

VAST Contest Dataset Use in Education

Mark A. Whiting¹, Chris North², Alex Ender², Jean Scholtz¹, Jereme Haack¹, Carrie Varley¹, Jim Thomas¹

¹Pacific Northwest National Laboratory, ²Virginia Tech, Department of Computer Science

ABSTRACT

The IEEE Visual Analytics Science and Technology (VAST) Symposium has held a contest each year since its inception in 2006. These events are designed to provide visual analytics researchers and developers with analytic challenges similar to those encountered by professional information analysts. The VAST contest has had an extended life outside of the symposium, however, as materials are being used in universities and other educational settings, either to help teachers of visual analytics-related classes or for student projects. We describe how we develop VAST contest datasets that results in products that can be used in different settings and review some specific examples of the adoption of the VAST contest materials in the classroom. The examples are drawn from graduate and undergraduate courses at Virginia Tech and from the Visual Analytics “Summer Camp” run by the National Visualization and Analytics Center in 2008. We finish with a brief discussion on evaluation metrics for education.

KEYWORDS: education, evaluation, synthetic data.

INDEX TERMS: K.3.2 [Computer and Information Science Education]: Curriculum; H.5.1 [Multimedia Information Systems]: Evaluation/Methodology

1 INTRODUCTION

Visual analytics appears in the classroom through the creation of visual analytics curriculum in universities and through special-purpose training classes. Some classes are information visualization classes being reformulated with a visual analytics consideration. Others are new classes specifically created to teach aspects of visual analytics. Visual analytics classes need realistic tasks and data that is similar to that used by professional analysts, so that students can best begin to appreciate the work of professional analysts.

This kind of task information and data is not readily available to schools. Problems and data are often classified or otherwise restricted due to confidentiality concerns. When problems tackled by analysts are available in the literature, the data associated with them are not usually present.

The National Visualization and Analytics Center (NVAC) (see <http://nvac.pnl.gov>) recognized the difficulties visual analytics researchers would face in the development of applications, requiring data to test and evaluate their systems. The Threat Stream Generator (TSG) project [24] was initiated in 2004 to develop approaches for generating realistic, synthetic test data and

to provide challenges and datasets for researchers across the community to use. The TSG team has been making tasks and datasets available primarily through the IEEE Visual Analytics Science and Technology (VAST) contest since its inception in 2006. There are now seven complex, heterogeneous challenges and datasets available for researchers to use for evaluation as a result of this contest work. The datasets have been downloaded over 600 times to date.

The challenges and datasets have also had a considerable life outside of the VAST contests. Researchers use them to help assess visual analytics software in government, commercial, and academic settings. A very exciting application of the challenges and datasets has been in support of visual analytics education. In this paper, we briefly review the design and creation of visual analytics challenges and processes and discuss their use outside of a contest setting. We describe our experiences in using VAST tasks and datasets in visual analytics coursework at Virginia Tech. A version of a VAST challenge and dataset was used at the 2008 NVAC Visual Analytics Summer Camp, which is also presented. Finally, we discuss evaluation challenges, as supported by VAST datasets in the education setting and discuss additional work needed for educational uses.

2 VISUAL ANALYTICS CHALLENGE AND DATASET DESIGN

The general process for creating a visual analytics challenge problem and synthetic data, such as that used for the VAST contest, is described in [24]. When a challenge and a dataset are reused in a different setting than a contest, it is vital to consider the requirements of the new application and what re-engineering is needed for both task and data.

2.1 Dataset Design Factors

Over the course of our research into the creation of visual analytics challenges and data, we rely on the factors depicted in Figure 1 to guide the development process.

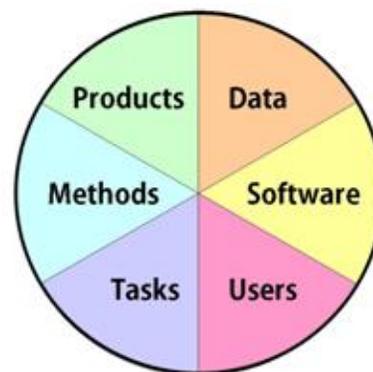


Figure 1: Dataset design factors

Each of these factors must be well-considered for the challenge and dataset to be successful in a particular application. “Data”

1. Email: firstname.lastname@pnl.gov
2. Email: north@vt.edu, aendert@vt.edu

encompasses the type, format, size, heterogeneity, and other characteristics of the challenge dataset. “Software” is the visual analytic application to be assessed. “Users” represent the group using the challenge and dataset, for example, university researchers. “Tasks” are the analyses to be performed with the visual analytic tool. “Methods” are the analytic techniques to be used. “Products” are the anticipated outputs of the visual analytics tools. All of these factors need to be considered to create a successful learning exercise. For example, a numeric dataset (as considered under the Data factor) cannot be analyzed by a text-processing software system (Software). Also, a requirement to use the “Analysis of Competing Hypotheses” method (Methods) will be impossible if it is not supported by the analysis tool (Software) or understood by those doing the analysis (Users). The difficulty in re-engineering a challenge and dataset varies depending on the original development goals compared to the new application goals.

2.2 Designing a Contest Dataset

The VAST 2007 contest featured a scenario involving ecoterrorism and illegal activities related to its endangered species theme [6]. The dataset “Blue Iguanodon” contains heterogeneous data, including news stories, blogs, images, mixed data spreadsheets, and background text. We provide a short description of the design of this dataset as it is one of the most complex contest datasets, and it is one that has been used by Virginia Tech in the classroom.

The VAST contests always have a scenario and embedded ground truth to aid in evaluation. Blue Iguanodon was inspired by the real-life problems of exotic animal smuggling, which is estimated by some to be the second largest black market after illegal drugs with sales estimated at \$10-20 billion [12].

The Blue Iguanodon scenario focuses on the activities surrounding Luella Vedric, a New York socialite, animal rights champion, but a secret marketer of illegal exotic animals, a trade that provides her with a sizeable income. One subplot involves African animals that come to Luella’s customers through a connection with a traveling circus owner named Abu Hassan. Abu moves his circus through several African countries, and animals of several species are added or removed from their performance registries along the way. The second subplot involves exotic fish smuggling. On top of the fish deals, the operation is also smuggling liquid cocaine from Peru into the U.S. via the water in the fish containers. The clever scheme has the packaging comprising two clear plastic bags, with the lining between the inside and outside bags containing the liquid cocaine. The third subplot involves the smuggling of wild chinchilla from Chile that is taking advantage of a huge swell in chinchilla popularity among pet owners. Cesar Gil, a chinchilla distributor in the Los Angeles area (a large chinchilla market), is also a

member of a radical animal rights movement, is distressed by the harvesting of wild chinchilla in Chile, and has decided to infect chinchilla with monkeypox to stop the harvest (i.e., sacrifice a few to save the many). His plan is remarkably successful, as people do not have immunization against smallpox (having been eradicated some years ago) which would have protected them against monkeypox.

Since the contest allowed five months for analysis, we believed that the complexity of the analytical tasks and data were appropriate for the contest. One comment we received from the previous contest was that some groups do not have entity extraction software available to them to assist in analyzing text data. Therefore, we provided pre-processed text, that is, an entity database derived from the news stories.

A large piece of the Blue Iguanodon dataset is text, so we spent considerable time on threat insertion for news articles. Threat insertion is the placement of clues in a text corpus, which may be discovered through analysis. In this case, the text corpus was a blog with posts focused on animal rights news. There are several aspects of threat insertion that must be addressed including:

- Total number of articles compared to number of articles containing clues
- Matching the format and writing style of inserted articles to pre-existing articles
- Making sure that inserted events fit into a timeline
- Amount of information revealed per clue
- Ensuring the total amount of information revealed “tells the story”

We use a tool called the Threat Definition Matrix (TDM) [24] to track this information and to help evaluate the overall complexity of the scenario and dataset (Figure 2). The TDM allows us to specify the scenario time, the subplot, a particular event being addressed, the dataset where the threat element is inserted, the type of cue, its subtlety, the data format, exactly what is being revealed in the plot, and a pointer to the location of this clue. The TDM is our primary tool for tracking the “who-what-when-where-why-and how” of the scenario.

It is difficult to provide quantitative metrics on the dataset, such as signal to noise of the clues to surrounding data, since clues have considerable qualitative features [24]. In the news article dataset, threat information was inserted into 24 of 1455 articles. This does not take into consideration the amount or importance of information inserted per item. In the main terrorist blog, every cartoon entry had relevant information. The VAST contest team felt that the handling of a blog plus images itself would prove a significant challenge itself.

Blue Iguanodon has a companion dataset named White Smilodon that was used in the VAST 2007 interactive session and at the NVAC Summer Camp evaluation session described below. White Smilodon is similar in structure to Blue Iguanodon, but has

Scenario Date	Plot	Event	Dataset	Type	Subtlety	Format	Elements	Link
8/19/2003	Chinchilla Bioterror	Statement by Faron	AR Blog	News article	Low	Text	Link of Faron to Collie	Article: Week-of-Mon-20030818.txt_23.xml
8/22/2003	Chinchilla Bioterror	Wild chinchilla poaching in Chile reported	AR Blog	News article	Low	Text	Chins are protected in Chile, but poaching is reported	Article: Week-of-Mon-20030818-1.txt_44.xml
9/1/2006	Chinchilla Bioterror	Gil starts a wild chin farm in LA area	AR Blog	News article	Medium	Text	Gil is kicking off his plot...the article is more like an ad	Article: Week-of-Mon-20030901-1.txt_36.xml

Figure 2: Threat Definition Matrix.

a simpler scenario, fewer news stories, fewer images, and was designed so that analysts could make significant progress in the three hours of the interactive session.

The TSG team has worked with instructors and others wishing to use the datasets in educational settings to ensure their proper application and that the necessary changes are made for their use. In general, we try to adjust for the following characteristics:

- Users: Students are just being introduced to visual analytics concepts.
- Software: Software will be constructed from scratch as part of the class.
- Tasks: Analyses must be performed during a very short classroom schedule.
- Methods: Students will have very limited or no training in analytical methods.
- Products: Class projects must conform to the instructor's goals, emphasizing the class requirements.

For the most part, the task and dataset changes involved ensuring the task complexity is reduced from that of the contest, the task must allow demonstration of the class requirements in the use of software, the data must be reduced in complexity and size, and there should be little expectation of the application of formal analytical methods. With careful considerations, the VAST datasets can then provide a valuable tool for visual analytics education.

3 PEDAGOGICAL FOUNDATION

There have been several requests to the NVAC TSG team by university professors wishing to use the datasets in a classroom setting to help teach visual analytics, information visualization, or related classes involving visualization. Several student teams have participated in the VAST challenges, and we believed that incorporating a challenge in a visual analytics class would be a good opportunity to expose more students to the kind of scenarios and data facing information analysts outside of the contest setting. Hands-on experience with solving practical real-world problems and exercises is an important component of technology education [21].

The use of the scenarios, tasks, and datasets allows the instructor to employ an anchored instruction approach [2][11][13] to teaching visualization and critical analytic thinking. We enhance learning about analysis “by creating environments that permit sustained exploration by students and teachers and enable them to understand the kind of problems and opportunities that experts in various areas encounter and the knowledge that these experts use as tools” [18]. Anchored instruction, a derivative of situated learning, emphasizes learning activities designed around realistic case studies or problem scenarios (the anchor), and materials that promote exploration by the learner, resulting in active learning of complex concepts [4]. For an anchored instruction practitioner, an anchor might be a video containing a complex problem with embedded data that should be used to solve the problem. For example, the movie “The Young Sherlock Holmes” was used as a primary anchor, and students were asked to examine the film in terms of causal connections, motives of the characters, and authenticity of the settings in order to understand the nature of life in Victorian England [23].

In our application, the anchor is the VAST scenario, tasks, and dataset. The context used by situated learning practitioners in the

late 1980s and early 1990s was videodisks so that students could experience an interactive visual format for their explorations. We substitute the visual analytics tool's environment for exploration of the VAST challenges, gaining the added advantage that along with learning about analysis, the advantages and disadvantages of visual tools can be examined. A very similar approach to the use of the VAST challenge data in visual analytics classes can be seen in the Vanderbilt Cognition and Technology Group's Jasper Woodbury Adventures, an anchored instruction series that focuses on mathematical problem finding and problem solving [2]. There are several related and derivative approaches for anchored instruction, which have led up to the current focus on the learning potential of digital games in the classroom (see the discussion in Van Eck [22]), that may be applicable to future VAST challenges that are considered for classroom use. Games are typically used in technology education to help students learn content in one of two fashions. Students either play a game designed by the instructor to reveal the content, or students design a new game that applies the content. The use of VAST challenge data in education can partially combine these approaches, in that students design new tools to help themselves play the game. That is, to solve the analytic scenario, students construct new visual analytics tools. Thus, it can be used to help students learn about both analytic methods and engineering methods.

4 VAST DATASETS IN UNIVERSITY COURSES

We present experiences and lessons learned from their usage in two Computer Science courses at Virginia Tech. The VAST datasets are applied in these courses from the perspective of constructing interactive software tools that support analysts in analyzing datasets like these. That is, these are courses primarily about tool building, and less so about the practice of analysis.

4.1 Graduate Course on Information Visualization

CS5764 Information Visualization (InfoVis) is a research-oriented graduate course that focuses on the design of visual representation and interaction techniques (<http://infovis.cs.vt.edu/cs5764/>). In the course, students learn about representations and interactions for different types of data, design principles and theoretical underpinning of visualization, implementation and evaluation methods, some existing tools, and a portion of the current literature. The course outline is loosely based on several publications [3][13][20].

The IEEE VAST 2007 Challenge, which used the Blue Iguanodon dataset [6], was assigned as the class semester project without modification. This dataset contains approximately 1700 documents, such as news stories and blogs, and a pre-processed entity database of approximately 8000 entities extracted from the text documents. Students worked in teams of three to four people to design and build visualization tools that would help them identify the ground-truth story hidden within the dataset. Unfortunately the course occurred after the submission deadline for the contest (fall semester), so students could not submit their solutions. Nevertheless, the dataset proved very valuable to the pedagogical goals of the course.

We initially believed that the project might be too difficult, but were surprised that three of the six teams succeeded in uncovering most of the solution with their tools. Two of these teams focused on network visualizations that enabled them to find important social networks or entity chains within the document collection. The third team created a novel approach for keyword queries with visualization of intersections between results. The remaining three teams got stuck examining surface-level distracters in the

dataset. The visualizations that were least helpful were those that focused primarily on geography or time attributes.

The assignment was challenging for the students. The text-intensive dataset is not a simple multidimensional tabular database that would afford straightforward visual encodings. Also, the pre-processed entity lists were automatically generated and so contained a significant amount of noise. For example, a single person's name might be in the list several times in different formats (e.g. different orders of first/middle/last names, initials, prefix, suffix, etc.). Hence, the data required students to undertake data processing steps that are not necessarily relevant to the specific goals of an Information Visualization course. However, it did give students an appreciation for the broader challenges associated with analysis of realistic data. The students rose to the challenge, but it would be better to slightly simplify the entity dataset in the future for courses like this one.

Because an important measure of the quality of a visualization is its support for finding answers, finding the dataset solution was included as a part of the students' final grade. We used the same answer submission template developed by the VAST Contest organizers, in which the students documented their analytic process and how their tools helped or hindered them in that process. To score their answer hypotheses, we partitioned the solution network provided by the TSG team into weighted components and used that to gauge the percentage of the solution that was found by the students. Points were not subtracted for other incorrect hypotheses. However, the majority of their grade was based on several deliverables throughout the semester that documented their progress in the requirements analysis, design, development, and evaluation phases, as well as the instructor's judgment of the overall technical quality of their final product based on course content.

We observed several benefits of using the Blue Iguanodon in this course:

- It forced the students to address the issue of scalability of data in visualization, and confront the design issues that arise when dealing with large datasets. In past offerings of the course where students could pick their own project topic, students tended to produce visualizations that could only display a small dataset, which is easy from a design perspective and not useful in practice. For example, in a previous semester we used a different dataset that contained only about 50 short intelligence snippets for the semester project, and this resulted in overly simplistic designs.
- It helped the students to realize the value of information visualization. Typically, there are several skeptical students in the course who prefer more automated approaches to data analysis. This project revealed that while computational approaches can identify many connections in the data, human-in-the-loop analysis is required to see the data, read the relevant documents, and understand the deeper subtle stories that make up the hidden ground truth (e.g. understanding the motivations of the scenario actors).
- The realistic nature of the dataset, combined with a hidden ground-truth solution, created an exciting and motivating challenge that the students enjoyed. The students felt connected to a larger agenda, and gained an appreciation for visual analytics and the domain of intelligence analysis. Since all of the student teams work on the same project topic, there was also a healthy competition and curiosity between the teams as they wanted to see what solution tools and answers the other teams discovered.

4.2 Undergraduate Course on Human-Computer Interaction

CS3724 Introduction to Human-Computer Interaction (HCI) is a practical undergraduate course, at the junior and senior level, focusing on the usability engineering process (<http://www.cs.vt.edu/undergraduate/courses/CS3724>). In the course, students learn about the scenario-based design approach to usability engineering [16], including methods for requirements analysis, design, and usability evaluation, as well as user interface implementation techniques and a module on information visualization design.

The project framework was similar to the graduate course. In this class we used the Stegosaurus dataset, which consists of approximately 240 documents, mostly news articles and a few maps and supporting documents, and a database of approximately 3000 extracted entities. For this course, the TSG team manually cleaned the entity database prior to the project assignment.

We chose this dataset for several reasons. It is significantly smaller in size than the Blue Iguanodon, and the entity data was carefully manually cleaned. The scenario also included a known critical event, which simplified the analytic process by providing a starting point. It could be solved in approximately 2-6 hours with standard tools. This was a good match for the course, which targets third year undergraduates, since it placed emphasis more on designing usable interfaces that support the overall analytic process, and less so on data visualization. Thus they could work towards solving the dataset with standard tools, and identify and address critical usability problems or areas for improvement along the way. Also, the dataset solution had not been publicly released, since it was used for the invitation-only live contest at IEEE VAST 2006, so students could not search online for the solution or results. Overall, the difficulty was about right.

A specific difficulty with this class was finding expert analysts for the students to observe during the early requirements analysis phase of the project. A solution that worked reasonably well was to have them observe the instructor's graduate students analyzing a similar dataset, and to observe each other as they analyzed the Stegosaurus data. We also supplemented this observation phase with other background materials about intelligence analysis, such as *The Psychology of Intelligence Analysis* [7] and *Intelligence Essentials for Everyone* [10]. However, while the background materials helped to set the mood, we found it more valuable for the purposes of the course to keep the students focused on specific issues associated with analyzing datasets like Stegosaurus, rather than the broad and general content of the background materials. In this sense, the dataset provided a specific concrete objective for the students' projects that was very helpful.

Eight of the 10 teams succeeded in solving the dataset ground-truth, and the other 2 teams found about 50% of the answer. Many of the teams developed tools that acted like a dynamic whiteboard to help them find connections, track their findings as node-link diagrams, and quickly link back to marked-up source documents (e.g. similar to Analyst's Notebook [8]). Various integrated search features were also found to be very helpful. A major differentiator was whether the node-link diagrams were manual (created by users) or automatic (parsed from the data), and students should be encouraged to consider how these seemingly opposite approaches might be combined.

The evaluation phase of their projects had two components: (1) a benchmark study of external users performing specific short tasks, and (2) a longitudinal, insight-based study of themselves as they analyzed the Stegosaurus dataset [17]. The second part was most enlightening for them as they witnessed first-hand how their tools helped or hindered their own analytic process. Searching for

the ground truth hidden in the data is what convinced them of the value of usability engineering. It helped them to understand that usability is not limited to learnability (making tools that are easy for novices to learn), but also includes expert performance (making tools that help experts solve hard problems over long periods of time).

At the end of the semester, 17 of the 40 enrolled students responded to our survey as follows. What percentage of the articles in the Stegosaurus dataset did they read to determine their hypothesis? Answers varied widely from 5% to 100% of the articles, but most said either 20% or 100%. Estimate the time to generate your hypothesis? Answers ranged from 1 to 10 hours, with an average of 4.1 hours. When asked about the most important lessons learned, most answered about the importance and difficulty of good usability engineering and teamwork. We believe the use of the Stegosaurus dataset played an important role in making usability engineering real to these students.

5 LESSONS LEARNED

At the beginning of the semester in each course, we used a separate smaller dataset as both an in-class exercise and follow-up homework assignment to introduce the students to the problem domain and spark discussion about potential visual analytics tools. The dataset was designed elsewhere for teaching intelligence analysis methods, and could be solved in a single class period. It contained approximately 50 short fictional intelligence snippets, including arrest reports, intercepted phone calls, and bank transactions. This exercise was very helpful and motivating. However, the smaller dataset was too different from the larger datasets. It focused on typical terrorist bomb plots and included more diverse data entities, such as phone numbers and bank accounts, whereas the Stegosaurus and Blue Iguanodon scenarios cover a broader spectrum of illicit activity such as drug smuggling and contain more homogenous text articles. Thus, while motivating, the exercise tended to cause some students to overly constrain their potential hypotheses in the larger datasets to bomb plots, and led to tool design ideas that were not helpful. So students should be cautioned or a more closely matched motivating exercise should be constructed in the future.

More directed background materials for the requirements analysis phase are needed. These materials should focus on how analysts analyze a dataset like this; perhaps a video of a complete scenario that students can examine in detail and derive requirements. Unfortunately this does not give students the opportunity to probe with questions. Since the number of students is much larger than the number of available domain experts, we plan to conduct live in-class observation exercises in the future.

At the end of the semester, a live competition on a second dataset (similar to the VAST live contest) would help the students to further convincingly evaluate their designs. We had decided against this because we thought that students would hardcode their tools to the assigned dataset and it would be too difficult to load a second dataset for live competition. However, from the survey of the undergraduates, we found that none of the responding student teams hard-coded their projects, and could have loaded another similar dataset with only a small amount of programming or database effort. Thus, we plan to add a live competition to the project in future offerings of these courses.

These datasets worked very well at both educational levels (undergraduate and graduate), and for different types of tool-building course topics (usability and visualization). Important characteristics of the datasets are the following:

- The ground-truth motivated students by the intellectual puzzle, enabled students to gauge their progress, supported evaluation and grading, and clearly demonstrated the value of good methods. For the students, these assignments were more intellectually satisfying than other assignments that ask students to simply look for something interesting in a dataset [9].
- The datasets are realistic and the topic domain is timely and culturally relevant. Students are curious about intelligence analysis in general.
- Getting the right size dataset is critical. If it is too small, students will generate toy solutions without understanding design challenges. If it is too large, students will not have enough time during the semester to complete. These datasets were appropriate.
- Providing a starting point for analysis in the dataset scenario helps students make initial progress on the project sooner, and enables more steady progress throughout (rather than a sudden serendipitous eureka), which better matches educational goals and methods.
- Clean data was important, as some students in the graduate course spent far too much time on data cleaning and processing, and then gave up on designing user interface features that would have been valuable.
- Solutions must not be publicly available (students will search for them), but available to instructors. This indicates the need for additional datasets for educational purposes that are not used in the open VAST Contests. We were able to maintain secrecy of the solutions for the above courses in two ways. For the graduate course, the VAST conference was held during the Fall semester in which the course occurred, and the VAST Challenge organizers agreed to withhold publication of the Challenge results on the web until the end of the semester. Clearly this approach does not scale to future courses. For the undergraduate course, we used a dataset from the invitation-only live contest at the VAST conference for which the data and solutions were never published publicly. This approach can scale to the long term, assuming that instructors and students do not post solutions on the web. The existence of a large pool of such datasets would mitigate the impact of students handing down “spoilers” to future students.
- There must be multiple similar datasets for use in in-class exercises, observations, and a final live competition. Because of the nature of the textual data type and task of these dataset scenarios, it is possible to create endless variations for repeated evaluations. The number of possible ground truths is essentially limitless. Answers cannot be found with an automated process or query, and finding the answer in one does not help with finding the answer in another. Yet, because the datasets are equivalent in type, it is easy for students to construct solution tools that can load additional datasets and help solve them, for further evaluation purposes. These characteristics are much harder to achieve when hiding ground-truth patterns in fictitious quantitative datasets.

The datasets were mostly homogeneous, consisting primarily of text news articles. Hence, useful quantitative visual mappings had to be generated from derived data, such as search hits and keyword counts, so there was less use of those types of visual encodings. Entity databases were useful for generating network oriented solutions. The more successful solution tools were those that integrated keyword or entity search features into the visual representation, focused on enabling users to build up their hypothesis, and focused on one primary representation, usually network oriented.

Several student teams were tempted to create comprehensive solutions that offered multiple views for all possible perspectives (network, time, geography, keywords, etc.), but quickly became overwhelmed and produced poor usability. Representations of time and geography were not very useful because they are not critical to the answer, the data was difficult to extract, and the date and location of the articles were not necessarily relevant to the events described in them. Students should be cautioned against overly complex solutions. Another alternative is to attempt a larger collaborative effort by the class as a whole to produce an integrated solution, with individual teams working on portions.

With this project assignment, one might expect a fairly homogenous set of solution tools created by the students. However, there was a surprising amount of diversity, indicating that there is much room for creativity with these datasets. Also, among the solutions that were similar (e.g. network oriented solutions), students could recognize how even minor design differences had significant impacts.

In evaluation, it appeared that the quality of the students' solution tool did not necessarily correlate with how much of the ground-truth answer they discovered. Finding the answer was a helpful measure, but certainly not the only measure. Given the educational setting, importance was placed on the usage of methods and techniques learned in the course, and meeting milestones during the semester. There are a few reasons for this effect:

- With Blue Iguanodon, there is no clear starting point in the scenario, so students must start by searching for anything 'interesting'. In some cases, students seemed to serendipitously stumble upon the key component of the answer, which then easily led to much of the rest of the answer. Thus, students either got most of the answer or none at all. Whereas, with Stegosaurus, the presence of starting evidence enabled students to make some initial progress quickly and then had to make several additional connections to fill in the entire story, so there was more steady progress and meaningful distribution of percent completion.
- Some students pursued directions in their design (e.g. geographic views) that turned out to be unhelpful for finding the ground-truth in this particular scenario. In some cases, their designs were solid and might be useful in other scenarios, but just not in this particular scenario. Some pursued high risk approaches that did not succeed well. More emphasis on requirements analysis phase could mitigate this, but clearly grading should not be based solely on finding the ground-truth.
- Some students are simply better analysts than others. Using themselves as test subjects is not ideal, although working in groups helps to mitigate the problem. Adding the urgency of a live competition at the end of the semester might also help to equalize abilities. But without ready access to a large

number of trained analysts, fair comparative evaluation remains an open problem.

6 OTHER IMPACTS

Applying these datasets in the courses also served to test and debug the datasets. Some students found formatting errors in the data that could have revealed the solution. Students also pointed out that the use of the fake country name Parazuela, while other names were real (e.g. Argentina), gave away that it was part of the solution. Students also pointed out that, while most of the ground truth was hidden in the many news articles, the additional materials in the datasets such as blogs or images sometimes made too obvious hints.

From an instructor's point of view, the use of these datasets in coursework not only served valuable pedagogical purposes, but also made organizing the course easier for the instructor. The datasets are readily available, easily fit into a semester project framework without the need to develop new materials for a new domain, and contain ground-truth solutions that help in the grading process. The course website linked directly to the VAST dataset materials. In comparison to previous semesters in which students could choose their own project topics, this approach saved a significant amount of the instructor's time that was previously spent preparing a set of potential project topics, guiding student teams into projects that would be appropriate for the course goals, and tracking the evolution of their topics throughout the semester to maintain appropriateness. It also saved a significant amount of course time (students' and instructor's time) at the beginning of the semester that was previously spent settling on project topics. Thus, students were able to make more progress on their projects earlier in the semester with the VAST datasets, resulting in more complete implementations and more meaningful evaluations by the end of the semester. The ground-truth solutions and score sheet templates also helped the instructor assess projects and provide meaningful feedback to the student teams. This form of evaluation helped to judge how the designed tools supported the process of analytics. Instructors should familiarize themselves with the dataset and solution at the beginning of the semester. A drawback was that graduate students did not have the freedom to select projects that would directly relate to their own thesis research (which would serve faculty in the department), but this was not common in previous semesters anyway. Overall, the VAST datasets are an excellent educational resource that instructors can effectively apply in their courses with very little effort. We found it to be successful, and plan to continue using the datasets in future course offerings.

The TSG team is aware that there is increasing interest in the use of datasets by educators and that challenge solutions for classroom use must not be publicly released if educators want to factor in solution accuracy as part of the students' assessment. Initially, our approach was to release solutions to the regular VAST challenges, but keep the solutions to the interactive VAST sessions (held at the symposium with limited participation) unpublished but available to instructors. In the 2008 Challenge, we limited the release of the solution to only the teams who participated in the challenge. Requests for the solution from other groups are reviewed on a case-by-case basis. While this does not ensure that the solution is not available, it does help limit distribution. Another possible solution is to create modified versions of the challenge scenario and datasets strictly for educational uses. A grander solution that will require additional research is to construct usable tools that enable instructors to modify or implant new ground truths within template datasets.

7 DATASETS AT THE NVAC SUMMER CAMP

In 2008, NVAC hosted the inaugural Visual Analytics Summer Camp [1]. This camp was a two-week educational immersive session on a broad range of topics in Visual Analytics. Just as dynamic as the field, the group of participants were diverse in that they came from government, industry, and academia, with background experience in analytics, research, development, and entrepreneurship. Attendees came from the National Air and Space Intelligence Center, Mercyhurst College, Simon Fraser University, Middlesex University, Virginia Tech, and other institutions. The curriculum was designed to provide a sampling of activities performed by professional analysts. These included watch and warn training, analytical writing, evaluation, and analytical tasks to perform using visual analytical tools. The students also provided input on their desired outcomes. Some of these included:

- To get introduced to the state-of-the-art in analyst requirements and tools to travel “the last 12 inches” from information to actionable knowledge.
- Understand more deeply the problem areas for VA as well as the breadth of its applications. Also, to network and gain experience with new tools in order to find areas for improvement and to focus my own research area.
- To understand what exactly visual analytics means and to have as a skill set to take with me to a position of intelligence analyst.

Throughout the two-week agenda, the participants were given the opportunity to get hands-on experience with a collection of analytic tools, the ability to network with research leaders in the field, and valuable discussion between analysts and developers (Figure 3).



Figure 3: IN-SPIRE expert assists a student at NVAC Summer Camp.

The tools used during the camp included Analyst’s Notebook, Starlight, IN-SPIRE, and Jigsaw [5]. In the evaluation session of the workshop, the students were assigned small groups to work through a VAST challenge with the White Smilodon dataset (used in the VAST 2008 interactive session) using Jigsaw, and two VAST contest committee members were teaching. While one group worked through the problem, another observed their activity with respect to the usability and utility of the tool. The session was held near the end of the workshop, so that students had some familiarity with both the tool and analysis processes.

To the credit of both the student group and to Jigsaw, the students were able to make considerable progress on the White Smilodon tasks in the two hour allocation. Since the students already knew each other fairly well from the week and a half they had spent together, they did not need the bonding period we usually see needed for the VAST interactive session, where analysts and tool builders are brought together for the first time to analyze a dataset. We observed groups making and debating hypotheses, playing “what-if” games, and experimenting with Jigsaw to dive deeper into the data, as well as to explore various functionality of interest. We also observed the evaluators asking questions about the software functionality by itself (e.g., “Does that feature do what you want it to do?”) and as related to the analytical reasoning process (“How did you reach that conclusion?”). Some of the evaluators expressed reluctance at times to interrupt the analysis teams’ thought processes, but found approaches to do their jobs, such as asking questions relevant to the specific interactions going on and finding people good at multi-tasking to ask (those who could both analyze and discuss the process). The student analysts were generally able to do their tasks and provide comments to the evaluators about functionality they would like to have had. We were prepared to give out a considerable hint as to how to start to analyze the data; however, we did not need to do this as each team found a productive entry way into the data.

When the session was over, we provided feedback to the students about the analysis, their solutions, and on the evaluation points they discovered. We were able to contrast their solutions with the groups from the VAST contest who had also analyzed the dataset. In return, the students provided valuable feedback on the exercise, the dataset, and the use of Jigsaw from their perspective as novice analysts.

Even though the Summer Camp was an intense, compressed educational experience, surveys at the end of the session included comments stating that the participants found the hands-on experience with the tools and the realistic exercises gave them a greater appreciation for visual analysis.

8 DEVELOPING EVALUATION METHODOLOGIES AND METRICS FOR EDUCATION

Conversations with instructors of visual analytics have convinced us that providing them with help in evaluating student projects is a worthwhile endeavor. There are three parts to the VAST contest evaluations: accuracy, interactive visualizations, and process. As described above, students were asked to provide these three components for evaluation. Assessing the accuracy is relatively easy and we provide the solutions to instructors upon request. It should be noted that in some cases, participants have found other possible solutions that we did not know were in the data. If they provide adequate evidence for their solution, it is counted as correct. In the future, this assessment may become easier as we are implementing an infrastructure as part of a National Science Foundation grant [15] that will provide accuracy feedback without providing the solution. Instructors will decide whether they want this feature to be available to their students.

Assessing the visualizations is more of a challenge [14]. There are guidelines for visualizations and for human-computer interactions that can be applied using a heuristic evaluation technique. A separate part of the class projects might be a heuristic evaluation of the teams’ projects. Assessing the utility of the visualizations in the context of conducting the analysis is not well understood [19]. For the actual VAST contests, we collect qualitative feedback from both professional analysts and visualization experts. Instructors will probably not be able to

obtain these resources for class projects. There are several possibilities that could be used for classroom projects. The graduate projects, described above, submitted descriptions of the process they used in their analysis including the information they found in each visualization. The process descriptions can be evaluated for efficiency: how many steps were used to arrive at the answer and how much information is gleaned from each visualization. Another possibility is to have the students first research the state of the art visualizations for the specific combination of data type and analysis to be performed. Students could then identify the limitations of that particular data type and design their visualizations to overcome these limitations.

The most challenging assessment is that of the utility of the visual analysis tool in the context of analysis. One problem described above is that early classroom exercises do not match well with what is required in the project. One solution would be to give the students the datasets previously used for the classroom project for a longer exercise. They could be required to use standard search tools and office tools to analyze this data. This would give them a baseline to use in identifying problems that their design would, hopefully, overcome. The developed systems could then be used to analyze the larger VAST contest datasets. An option would be to include a “normative process” with the solution to any given contest data. After students had finished their analysis, they could compare their process to the normative process and submit this as part of a “self-evaluation.”

Helping instructors to provide better feedback for student projects should be focused on helping both the instructors and students understand the current analytic processes and current problems faced by the analysts. While providing the datasets and scenarios is a necessary piece of this, we need to also provide an understanding of the processes used by analysts.

9 CONCLUSION

The experiences at Virginia Tech and at the Visual Analytics Summer Camp illustrate the usefulness of the VAST contest challenge tasks and datasets in educational settings. Students have benefitted from practical exercises, learned about visual analysis, and were able to apply their new knowledge in the development of analytical software in their projects. From the Virginia Tech experience, we learned that there are various difficulties in having students pick up the task descriptions and datasets alone to attack the tasks. It may help to package the VAST contest with educational materials such as descriptive literature and videos, to support future classroom work. Finally, we have an on-going research need for metrics development for education. Activities like the VAST contest and work through NSF as in the SEMVAST project, will contribute to this in the future.

ACKNOWLEDGMENTS

This work was supported in part by the National Visualization and Analytics CenterTM (NVACTM) located at the Pacific Northwest National Laboratory in Richland, WA. The Pacific Northwest National Laboratory is managed for the U.S. Department of Energy by Battelle Memorial Institute under Contract DE-AC05-76RL01830.

REFERENCES

[1] C. Almquist. Visual Analytics Summer Camp. Downloaded 3/26/2009. http://nvac.pnl.gov/vacviews/VACViews_nov08.pdf.
 [2] J. Bransford. Anchored Instruction. In *The Adventures of Jasper Woodbury*. Retrieved 7/9/09 from

<http://peabody.vanderbilt.edu/projects/funded/jasper/intro/Jasperintro.html>
 [3] S. Card, J. Mackinlay, and B. Shneiderman. *Readings In Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers, January 1999.
 [4] Cognition and Technology Group at Vanderbilt. Anchored instruction and its relationship to situated cognition. *Educational Researcher*, 19(6): 2-10, 1990.
 [5] C. Goerg, L. Zhicheng, P. Neel, S. Kanupriya, and J. Stasko. Jigsaw meets Blue Iguanodon - The VAST 2007 Contest, *IEEE Symposium on Visual Analytics Science and Technology 2007*. October 30 – Nov. 1. Sacramento, CA, USA. 2007.
 [6] G. Grinstein, C. Plaisant, S. Laskowski, T. O’Connell, J. Scholtz, and M. Whiting. VAST 2007 Contest-Blue Iguanodon, *IEEE Symposium on Visual Analytics Science and Technology 2007*. October 30 – Nov. 1. Sacramento, CA, USA. 2007.
 [7] R. Heuer. *Psychology of Intelligence Analysis*. Center for the Study of Intelligence, Central Intelligence Agency, 2001.
 [8] I2. Analyst’s Notebook. Downloaded 3/26/2009. http://www.i2.co.uk/products/analysts_notebook/.
 [9] A. Kerren, J. Stasko, and J. Dykes. Teaching Information Visualization. In A. Kerren, J. Stasko, J.-D. Fekete, C. North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, pp. 65-91, 2008.
 [10] L. Krizan. *Intelligence Essentials for Everyone*. Joint Military Intelligence College, Occasional Paper Number Six, Washington, DC, 1999.
 [11] Learning Technologies at Virginia Tech. 2009. Teaching Models, <http://www.edtech.vt.edu/edtech/id/models/index.html>.
 [12] S. Lovgren. Wildlife Smuggling Boom Plaguing L.A., Authorities Say. *National Geographic News*. Downloaded 3/25/2009. <http://news.nationalgeographic.com/news/2007/07/070725-animal-smuggle.html>.
 [13] C. North. Information Visualization. *Handbook of Human Factors and Ergonomics*, 3rd Edition, G. Salvendy (editor), New York: John Wiley & Sons, pg. 1222-1246, (2005).
 [14] C. Plaisant. 2004. The challenge of information visualization evaluation. In *Proc. of the Working Conference on Advanced Visual Interfaces (2004)*. AVI '04. ACM, New York, NY, 109-116.
 [15] C. Plaisant. Scientific Evaluation Methods for Visual Analytics Science and Technology (SEMVAST). Downloaded 3/26/2009. <http://www.cs.umd.edu/hcil/semvast/>
 [16] M. Rosson and J. Carroll. *Usability Engineering: Scenario-Based Development of Human-Computer Interaction*, Morgan Kaufmann, 2002.
 [17] P. Saraiya, C. North, V. Lam, and K. Duca. An Insight-based Longitudinal Study of Visual Analytics. *IEEE Trans. on Visualization and Computer Graphics*, 12(6): 1511-1522, Nov. 2006.
 [18] P. Smith and A. Pellegrini. *Psychology of Education: Major Themes*. Taylor & Francis, London, 2000.
 [19] J. Scholtz. Beyond Usability: Evaluation Aspects of Visual Analytic Environments. *IEEE Symposium on Visual Analytics Science and Technology 2006*. October 31 – Nov. 2. Baltimore, MD, USA. 2007.
 [20] R. Spence. *Information Visualization: Design for Interaction*, 2nd Edition, Addison-Wesley, 2007.
 [21] A. Tucker. Strategic directions in computer science education. *ACM Comput. Surv.* 28, 4 (Dec. 1996), 836-845.
 [22] R. Van Eck. Digital Game-Based Learning: It’s Not Just the Digital Natives Who are Restless. *EDUCAUSE Review*, 41(2), 16-30, 2006.
 [23] N. Vye. The Effects of Anchored Instruction for Teaching Social Studies: Enhancing Comprehension of Setting Information. *Annual Meeting of the American Education Research Assoc*, Boston, 1990.
 [24] M. Whiting, J. Haack, and C. Varley. Creating realistic, scenario-based synthetic data for test and evaluation of information analytics software. In *Proc BELIV '08*. ACM, New York, NY, 1-9, 2008.