theory, design principles, methods for visual representation, and interaction techniques.

The VAST Challenge provides a specific domain problem scenario to which students apply and practice the HCI or visualization methods that they learn in these courses. Because the Challenge provides a novel scenario, it requires students to synthesize their learned knowledge to create novel approaches. Furthermore, reflecting upon their processes and outcomes enables students to evaluate the appropriateness and successfulness of the learned methods.

This use of visual analytics challenges fits into our educational agenda called "CS for CSI", which seeks to motivate students that computer science is an important field skill in supporting popular domains such as investigative analysis or crime scene investigation.

4.1 Project Structure

Students are assigned a semester-long team project with the objective of creating a software tool that can help analysts solve the VAST Challenge. In this case, the pedagogical goal is not the analytics to solve the challenge per se, but rather the methods employed to create the tool. Thus, the students do not submit materials to the actual VAST Challenge. Instead, we use multiple similar previous Challenge datasets as test cases, and the official Challenge ground-truth solutions to support pedagogical assessment.

The project structure consists of five phases:

1. Requirements analysis: Understand the needs of the analytic user tasks and datasets involved in the Challenge scenario, by attempting to solve it using standard existing tools.

2. Design prototype: Design and implement a new user interface or visualization that will satisfy a need identified in step 1.

3. Evaluation and refinement: Use the new tool to attempt to solve a second similar Challenge scenario over a long period of time, refining the design of the tool throughout the process.

4. Live contest: Evaluate the final tool design by attempting to solve a third similar Challenge scenario during a single session in a competition between all student teams.

5. Presentation: Student teams present results to the entire class.

The students are evaluated on the following criteria in descending order of importance: (1) methods, including processes and techniques, applied; (2) the final design of their software tool; (3) the correctness of their Challenge solutions generated using their tool. The use of previous Challenges with published ground-truth solutions enables immediate assessment feedback to the students. This helps to reinforce course content, and provides a practical mechanism for grading. We score them using metrics similar to those used by the VAST Challenge organizers[42,43].

The Live Contest phase is a recent successful addition to the project structure[44], and is modeled after the analogous live event that took place at the VAST Conference in 2006 and 2007[42,43]. The goal of this phase is to stress test the students' newly designed tools in a time-limited high-pressure environment. Our competition occurs at the end of the semester as a separate 2-3 hour evening session that combines both classes (graduate and undergraduate). Student teams compete against each other in solving a new Challenge scenario that they have not seen before, but is similar in format to the scenarios they tested on earlier in the project. Each team uses its own software tool that they designed during the project. At the end of the session, teams submit their hypothesized solutions to the instructor. Then, all teams meet together to compare results, and winners are announced.

4.2 Examples

The student teams have produced a diverse variety of tools (see Figures 3 and 4). Some focus on the foraging and others focus on the synthesis portions of the sensemaking process. Some focus on visualizations of derived attributes extracted from text data, while others focus on search queries and interactions directly with textual information. Some emphasize automatic data-driven visualizations, while others emphasize user-driven diagramming. Common approaches include variations of network visualization, tag clouds, timelines, geographic overlays, notes organizers, text highlighting, facet browsing, and complex query interfaces. The teams typically are able to discover 50-100% of the ground-truth plot using their tools. An open question is whether we could arrange the project such that all the teams work together to produce an integrated tool. This could expose the students to a larger engineering project, but would reduce each team's autonomy in pursuing their own interests.

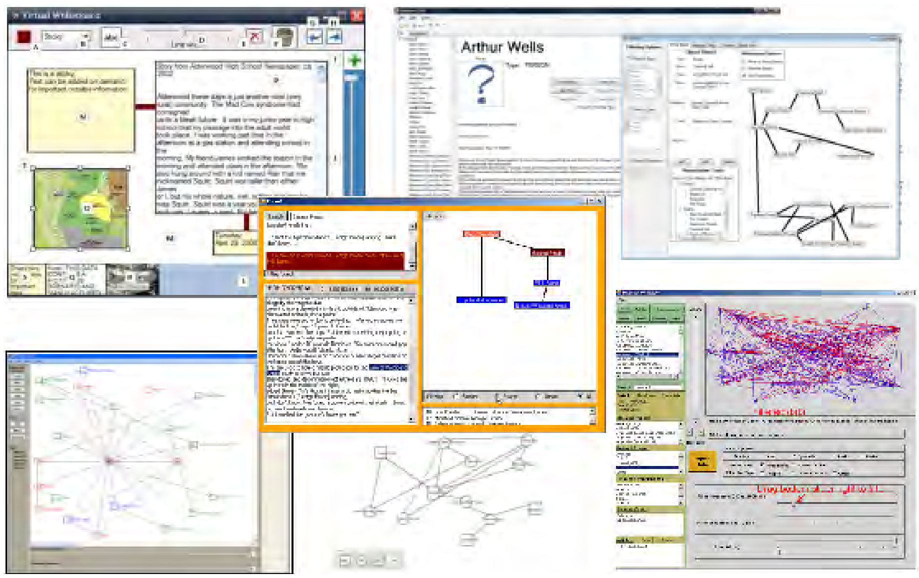

Figure 3: Examples of tools produced by undergraduates in the Human-Computer Interaction course.

Beyond learning the course material, the use of the VAST Challenge also serves to involve students in visual analytics research. A graduate student team that won the Live Contest in their class published a paper about their novel system called VizCept[45], which effectively supports small-team collaborative analytics. Other student projects have served to initiate student research theses and their subsequent related publications[46,47,48,49,50].

4.3 Results

To help assess the pedagogical value of the project structure, we conduct surveys specifically about the project at the conclusion of the most recent offering of the courses. The questions asked for free-text answers, which were then categorized using open-coding, counted, and sorted.

What were the most important lessons you learned in doing the project? The

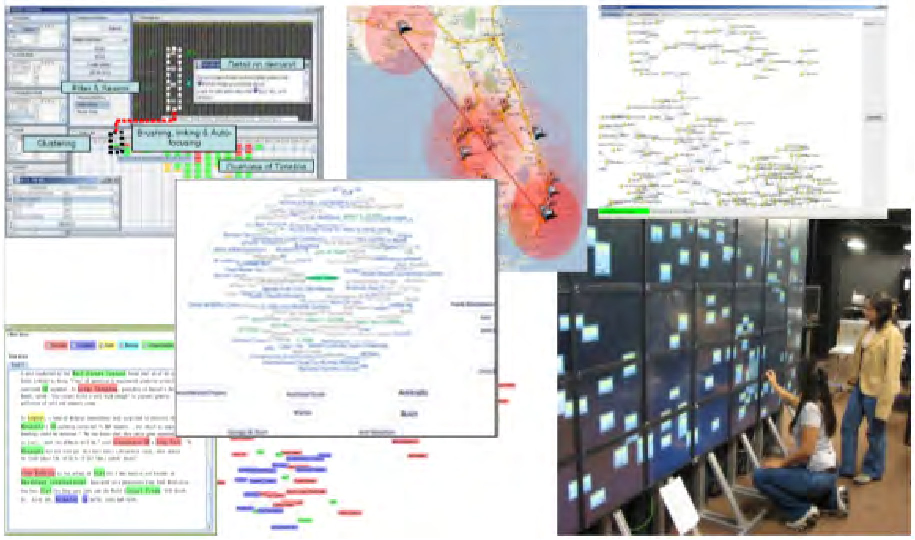

Figure 4: Examples of tools produced by graduates in the Information Visualization course.

undergraduates' top answers were: usability evaluation, teamwork, requirements analysis, new design ideas, iterative design, and user interface implementation techniques. The graduates' top answers were: analytic process, visual perception, collaborative visualization, designing for scalability, and specific visualization techniques. These responses are closely aligned with the pedagogical goals of the courses, indicating that the project is serving the desired intent.

Were the Challenge scenarios interesting? Sixteen of eighteen undergraduate respondents answered affirmatively, and the top cited reasons were: informative about analytics, entertaining, good topic, and realistic. Nine of ten graduate respondents answered affirmatively, with top cited reasons: informative about analytics, helped design, engaging, realistic.

Did the scenarios help you learn more about how your software would be used? Fifteen of eighteen undergraduates answered affirmatively, citing that the scenarios were important to enabling them to identify new requirements for their software and to evaluate their progress. All ten graduate respondents answered affirmatively, citing that the scenarios helped them identify user tasks and roles, evaluate usability, and understand various real problems in analytics.

What did you learn specifically in the live contest? The undergraduates' top answers were: identified changes to their design, their tool actually helped, stress testing, teamwork, and variations in analytic processes. The graduates' top answers were: synchronous collaboration, tracking hypotheses, frustrations of time pressure, variations in analytic processes, and specific needs such as search and filter.

Overall, the live contest in phase 4 was the students' favorite part of the project. The students found this form of evaluation extremely enlightening about the capability and usability of their tools. Being their own users enabled them to experience the usability problems and effects first-hand. Combined with the time pressure of the contest, being expert users enabled them to focus on the performance efficiency of their tools for gaining rapid insight into complex data. They felt the contest represented "real" usage, and clearly exposed the successfulness of their tool design.

5 Lessons learned

In this section we would like to discuss different opportunities and pitfalls of augmenting the educational curriculum with the VAST Challenge. While most of them apply to either of the previously described course types, certain points are particular for dedicated courses (see Section 3) or integrated courses (see Section 4).

5.1 Opportunities

As outlined in Section 2.1, the use of the VAST Challenge offers the great opportunity to readily integrate **high-level learning objectives and knowledge types**, which are otherwise difficult to cover. In contrast to a classical course that primarily focuses on *remembering*, *understanding*, and *applying* factual and conceptual knowledge, the VAST challenge not only allows students to reach higher cognitive processes such as *analyze*, *evaluate*, and *create*, but also enables students to address all knowledge types contained in the Revised Bloom's Taxonomy. Yet, there is a lot more to the use of the VAST Challenge.

Motivational Aspects The practical work and challenge setting are very motivating for students and they usually end up doing much more than the course would have required them to do. In addition, it may be a good opportunity to get students involved with the use of cutting-edge visualization and interaction technology, like gigapixel displays or touch tables, which is another motivating element.

Problem Solving and Creativity The challenge tasks do neither have obvious solutions nor suggest obvious analysis strategies. Students have to continuously search for, test, evaluate, and revoke or refine analysis strategies. At the beginning, this involves becoming familiar with existing algorithms and tools, as described in Section 2.3. Students have to find out by themselves which algorithm and tool is suitable for their purposes and, more importantly, where the limitations are. Finally, their creativity for problem solving is required and trained. They have to reason about novel solutions based on their acquired knowledge and skills, a task that is both demanding and insightful.

Soft Skills Different kinds of soft skills are trained. Students learn that team work and a good organization and distribution of work load is important to get things done in a limited amount of time. Basic project management skills can be obtained in a practical context that come close to what is required in real-world research and industry projects.

In a *dedicated course*, writing a Grand Challenge debrief is good training to abstract from technical details and think beyond the mere analysis, turning findings into actionable knowledge. This important step is not always easy for students, but provides them with a holistic view of their work.

An *integrated course*, in contrast, has the advantage that the ground truth is available right from the start. Thus, lecturers and supervisors can give more informed guidance during the course and have less work load assessing the quality of student solutions, as they know what the challenge requires students to find out. From the students' perspective it enables them to evaluate their solutions against the ground truth, which motivates students to improve their methods and recognize the value of the course content in doing so. With a growing history of VAST Challenges, students have the opportunity to test their methods against multiple similar challenge problems and multiple dataset sizes. Hence, the Challenge lends itself to integration into many different topical courses where such test cases are needed, including visualization, usability engineering, and data mining. Instructors must be careful, though, that some of the official solutions to past Challenges are publicly available online, and so some students might be tempted to cheat in a class competition. Ultimately, the Challenges push students to become highly qualified tool builders, which is an important goal of most integrated courses in engineering.

5.2 Pitfalls

Of course, there are also some pitfalls that lecturers and supervisors should always bear in mind.

Student's Pre-Knowledge Missing skills or different levels of pre-knowledge among the course attendees can hinder the progress of solving the challenge. Moreover, students are required to work and learn quite independently and it is important to prevent them from taking a wrong direction. Especially in the *dedicated course*, students should be advanced and have a solid background knowledge in order to be able to fulfill the goal of the course, namely handing in a coherent solution in time.

Increased Supervision Effort Often basic domain knowledge is required for solving the challenge tasks, that may refer to diverse fields such as the analysis of computer networks, genes, text documents, geo-spatial and movement data, time series, social networks, medical records, phone records, videos, and combinations of aforementioned data types. As it cannot be expected that students have this knowledge it must be conveyed throughout the course. In the worst case this implies that lecturers and supervisors must acquire by themselves the respective knowledge about the domain and domain-specific computer science methodology, which can heavily increase their work load. Another problem is the limited scalability of the course, which especially for the *dedicated course* is an issue. In that case, the supervision effort implies regular meetings and feedback sessions, so that either only a limited number of student teams can be properly supervised or additional faculty staff has to get involved. In the optimal case, several supervisors are available for the course who cover different backgrounds.

Scope Creep In the case of the *integrated course*, these pitfalls play out slightly differently. The Challenges often require multiple skills, but an integrated course typically focuses on one such skill, such as visualization or data mining, and it is unreasonable to expect all necessary prerequisites. Thus the assigned project must be carefully scoped to reduce complexity, perhaps by providing pre-processed data or smaller datasets. Similarly, students may become overwhelmed by the feature-creep if they attempt to build tools that aim for comprehensive solutions, and should be cautioned to focus their attention on a well-defined portion of the larger problem. Some ambitious students attempt high-risk solutions using novel technologies such as gigapixel displays that may be outside the scope of the course content, but often fail to complete their project due to the extra learning load, and can require alternative evaluation metrics. Projects that carefully define their scope early, with approval feedback from the instructor, are the most successful.

6 Conclusions

As shown, augmenting the curriculum with a realistic data analysis scenario brings a number of educational benefits. While the classical model of lectures has some advantages, it is known that it has to be complemented by more practical educational elements. Traditionally, these come in the form of hands-on exercises and tutorials. In a sense, the VAST Challenge is an exercise that cannot be solved within a couple of hours, but rather requires weeks of continuous incremental work. Apart from presupposing a proper organization and planning, it brings an additional dimension to teaching and learning: It leaves a much wider space for creativity, as it is not quite clear beforehand what a good solution could look like. Students thus get to explore the whole space of possibilities and are forced to reason about which of the numerous algorithms and techniques they have been taught might be suitable. In the end, creative combinations often lead to good results. Going through such a process of high-level problem solving trains students in a way of thinking that is important both in science and industry: Beyond the mere reproduction of knowledge, they learn to translate it into novel applications.

Acknowledgement

We thank Franz Wanner for his pedagogical input to this paper. This research was partly funded by the DFG (German Research Foundation) within the priority research program SPP-1355 "Scalable Visual Analytics", by the BMBF (German Federal Ministry of Education and Research) within the research project "Visual Analytics for Security Applications", by NSF FODAVA grants CCF-0937133 and CCF-0937071, and the DHS VACCINE Center of Excellence.

References

1. Bloom B, Engelhardt M, Furst E, Hill W, Krathwohl D. Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. Longmans; 1956.
2. Anderson L, Krathwohl D, Airasian P, Cruikshank K, Mayer RP, Pintrich P. P., Raths, J. and Wittrock, M.(2001). A taxonomy for learning, teaching and assessing: A revision of Blooms Taxonomy of Educational Objectives. New York: Longman;.
3. Tufte ER. Envisioning Information. Graphics Press; 1990.
4. Card SK, Mackinlay JD, Shneiderman B. Readings in information visualization: using vision to think. Morgan Kaufmann; 1999.
5. Ware C. Information visualization: perception for design. 3rd ed. Morgan Kaufmann; 2012.
6. Ward M, Grinstein G, Keim D. Interactive data visualization: foundations, techniques, and applications. AK Peters, Ltd.; 2010.
7. Kerren A, Stasko J, Dykes J. Teaching Information Visualization. In: Information Visualization. vol. 4950 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg; 2008. p. 65–91.
8. Berthold M, Borgelt C, Höppner F, Klawonn F. Guide to intelligent data analysis: how to intelligently make sense of real data. Springer; 2010.
9. Fayyad UM, Wierse A, Grinstein GG. Information visualization in data mining and knowledge discovery. Morgan Kaufmann Pub; 2002.
10. Spence R. Information Visualization: Design for Interaction. 2nd ed. Pearson, Prentice Hall; 2007.
11. Lazar J, Feng JH, Hochheiser H. Research methods in human-computer interaction. Wiley; 2010.
12. Johnson-Laird P. How We Reason. Oxford University Press; 2009.
13. Heuer RJ. Psychology of Intelligence Analysis. Center for the Study of Intelligence; 1999.
14. Heuer RJ, Pherson RH. Structured Analytic Techniques for Intelligence Analysis. CQ Press College; 2010.

15. Pirolli P, Card S. The Sensemaking Process and Leverage Points for Analyst Technology as Identified through Cognitive Task Analysis. 2005;.
16. Esser JK. Alive and Well after 25 Years: A Review of Groupthink Research. Organizational Behavior and Human Decision Processes. 1998 February;73(2-3):116–141.
17. Thomas JJ, Cook KA. Illuminating the Path: The Research and Development Agenda for Visual Analytics. National Visualization and Analytics Ctr; 2005.
18. Keim DA, Kohlhammer J, Ellis G, Mansmann F, editors. Mastering The Information Age-Solving Problems with Visual Analytics. Eurographics; 2010.
19. Georgia Institute of Technology's "Visual Analytics Digital Library": http://vadl.cc.gatech.edu;.
20. KNIME: http://www.knime.org;.
21. WEKA: http://www.cs.waikato.ac.nz/ml/weka/;.
22. RapidMiner: http://www.rapidminer.com;.
23. PERL: http://www.perl.org;.
24. Python: http://www.python.org;.
25. R: http://www.r-project.org;.
26. Tableau: http://www.tableausoftware.com;.
27. IBM Many Eyes: http://www-958.ibm.com/software/data/cognos/manyeyes/;.
28. D3 Data-Driven Documents: http://www.d3js.org;.
29. Bostock M, Ogievetsky V, Heer J. D^3 Data-Driven Documents. Visualization and Computer Graphics, IEEE Transactions on. 2011;17(12):2301–2309.
30. Prefuse: http://www.prefuse.org;.
31. Heer J, Card SK, Landay JA. Prefuse: a toolkit for interactive information visualization. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM; 2005. p. 421–430.
32. Piccolo2D: http://www.piccolo2d.org;.
33. Bederson BB, Grosjean J, Meyer J. Toolkit Design for Interactive Structured Graphics. IEEE Trans Software Eng. 2004;30(8):535–546.
34. GraphViz: http://www.graphviz.org;.
35. Jigsaw: http://www.cc.gatech.edu/gvu/ii/jigsaw/;.
36. Stasko JT, Görg C, Liu Z. Jigsaw: supporting investigative analysis through interactive visualization. Information Visualization. 2008;7(2):118–132.
37. Bertini E, Buchmüller J, Fischer F, Huber S, Lindemeier T, Maass F, et al. Visual analytics of terrorist activities related to epidemics. In: 2011 IEEE Conference on Visual Analytics Science and Technology, VAST 2011, Providence, Rhode Island, USA, October 23-28, 2011. IEEE; 2011. p. 329–330.
38. Fayyad UM, Piatetsky-Shapiro G, Smyth P. The KDD Process for Extracting Useful Knowledge from Volumes of Data. Commun ACM. 1996;39(11):27–34.
39. Keim D, Andrienko G, Fekete JD, Görg C, Kohlhammer J, Melançon G. Information Visualization. Berlin, Heidelberg: Springer-Verlag; 2008. p. 154–175.

40. Bak P, Rohrdantz C, Leifert S, Granacher C, Koch S, Butscher S, et al. Integrative visual analytics for suspicious behavior detection. In: Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, IEEE VAST 2009, Atlantic City, New Jersey, USA, 11-16 October 2009, part of VisWeek 2009. IEEE; 2009. p. 253–254.
41. Astefanoaie A, Bozianu R, Broghammer M, Jungnickel R, Rohrdantz C, Schniertshauer J, et al. Visual analytics of a pandemic spread. In: Proceedings of the IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2010, Salt Lake City, Utah, USA, 24-29 October 2010, part of VisWeek 2010. IEEE; 2010. p. 277–278.
42. Costello L, Grinstein GG, Plaisant C, Scholtz J. Advancing user-centered evaluation of visual analytic environments through contests. Information Visualization. 2009;8(3):230–238.
43. Plaisant C, Grinstein GG, Scholtz J, Whiting MA, O'Connell TA, Laskowski SJ, et al. Evaluating Visual Analytics at the 2007 VAST Symposium Contest. IEEE Computer Graphics and Applications. 2008;28(2):12–21.
44. Whiting MA, North C, Endert A, Scholtz J, Haack J, Varley C, et al. VAST contest dataset use in education. In: Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, IEEE VAST 2009, Atlantic City, New Jersey, USA, 11-16 October 2009, part of VisWeek 2009. IEEE; 2009. p. 115–122.
45. Chung H, Yang S, Massjouni N, Andrews C, Kanna R, North C. VizCept: Supporting synchronous collaboration for constructing visualizations in intelligence analysis. In: Proceedings of the IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2010, Salt Lake City, Utah, USA, 24-29 October 2010, part of VisWeek 2010. IEEE; 2010. p. 107–114.
46. Endert A, Fiaux P, North C. Semantic interaction for visual text analytics. In: Konstan JA, Chi EH, Höök K, editors. CHI Conference on Human Factors in Computing Systems, CHI '12, Austin, TX, USA - May 05 - 10, 2012. ACM; 2012. p. 473–482.
47. Andrews C, North C. Analysts Workspace: An Embodied Sensemaking Environment For Large, High-Resolution Displays. In: Proceedings of the IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2012, Seattle, Washington, USA, 14-19 October 2012, part of VisWeek 2012. IEEE; 2012. p. 123–131.
48. Fiaux P. Solving Intelligence Analysis Problems using Biclusters; 2012. MS Thesis, Virginia Tech.
49. Hossain MS, Butler P, Boedihardjo AP, Ramakrishnan N. Storytelling in entity networks to support intelligence analysts. In: Yang Q, Agarwal D, Pei J, editors. The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, Beijing, China, August 12-16, 2012. ACM; 2012. p. 1375–1383.
50. Maiti D. Multiset Model Selection and Averaging, and Interactive Storytelling; 2012. PhD Dissertation, Virginia Tech.
51. Krathwohl DR. A Revision of Bloom's Taxonomy: An Overview. Theory into Practice. 2002;41(4):pp. 212–218.

A Dimensions of the Revised Taxonomy

A. **Factual Knowledge** – The basic elements that students must know to be acquainted with a discipline or solve problems in it.
Aa. Knowledge of terminology
Ab. Knowledge of specific details and elements

B. **Conceptual Knowledge** – The interrelationships among the basic elements within a larger structure that enable them to function together.
Ba. Knowledge of classifications and categories
Bb. Knowledge of principles and generalizations
Be. Knowledge of theories, models, and structures

C. **Procedural Knowledge** – How to do something; methods of inquiry, and criteria for using skills, algorithms, techniques, and methods.
Ca. Knowledge of subject-specific skills and algorithms
Cb. Knowledge of subject-specific techniques and methods
Cc. Knowledge of criteria for determining when to use appropriate procedures

D. **Metacognitive Knowledge** – Knowledge of cognition in general as well as awareness and knowledge of one's own cognition.
Da. Strategic knowledge
Db. Knowledge about cognitive tasks, including appropriate contextual and conditional knowledge
Dc. Self-knowledge

Table 2: Structure of the Knowledge Dimension of the Revised Bloom Taxonomy According to Krathwohl[51].

1.0 **Remember** – Retrieving relevant knowledge from long-term memory.
1.1 Recognizing
1.2 Recalling

2.0 **Understand** – Determining the meaning of instructional messages, including oral, written, and graphic communication
2.1 Interpreting
2.2 Exemplifying
2.3 Classifying
2.4 Summarizing
2.5 Inferring
2.6 Comparing
2.7 Explaining

3.0 **Apply** – Carrying out or using a procedure in a given situation.
3.1 Executing
3.2 Implementing

4.0 **Analyze** – Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose.
4.1 Differentiating
4.2 Organizing
4.3 Attributing

5.0 **Evaluate** – Making judgements based on criteria and standards.
5.1 Checking
5.2 Critiquing

6.0 **Create** – Putting elements together to form a novel, coherent whole or make an original product.
6.1 Generating
6.2 Planning
6.3 Producing

Table 3: Structure of the Cognitive Process Dimension of the Revised Bloom Taxonomy According to Krathwohl[51].