

Analysis Without Data: Teaching Students to Tackle the VAST Challenge

Edward W He, Daniel Tolessa, Ashley Suh, Remco Chang

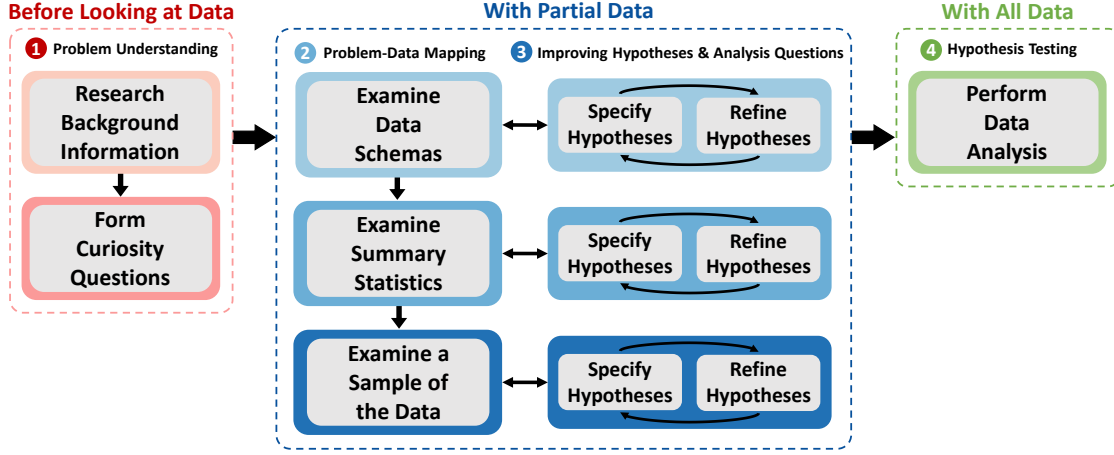


Fig. 1. Our four-stage proposed guideline, HDAV (**H**ypothesis-**D**riven **A**pproach for **V**AST), for students tackling VAST Challenges. HDAV follows the practice of pre-registering hypotheses before visual analysis and experimental design [4, 8], encouraging students to critically consider the *analytic questions* prior to analyzing the full dataset. Only after a student has fully articulated their hypotheses should they move to performing data analysis. The final arrow denotes that once the data analysis process begins, students should not return to forming hypotheses. Doing so could result in HARKing (Hypothesizing After Results are Known) [11] and false discoveries [32].

Abstract— The VAST Challenges have been shown to be an effective tool in visual analytics education, encouraging student learning while enforcing good visualization design and development practices. However, research has observed that students often struggle at identifying a good “starting point” when tackling the VAST Challenge. Consequently, students who could not identify a good starting point failed at finding the correct solution to the challenge. In this paper, we propose a preliminary guideline for helping students approach the VAST Challenge and identify initial analysis directions. We recruited two students to analyze the VAST 2017 Challenge using a hypothesis-driven approach, where they were required to pre-register their hypotheses prior to inspecting and analyzing the full dataset. From their experience, we developed a prescriptive guideline for other students to tackle VAST Challenges. In a preliminary study, we found that the students were able to use the guideline to generate well-formed hypotheses that could lead them towards solving the challenge. Additionally, the students reported that with the guideline, they felt like they had concrete steps that they could follow, thereby alleviating the burden of identifying a good starting point in their analysis process.

1 INTRODUCTION

The VAST Challenges¹ emulate real-world visual analytic tasks and problems, making them an effective tool in visual analytics education. In classroom settings, researchers and educators have found that the VAST Challenge enables situated learning, which emphasizes learning around realistic case studies and promotes exploration by the students, resulting in active learning of complex concepts [21, 31]. However, the same research has also found that the students’ performance can vary because of a lack of clear starting points. This results in the students exploring the data seemingly randomly and stumbling upon the right (and sometimes wrong) answers:

... There is no clear starting point in the [Blue Iguanodon] 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. (Whiting et al. [31])

Searching unmethodically for anything “interesting” can be costly and threaten the validity of a student’s analysis. While it is possible to discover interesting patterns in the data can represent the ground-truth, they can also be spurious and meaningless. In a study with university students, Zraggen et al. [32] found that over 60% of the students’ findings in open-ended data exploration with a visual analysis tool consist of false discoveries (Type I error where the found insights are wrong) or false omissions (Type II error where the correct answers are omitted). Without guidance to identify better “starting points,” students may fall into these common pitfalls when solving VAST Challenges.

In this paper, we present a preliminary guideline for helping students in a visual analytics classroom to overcome the initial steps of tackling a VAST Challenge. Motivated by the above work [31, 32], our key insight in forming this guideline is to discourage students from immediately trying to explore the data. Instead, we want to encourage students to spend time on the *analytics* aspect of the challenge – focusing on reasoning through the problem and tasks before digging into the data.

This approach of methodologically performing data analysis is inspired by recent work in pre-registering hypotheses before visual analysis [12, 26] and experimental designs [4]. In experimental studies, pre-registration can prevent p-hacking [7], HARKing (Hypothesizing After Results are Known) [11], and postdiction (as opposed to prediction) [18]. In exploratory data analysis, researchers have found

• Edward W He, Ashley Suh, and Remco Chang are with Tufts University. E-mails: {edward.he, ashley.suh, remco.chang}@tufts.edu.
• Daniel Tolessa is with Virginia Tech. E-mail: danielmt@vt.edu.

¹<https://vast-challenge.github.io/2022/about.html>

that asking analysts to describe their hypotheses prior to performing data analysis encourages the analysts to investigate the data in more detail [3], which can in turn result in deliberate analysis outcomes [14].

Based on these collective findings, we propose a Hypothesis-Driven Approach for VAST Challenges (HDAV). HDAV is designed to guide students through the initial understanding and reasoning processes of a VAST challenge, in order to thoughtfully form and refine hypotheses before conducting analysis. As evidenced by Whiting et al. [31] and Zraggen et al. [32], such an approach can prevent students from serendipitously finding the ‘correct’ answer, jumping to spurious conclusions, or pigeonholing themselves into their first ‘interesting’ finding. HDAV consists of four stages that reflect a **student’s incremental access to the data for data analysis**. These four stages include: (1) Problem Understanding, (2) Problem-Data Mapping, (3) Improving Hypotheses and Analysis Questions, and finally (4) Data Analysis.

We follow an iterative approach in developing the guideline. In the first phase, two undergraduate students (who are co-authors of this paper) investigated the 2017 VAST Challenge Mini-Challenge 1 [30] following the pre-registration practice of articulating their analysis questions and hypotheses while minimizing their exposure to the raw data. Similar to the process of *coding* in qualitative studies [16], the two students examined the VAST Challenge independently and met over multiple sessions to achieve internal agreement between their listed analysis questions, as well as their overall analysis approach. In the second phase, the research team (including all co-authors of this paper) reflected on the students’ process of formulating and refining analysis questions, resulting in our proposed guideline. Finally, the same two undergraduate students applied the guideline to a new VAST Challenge (in this case, the 2016 VAST Challenge Mini-Challenge 2 [6]). Role-playing as potential students in a visual analytics course, the two students documented their experience and reported the benefits and the shortcomings of using the guideline.

We note that the guideline reported in this paper is preliminary. Future controlled experiments with more students in a visual analytics course is necessary to validate the efficacy of our guideline. Therefore, we present our guideline as a series of concrete steps that can help students and educators reason about the *analytical* questions for a VAST Challenge independently from the data analysis (shown in Figure 1). Our findings from using the guideline in the 2016 VAST Challenge suggests that students can benefit from a more methodological approach to starting analysis than “searching for anything interesting” [31], and motivates future work for visual analytics education.

2 RELATED WORK

Our proposed guideline is based on past literature on the use of the VAST Challenge and the authors’ experience with a university-level visual analytics course. The guideline differs from, but is consistent with, the existing models in visual analytics. In this section we describe past work in these two areas and how our work relates to them.

2.1 VAST Challenge in Education

The IEEE Visual Analytics Science and Technology (VAST) Challenge was first introduced in 2006 as a contest to advance the science of visual analytics. The VAST Challenges have evolved to serve as a benchmark for the visual analytics community to experiment with new interactive systems, algorithms, and evaluation techniques. The Challenges are typically modeled after real-world scenarios and each challenge comes with a “ground-truth” solution [5]. Each challenge typically consists of three independent mini-challenges (MCs) that reflect different problem domains (e.g., temporal analysis, image analysis, text analysis, geospatial analysis, etc.). Solving all three mini-challenges will reveal the higher-level “story” of the overall challenge [5, 24].

As a platform for evaluating visual analytics environments [23], the VAST Challenge holds many benefits when used in an education context. Whiting et al. [31] reported an early experiment at Virginia Tech where the VAST Challenge was used in two Computer Science courses (for graduate and undergraduate students respectively). Using the 2007 VAST Challenge, the authors found it encouraged active

learning and fostered critical thinking, as the concept of a “ground-truth” to be discovered by the students was an intriguing intellectual puzzle. Additionally, the realistic nature of the challenge motivated the students to apply what they had previously learned in class [31]. These findings were confirmed by another study in 2013 at the University of Konstanz where the researchers used the Bloom’s taxonomy of Educational Objectives and demonstrated how the VAST Challenge enabled the integration of different learning objectives [21].

However, while both studies reported positive outcomes from using the VAST Challenge for education, they also noted potential difficulties and pitfalls. From the perspective of the instructor, the use of the VAST Challenge increased supervision effort and raised questions about how to design a consistent grading rubric. From the perspective of the students, both studies reported the importance and the high cost of the students taking a “wrong direction,” especially during the initial phase of the analysis. The goal of our paper is to provide a preliminary guideline that can help students avoid such pitfalls in the initial phase and increase their chances of successfully solving VAST Challenges.

2.2 Visual Analytics Models and Guidelines

There have been a number of models (or workflows) proposed for the visual analytics community. These models have defined the process of visual analytics and identified opportunities to improve current practices. Some of the more notable models include the visual analytics “diamond” model by Keim et al. [10], Card and Pirolli’s Sensemaking model [19], van Wijk’s value of visualization model [29], Sacha et al.’s knowledge-generation model [22], and the more recent models of visual analytics as (machine learning) model building and exploration [1, 2].

However, while these models have been invaluable to the visual analytics community, generally they have been designed for researchers and are “descriptive” in nature. For students who are not well-versed in visual analytics literature, these descriptive models are insufficient in helping them identify a good “starting point” when addressing a complex visual analytics problem like the VAST Challenge.

To develop a “prescriptive” guideline that is more suitable for students, we draw inspiration from the existing visual analytics models and the Design Study Methodology paper by Sedlmair et al. [25]. The goal of the guideline is to have a linear process that a student can follow where each step of the process is associated with concrete activities and examples. Figure 1 illustrates the workflow of our proposed guideline. In the sections below, we describe the process for generating the guideline and the explanations for each of the steps.

3 METHODOLOGY FOR TACKLING THE VAST CHALLENGE

Our goal is to develop a guideline that students in a university-level visual analytics course can follow when working through a VAST Challenge. To do so, we recruited two student participants to complete a VAST Challenge, with specific constraints and criteria, and rigorously document their process. The resultant guideline outlines a step-by-step workflow that includes: (1) concrete instructions for students to follow, (2) specific examples from the 2017 VAST Challenge to provide clarity, and (3) additional suggestions for steps that students may have trouble completing. We describe the process for developing the guideline in terms of the material, the participants, the task, and the procedures.

3.1 Material

The problem used for this task was the 2017 VAST Challenge² Mini-Challenge 1 (MC1). This challenge is based on a fictitious nature preserve where the local Red Crested Blue Pipet (bird) resides, and a (fictitious) problem in the preserve that must be solved: Recently, it has been discovered that the nesting pairs of the Blue Pipets have been decreasing. The local Ornithology Preservation Society is concerned about the decrease in these nesting pairs and would like to identify possible reasons for this. The goal of MC1 is to identify if there is a link between the traffic going through the park and the decrease in nesting pairs. The questions provided by the challenge are:

Q1 Describe daily patterns of vehicles traveling in the park.

²<http://vacommunity.org/VAST+Challenge+2017>

Level	Criterion
0	No explanation provided. A nonsense statement, a question, an observation, or a single inference about a single event, person, or object.
1	An inappropriate explanation. For example, "...because it's magic" or "...because the man pushed a button."
2	Partial appropriate explanation. For example, incomplete reference to variables, a negative explanation, or analogy.
3	Appropriate explanation relating at least two variables in general or nonspecific terms.
4	Precise explanation with qualification of the variables. A specific relationship between the variables is provided. This is a hypothesis.

Table 1. Sunal and Haas' *Hypothesis Quality Scale* [27] for writing and evaluating students' hypotheses. Level 4 and beyond is considered a valid hypothesis.

Q2 Describe patterns of vehicles traveling through the park over a longer period of time (multiple days).

Q3 Identify unusual patterns of activity.

MC1 data includes: (1) a map of the nature reserve (as an image), and (2) a tabular dataset consisting of sensor-log information for vehicles visiting and driving through the preserve. Each time a car passes through one of the sensor locations, the time-stamp, sensor-ID, car-ID, and car-type are recorded as a row in the data. In addition to the data, MC1 includes a data description and a list of questions that a contestant is required to complete (referred to as the "answer sheet").

3.2 Participants

Two undergraduate students majoring in Computer Science (who are co-authors of this paper) served as the "participants" in tackling the 2017 VAST Challenge. The students were recruited as part of a summer undergraduate research appointment. The two students had not previously taken a visual analytics course and had little experience with visual analytics systems (e.g. Tableau). As such, we consider these participants "typical students" with the average criteria of those who might be taking a visual analytics course at the university level.

3.3 Tasks

The two participants were tasked with approaching the VAST Challenge with the following constraints:

Generate Analytic Questions and Hypotheses: Following the concept of pre-registration [4], the two participants were instructed to generate analysis questions or hypotheses related to the challenge before performing data analysis. Questions and hypotheses were specified to be related to Q1-Q3 from the challenge. For example, a question could be on what could qualify as an "unusual pattern of activity" (MC1 Q3).

Limit the Use of Raw Data: To mitigate potentially spurious findings [32], the participants were asked to generate as many analytic questions as possible before looking at the available data. After exhausting all relevant questions, they were then permitted to look at parts of the raw data. For example, the participants could look at the schema of the data or a few rows of the data. Each time the participant looked at some aspect of the data, they were instructed to generate additional questions and hypotheses, but *not* attempt to validate them. This process was repeated until participants reported they had done all they could with hypothesis generation without performing data exploration or analysis.

Turning Open-ended Questions into Formal Hypotheses: To ensure that the participants were forming meaningful analytic questions and hypotheses that could eventually be tested to solve the challenge, the participants were asked to refine any open-ended questions into formalized hypotheses using either the hypothesis grammar by Suh et al. [26] or the Hypothesis Quality Scale by Sunal and Haas [27] (see Table 1). An example of this process is discussed in Section 4.3.

Careful Documentation: Each participant was asked to keep a detailed research journal of: (1) the questions or hypotheses that they generated, (2) parts of the data they examined, (3) tools that they used to examine the data (e.g., Excel, Tableau), (4) how the data was examined (e.g., screenshots of tables or visualizations), and (5) additional notes and observations about their process.

Iterative and Collaborative Refinement: The two participants were asked to complete all components of the task independently. After both had completed a round of forming analytic questions for the challenge, they were asked to meet with the other person and compare their research journals. After each meeting, the participants were asked to return to the task and refine their questions and hypotheses based on the previous discussion. This iterative process continued until the two participants agreed that their two lists of questions and hypotheses were sufficiently refined, exhaustive, and consistent with each other's.

3.4 Procedure

The two participants (henceforth referred to as PID1 and PID2) completed the task over the span of 2 weeks. The participants initially took different approaches. After reading the MC1 challenge (and without looking at the data), PID1 began the task by performing background research into general bird nesting behavior (e.g., by reading online articles) while PID2 immediately started posing questions and hypotheses.

It took the participants three iterations before they agreed that their two lists of questions and hypotheses were consistent with each other's. For example, during the discussion after their first iteration, the participants noted that PID1's questions were strictly focused on unusual patterns of activity (MC1 Q3), and did not generalize to the normal behavior of vehicle activity. On the other hand, PID2's questions addressed each component of MC1, but were much broader than PID1's questions. Therefore, in the next iteration of the task, PID1 spent time generating questions for the remainder of MC1 while PID2 spent time refining his questions to better categorize unusual activity.

The two varied methods in which the participants posed questions is consistent with prior literature where researchers documented the different ways that students approach the VAST Challenge [31]. Over the course of two additional iterations, the participants were able to come to an agreement that their lists of questions and hypotheses were the same (or sufficiently similar) in both coverage and specificity.

The participants consolidated their research journals to construct the initial draft of the guideline. This guideline underwent further editing and refinement by the research team (the co-authors of this paper). In the section below we describe our proposed guideline for students to approach a VAST Challenge, using the iterative process conducted and fine-tuned by the task participants.

4 THE GUIDELINE

The guideline aims to provide a structured step-by-step process to guide students towards solving a VAST Challenge. With the focus on encouraging the students to critically analyze the problem before performing data exploration and analysis, this guideline consists of 3 parts: (1) problem understanding, (2) problem-data mapping, and (3) hypothesis formation. Figure 1 illustrates the workflow of the guideline.

4.1 Before Looking at the Data: Problem Understanding

The first stage of the guideline is to carefully review all materials of the VAST Challenge and brainstorm about the possible hypotheses that can be formed, without looking at or analyzing the data. VAST Challenges come with a wealth of contextual information in addition to the raw data: A typical challenge includes a background story about the scenario, an "answer sheet" that contains the questions that a contestant needs to complete, and (sometimes) data dictionaries that explain the dataset. Even before looking at the data, a student can begin to critically reason about the challenge and form questions and hypotheses that will guide the eventual solution.

4.1.1 Background Research

In order to have a good understanding of the problem, students should perform some background research to familiarize themselves with the story in the challenge. For example, understanding the roles of a park ranger in a US nature reserve will help a student better assess the behavioral patterns of the ranger vehicles in the 2017 Challenge.

Most critically, the students need to carefully review the questions in the "answer sheet." Often the answer sheet provides hints to the possible solution. For example, Q1 of the 2017 Challenge MC1 states:

Characterize the patterns by describing the kinds of vehicles participating, their spatial activities (where do they go?), their temporal activities (when does the pattern happen?), and provide a hypothesis of what the pattern represents (for example, if I drove to a coffee house every morning, but did not stay for long, you might hypothesize I'm getting coffee "to-go")

This question indicates that finding the solution requires a geospatial-temporal analysis grouped by vehicle types. Without looking at the data, the student can reasonably assume what the data will contain (e.g., timestamps for different vehicles across regions of the park) and the types of visualizations (e.g., time series charts or choropleths) that would be helpful for solving the challenge.

4.1.2 Curiosity Questions

After reviewing the background material, the students should write down any initial questions inspired by the challenge. We refer to these as "curiosity questions." Curiosity questions are questions that students can pose about a scenario without any restriction or concern on what the answer *ought* to be. This step is beneficial as it aids with brainstorming and forming ideas without worrying about the quality or the validity of those questions. In a collaborative classroom setting, it would also encourage team members to adjust to each other's prior beliefs about the scenario. Example curiosity questions generated by the two participants include:

- What time do cars usually enter the park?
- How many less nesting pairs are there?
- Is there an increase in noise disturbance?
- Where in the park are birds most active?

4.2 Peeking at the Data: Problem-Data Mapping

Once the students have understood the scenario of the challenge and posed their curiosity questions, the next step is to map the questions to the data. Similar to the concept of pre-registration, the students will not have full access to the raw data when generating analysis questions and initial hypotheses. Instead, the students will have step-by-step incremental access to partial data (we refer to this as "peeking" at the data), in which they examine: (Step 1) the data schema, (Step 2) summary statistics, and (Step 3) a sample of the data.

4.2.1 Step 1: Examining the Data Schemas

During the initial step, students are allowed to look at the data schemas. Data schemas provide essential information about the dataset, such as the number of columns, their names, and their data types. Examining the schemas will help the students recognize which of their curiosity questions can actually be answered by the data that is provided.

For example, in Section 3's exercise, the participants were tasked with discovering the cause of the decrease in Pipets' nesting while completing the 2017 VAST MC1. Unsurprisingly, most of the curiosity questions posed by the participants related to causal reasoning about the nesting decrease (examples provided in Section 4.1.2). However, upon examining the data schema, the participants recognized that no nesting information is provided in the data – the data only contains sensor information about traffic in the nature reserve (see Section 3.1). At this point, the participants refined their initial questions and switched from asking causal questions to asking questions about detecting anomalies, or abnormal patterns, in the data.

4.2.2 Step 2: Examining the Summary Statistics

Summary statistics are used to provide a holistic view of the data as simply as possible. In Section 3's exercise, we constrained summary statistics to include central tendency (mean, mode, median), spread (standard deviation, range, percentile), shape (distribution), in addition to the characteristics and structure of the data: cardinality, number of rows, and completeness (if there are missing values).

In the second step of peeking at the data, the students should review the next draft of their questions with the summary statistics in mind. The students might realize that the size of the data is very large and scalability could become a concern, or that there are missing values in

the data that need to be addressed. Most importantly, students should recognize whether their questions might not be answerable without performing some kind of data transformation. Data transformation could require transforming the values within the same column (e.g., normalizing the data to fit the range of [0, 1]), performing data joining (e.g., merging two tables in a database into a single table), or deriving new columns entirely.

For example, to complete the 2017 VAST MC1, a question pertaining to whether cars are speeding in the preserve will require calculating the distances between sensors – which can only be done by examining the map – and the elapsed time between two consecutive sensor readings of the same car. At the end of this step, students should have an updated list of questions and potential hypotheses, as well as their strategies for data transformation that are necessary to answer their questions.

4.2.3 Step 3: Examining a Sample of the Data

During the third step, students should have access to a sample of the data. The purpose of this step is to do a "trial run" of the full analysis experience so that the students can: (1) generate initial visualizations and analysis scripts based on the sample data, and (2) assess whether the tools that they have are sufficient for their analysis needs.

As an example, in the 2017 VAST MC1, the students should recognize that the map (as an image) will be difficult to work with, but analysis of which will be important. Road information (e.g., connectivity and distances between sensor locations) can only be extracted by examining the map. However, as there is no standard mechanism to extract such information, the students should begin to think about whether they want to manually compute the needed data or write image-processing tools to do automatic extraction.

At the end of this step, the student should have a better sense of how their questions and hypotheses can be addressed with the data. They should also be able to document the tools that are necessary to complete their analysis and estimate the difficulty of completing remaining tasks.

4.3 Improving Hypotheses and Analysis Questions

Most undergraduate students in a visual analytics course do not have formal training in business or intelligence analysis, which is often a strong component in solving VAST Challenges. Without such training, students can have difficulty coming up with "good" analysis questions and thus testable hypotheses.

For example, in a previous visual analytics university course where students were given the 2017 VAST MC1, a student submitted the following hypothesis to Q3: "*certain visitors are irregularly active in certain regions of the park, causing the decreased nesting.*" Although this hypothesis matches the problem description (a geospatial-temporal analysis of vehicles), this hypothesis lacks specificity for further analysis (i.e., who are the "*certain visitors*," what are the "*certain regions*," and how to define "*irregular active*?").

The purpose of this stage is to help the students generate, assess, and refine their analysis questions and hypotheses. As shown in Figure 1, this exercise happens at the end of each step of the Problem-Data Mapping stage. The goal is to ask the students to reflect on their analysis questions and refine them into formalized hypotheses before moving to the next step in the process.

4.3.1 Hypothesis Specification

One way to assess the quality of an analysis question is to formalize the question as a scientific hypothesis. A valid scientific hypothesis states a qualified or quantified relation between two variables and is falsifiable (i.e., it has a binary true or false outcome) [20]. Rich literature has been contributed on assessing and refining the quality of a hypothesis from the science education community [15, 17, 28]. For the purpose of our project, we use the Hypothesis Quality Scale proposed by Sunal and Haas [27] (shown in Table 1).

To help students with the hypothesis specification process, consider the following example where the student might start with the following analysis question:

- *When do cars generally enter the park?*

This represents a Level-0 hypothesis in the Hypothesis Quality Scale in that it is a question without explanation. An improvement over this question might be:

- *Do cars enter the park mostly in the mornings?*

This question might represent a Level-2 hypothesis in that it has a true or false answer, but it is incomplete in its references. Another iteration of improvement results in:

- *There are more cars entering the park between 5-8am than any other three hour period.*

This hypothesis is testable and specific, representing a Level-4 hypothesis in the Quality Scale.

4.3.2 Hypotheses Refinement

Representing an analysis question formally as a hypothesis has the additional advantage that it can be further refined. For example, starting with the previous hypothesis, a student can now alter parts of the hypothesis to afford more flexibility in the analysis:

- *There are more cars entering the park between 5-9am than any other four hour period.*

In addition to changing the value ranges, the students can also add additional specifications:

- *There are more 2-axle cars entering the park between 5-9am than any other car type in a four hour period.*

Students are encouraged to reflect on the hypothesis refinement process with respect to the VAST Challenge questions. In the case of the 2017 VAST MC1 Challenge – which hints at a geospatial-temporal analysis grouped by car-types – the students should investigate refining their hypotheses along those three dimensions.

4.4 Data Analysis

The last step of the guideline is for the students to perform their planned data analyses in solving the VAST Challenge using the hypotheses that they have developed in the previous stages. In some courses, the students use existing tools to solve the challenge (e.g., the courses at Virginia Tech [31]) while other courses require the students to develop novel visual analytics systems that are tailored for the challenge (e.g., the visual analytics course at University of Konstanz [21]).

Regardless of the structure of the course, it is important to note that once the students start the data analysis process, it is not advisable for them to return to the early stages of the guideline. Posing research questions and hypotheses after data analysis runs the risk of HARKing and Type I errors (false discoveries) in the analyses [32].

Whether a course instructor should strictly disallow students from back-tracking depends on the course's learning objectives. One could argue that back-tracking should be allowed, as students should be free to make mistakes in a classroom setting as long as they recognize these mistakes and can correct them. Regardless of the instructor's decision on this matter, our guideline makes clear of the danger of back-tracking and strongly discourages the students from starting the data analysis process before being certain of their hypotheses.

We note that our guideline does not specify how data analysis should be conducted, instead, specifics should be justified by the student team in a write-up or predetermined by the course instructor. We discuss future work in which the guideline can be extended to include performing data analysis for students in Section 7.

5 APPLYING THE GUIDELINE

We perform a preliminary assessment of our guideline by applying the proposed workflow to a different VAST Challenge.

Participants: The same two students who analyzed the 2017 VAST Challenge, described in Section 3, participated in this exercise.

Material: We use the 2016 VAST Challenge Mini-Challenge 2 (MC2). This challenge presents a fictitious situation set on the island of Kronos where the company GASTech has just moved into a new building. The new building is equipped with sensors that detect a variety of building characteristics, such as temperature and chemical concentrations. GASTech has also implemented a new security procedure that requires

staff to carry *prox cards*, which are read by mobile and stationary sensors located throughout the building and logged in a database. The goal of this challenge is to identify both typical and atypical patterns or areas of concern, given two weeks of building and prox card data.

We note that the 2016 VAST MC2 is more complex than the 2017 VAST MC1. Instead of only one tabular dataset with one map, the 2016 MC2 contains 8 tables, 10 JSON documents, and 9 map figures.

Task: The two participants were tasked with analyzing the 2016 Challenge MC2 independently (no communication occurred between the two until the end of the task). Similar to the previous exercise, the participants were required to document detailed notes on their analysis and hypothesis refinement process – but in this case, their approach succinctly followed the proposed guideline. The participants were only required to reach Step 3 (examining data samples) of Problem-Data Mapping (Section 4.3), and did not perform any data analysis in the completion of their task.

Outcomes: The two participants spent approximately 8 days to complete this task (compared to 2 weeks in the previous exercise). They were consistent in the time spent in each stage of the guideline. Both spent 1-2 days generating their curiosity questions, 1-1.5 days between Step 1 (examining data schemas) and Step 2 (examining summary statistics) of Problem-Data Mapping, and 2-3 days in Step 3 (examining data samples). In total, the participants spent 2-3 days specifying and refining their analysis questions into formalized hypotheses.

6 OUTCOMES AND LESSONS LEARNED

Using our guideline to solve the 2016 VAST MC2 provided a preliminary assessment of how this guideline could be helpful for future visual analytics students. On the positive side, the two participants reported that using the guideline resulted in a clearer and more streamlined experience where each step of the analysis process was concretely laid out. However, the two participants also noted areas for improvement, which we document in this section.

6.1 Quality and Quantity of the Generated Hypotheses

We first examine the quantity and quality of the hypotheses generated by the two participants (PID1 and PID2 respectively).

PID1 generated 24 curiosity questions after conducting initial research into the challenge. After peeking at the data in Steps 1 and 2, PID1 added 18 new analytic questions. PID1 was then able to form 20 hypotheses after Step 3, and finally formalized 29 total hypotheses after refinement. However, PID1 noted that many of the 29 hypotheses have “specifications” that can be exchanged to pose different hypotheses (examples shown in blue in Section 4.3.2). PID2 started with 8 curiosity questions after finishing initial research into the challenge. After peeking at the data in Steps 1-3, PID2 had 14 additional questions. In total, PID2 generated 15 hypotheses from 22 analytic questions.

Both participants generated hypotheses that could lead to discovering the ground truth (or valid answer) to the challenge. For example, a ground truth provided in the 2016 VAST MC2 solution is that delivery workers enter/exit through the docks. A relevant hypothesis our participants came up with was “There is a staff member who starts/ends the day not on floor 1, zone 1 (the main entrance).” An evaluation of this hypothesis as true will help reveal delivery members that follow an “atypical pattern,” as described by the challenge.

Another ground truth (or valid answer) provided in the solution sheet is that hazium concentration spikes on F3Z1 on June 3 at 7AM, as well as on June 11 at 8PM. A relevant hypothesis formed by the participants was “Between 12-1PM on 5/31 there is an entry for hazium concentration over 0.” In this case, “Between 12-1PM” can be exchanged for any 1-hour period and “5/31” can be exchanged for any date. An evaluation of these hypotheses, using T as the correct time frame, will help the participants identify hazium concentration spikes.

Although our participants did not perform any data analysis to complete the 2016 VAST MC2, our preliminary findings suggest that, by following the guideline, future students will be able to identify good “starting points” for further investigation while following best practices in pre-registration and p-hacking prevention.

6.2 Positive Lessons Learned

The two participants reported three ways that the guideline helped with their analysis task.

A Streamlined Process: The participants found that the guideline provided a streamlined process for them to follow, resulting in time saved by the task being simpler to tackle. For the 2017 VAST Challenge, both participants had difficulty knowing where to begin analysis after reading the questions – one started doing related research, while the other began posing questions. In the second task with the 2016 VAST Challenge, the procedures outlined in the “Problem Understanding” stage (Section 4.1) was helpful in prescribing a place to start.

Concrete Directions for Each Step: The participants noted that the guideline provided concrete directions for each step of the analysis process. Since “peeking at the data” (Section 4.2) was described to have three steps between examining the data schemas, the summary statistics, and a data sample, it was easier for the participants to know when and how much of the data they were allowed to access. The clear distinction between the three steps also prevented the participants from looking at the full dataset before they had the chance to reflect on all possible questions, hypotheses, and analysis tasks.

Guidance for Hypothesis Specification and Refinement: The participants found the guidance for hypothesis specification and refinement (Section 4.3) to be useful in two ways. First, following the process resulted in clearer hypotheses that were easy to share and communicate with others. Second, with the use of the formal hypotheses, it was easier for the participants to consider alternatives via refinement. For example, the participants were able to make the hypotheses more or less specific by considering different subsets of the data.

6.3 Areas for Improvement

While the two participants were positive about the guideline overall, they also noted two areas where it is incomplete or can be improved.

How Many Hypotheses are Enough? The 2016 VAST MC2 is more complex than the 2017 VAST MC1: 2017 VAST MC1 contains only one tabular dataset, while the 2016 VAST MC2 has 8 tables that need to be joined prior to analysis. As a result of the increased data complexity, the participants found that the guideline’s suggestion of “exhaustively generate as many hypotheses as possible” was not attainable. A practical method to help students decide when they can move on to the next step will be necessary to address the issue of data complexity.

How Good are the Hypotheses? Related, if the students cannot realistically generate all hypotheses, they need to be provided a way to rank the hypotheses and determine which ones to keep. A strategy that the two participants used to combat this challenge was to group similar analysis questions and hypotheses into semantically meaningful clusters. In future iterations of this guideline, we will look to incorporate this as a possible strategy for students to follow.

7 LIMITATIONS AND FUTURE WORK

We acknowledge that this guideline is only at a preliminary phase and is not ready for wide adoption without further considerations. As the guideline is generated by the same two students who tested the guideline, the lessons learned can be biased by their personal experience. For example, although the two students were much faster in solving the second VAST Challenge using the guideline (8 days compared to 14 days), the speed-up could be due to their experience and not the guideline. Similarly, we presume that there are other areas for improvement. The two points raised by the students both relate to the applicability of the guideline with regards to the changes in data complexity. We imagine that if the guideline was used by a student who has never seen a VAST Challenge before, they would likely have additional suggestions or barriers faced.

In future work, we plan to refine the guideline by incorporating examples and lessons learned from the 2016 VAST MC2 exercise. Additionally, we will add prescriptive details on performing analysis given the students’ completion of their hypotheses (i.e., following Section 4.3). Further documentation for the guideline will be required

to prepare it for dissemination to students in a visual analytics course. Finally, we plan to conduct a controlled study in a visual analytics course that the co-author teaches at Tufts University in Spring 2023 to assess the effectiveness and limitations of the guideline in helping students tackle the VAST Challenge.

8 DISCUSSION

Although our proposed guideline is preliminary, our experience in developing the guideline has raised some interesting questions related to how VAST Challenges should be approached and whether the guideline can be generalized to visual analytics tasks beyond the use of the VAST Challenge in a classroom setting.

Data-Driven or Hypothesis-Driven? While there has been no other guidelines in teaching students how to solve the VAST Challenge, past research seems to suggest that the students begin addressing a challenge by first performing data exploration (e.g., [31]). Through the interactive exploration, the students form insights, come up with hypotheses, and use the visual analytics tool to refine and evaluate those hypotheses.

Our proposed hypothesis-driven guideline to solving the VAST Challenge is a departure from this approach. We are encouraged by our preliminary finding that the students were able to generate hypotheses that were on the right track to discovering the ground truth without having access to the full dataset. However, we note that this positive finding is not unique and has been studied with similar success in prior research [3, 8, 9, 13, 14]. Our work adds to this growing area of research and suggests the possibility of using a hypothesis-driven approach, as opposed to data-driven, for education purposes.

Generalizability Beyond the VAST Challenge? We consider the possibility of generalizing the proposed guideline to help analysts perform visual data analysis beyond the VAST Challenge. VAST Challenges are modeled after real-world scenarios and have been touted to reflect intelligence analysis in practice [5, 24]. However, we observe that there are also differences that could prevent the guideline from being applicable to general data analysis tasks. First, VAST Challenges typically provide questions that hint at the final solution. In the 2017 VAST MC1, the questions indicate that the correct answer involves analyzing the data by first grouping the rows by car-types. In real-world analysis tasks, it is not always the case that such instructions are available.

Second, real-world data are often much more complex than what is given in VAST Challenges. When switching from the 2017 VAST MC1 to the 2016 VAST MC2, we observed a significant increase in the number of possible hypotheses that the students could have formed. With even larger and more complex data, our proposed hypothesis-driven guideline would likely reach its limit, as an analyst could not systematically and exhaustively generate all possible hypotheses. We speculate that there are ways to mitigate this problem (e.g., by clustering the hypotheses as the students did). We leave the challenge of generalization for our guideline as future work.

9 CONCLUSION

In this paper, we present a preliminary guideline to help students in a university-level visual analytics course tackle a VAST Challenge. Our work is based on the observation that without guidance, students currently struggle with identifying a good “starting point” when tasked with analyzing VAST Challenges. Our guideline is developed based on our experience solving the VAST 2017 Mini-Challenge 1 following a hypothesis-driven approach. After completion, we applied the guideline to the VAST 2016 Mini-Challenge 2 to assess its benefits and limitations. The preliminary results from this exercise suggest that the guideline provides the students with a streamlined approach and concrete steps to follow, reducing the students’ time spent in the early phase of the task by identifying good analysis directions to pursue.

ACKNOWLEDGMENTS

The authors wish to thank the NSF Diamond REU Program at Tufts University (2149871) and NSF grants 1855886, 1940175, 1452977, 1939945, 2118201 for supporting this work.

REFERENCES

- [1] N. Andrienko, T. Lammarsch, G. Andrienko, G. Fuchs, D. Keim, S. Miksch, and A. Rind. Viewing visual analytics as model building. In *Computer graphics forum*, vol. 37, pp. 275–299. Wiley Online Library, 2018.
- [2] D. Cashman, S. R. Humayoun, F. Heimerl, K. Park, S. Das, J. Thompson, B. Saket, A. Mosca, J. Stasko, A. Endert, et al. A user-based visual analytics workflow for exploratory model analysis. In *Computer Graphics Forum*, vol. 38, pp. 185–199. Wiley Online Library, 2019.
- [3] I. K. Choi, T. Childers, N. K. Raveendranath, S. Mishra, K. Harris, and K. Reda. Concept-driven visual analytics: an exploratory study of model- and hypothesis-based reasoning with visualizations. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pp. 1–14, 2019.
- [4] A. Cockburn, C. Gutwin, and A. Dix. Hark no more: On the preregistration of chi experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173715
- [5] K. Cook, G. Grinstein, and M. Whiting. The vast challenge: History, scope, and outcomes: An introduction to the special issue, 2014.
- [6] R. J. Crouser, K. Cook, J. Fallon, G. Grinstein, K. Liggett, D. Nebesh, M. A. Whiting, and K. Whitley. Vast challenge 2016: Streaming visual analytics. In *IEEE Conference on Visual Analytics Science and Technology*, 2016.
- [7] P. Dragicevic. Fair statistical communication in hci. In *Modern statistical methods for HCI*, pp. 291–330. Springer, 2016.
- [8] J. Hullman and A. Gelman. Designing for interactive exploratory data analysis requires theories of graphical inference. 2021.
- [9] E. Jun, M. Daum, J. Roesch, S. Chasins, E. Berger, R. Just, and K. Reincke. Tea: A high-level language and runtime system for automating statistical analysis. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 591–603, 2019.
- [10] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In *Information visualization*, pp. 154–175. Springer, 2008.
- [11] N. L. Kerr. Harking: Hypothesizing after the results are known. *Personality and social psychology review*, 2(3):196–217, 1998.
- [12] Y.-S. Kim, L. A. Walls, P. Krafft, and J. Hullman. A bayesian cognition approach to improve data visualization. CHI '19, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300912
- [13] Y.-S. Kim, L. A. Walls, P. Krafft, and J. Hullman. A bayesian cognition approach to improve data visualization. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pp. 1–14, 2019.
- [14] R. Koonchanok, P. Baser, A. Sikharam, N. K. Raveendranath, and K. Reda. Data prophecy: Exploring the effects of belief elicitation in visual analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2021.
- [15] K. A. Kroeze, S. M. van Den Berg, A. W. Lazonder, B. P. Veldkamp, and T. de Jong. Automated feedback can improve hypothesis quality. In *Frontiers in Education*, vol. 3, p. 116. Frontiers, 2019.
- [16] K. M. MacQueen, E. McLellan, K. Kay, and B. Milstein. Codebook development for team-based qualitative analysis. *Cam Journal*, 10(2):31–36, 1998.
- [17] Y. G. Mulder, A. W. Lazonder, and T. de Jong. Finding out how they find it out: an empirical analysis of inquiry learners' need for support. *International Journal of Science Education*, 32(15):2033–2053, 2010.
- [18] B. A. Nosek, C. R. Ebersole, A. C. DeHaven, and D. T. Mellor. The pre-registration revolution. *Proceedings of the National Academy of Sciences*, 115(11):2600–2606, 2018.
- [19] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, vol. 5, pp. 2–4. McLean, VA, USA, 2005.
- [20] K. Popper. *The logic of scientific discovery*. Routledge, 2005.
- [21] C. Rohrdantz, F. Mansmann, C. North, and D. A. Keim. Augmenting the educational curriculum with the visual analytics science and technology challenge: Opportunities and pitfalls. *Information Visualization*, 13(4):313–325, 2014.
- [22] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE transactions on visualization and computer graphics*, 20(12):1604–1613, 2014.
- [23] J. Scholtz, C. Plaisant, M. Whiting, and G. Grinstein. Evaluation of visual analytics environments: The road to the visual analytics science and technology challenge evaluation methodology. *Information Visualization*, 13(4):326–335, 2014.
- [24] J. Scholtz, M. A. Whiting, C. Plaisant, and G. Grinstein. A reflection on seven years of the vast challenge. In *Proceedings of the 2012 BE-LIV Workshop: Beyond Time and Errors-Novel Evaluation Methods for Visualization*, pp. 1–8, 2012.
- [25] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE transactions on visualization and computer graphics*, 18(12):2431–2440, 2012.
- [26] A. Suh, Y. Jiang, A. Mosca, E. Wu, and R. Chang. A grammar for hypothesis-driven visual analysis, 2022. doi: 10.48550/ARXIV.2204.14267
- [27] C. S. Sunal and M. E. Haas. *Social Studies for the Elementary and Middle Grades: A Constructivist Approach*. ERIC, 2002.
- [28] W. R. Van Joolingen and T. De Jong. Supporting hypothesis generation by learners exploring an interactive computer simulation. *Instructional Science*, 20(5):389–404, 1991.
- [29] J. J. Van Wijk. Views on visualization. *IEEE transactions on visualization and computer graphics*, 12(4):421–432, 2006.
- [30] M. A. Whiting, K. Cook, R. J. Crouser, J. Fallon, G. Grinstein, J. Haack, C. Henderson, K. Liggett, D. Staheli, J. Strasburg, et al. Vast challenge 2017: Mystery at the wildlife preserve. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 173–178. IEEE, 2017.
- [31] M. A. Whiting, C. North, A. Endert, J. Scholtz, J. Haack, C. Varley, and J. Thomas. Vast contest dataset use in education. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pp. 115–122. IEEE, 2009.
- [32] E. Zraggen, Z. Zhao, R. Zeleznik, and T. Kraska. Investigating the effect of the multiple comparisons problem in visual analysis. In *Proceedings of the 2018 chi conference on human factors in computing systems*, pp. 1–12, 2018.