Project in Visualization and Data Analysis: Experiences in Designing and Coordinating the Course

K. Kucher¹, R. M. Martins¹, and A. Kerren¹,²

¹Department of Computer Science and Media Technology, Linnaeus University, Sweden
²Department of Science and Technology, Linköping University, Sweden

Abstract

Visual analytics involves both visual and computational components for empowering human analysts who face the challenges of making sense and making use of large and heterogeneous data sets in various application domains. In order to facilitate the learning process for the students at higher education institutions with regard to both the theoretical knowledge and practical skills in visual analytics, the respective courses must cover a variety of topics and include multiple assessment methods and activities. In this paper, we report on the design and first instantiation of a full term project-based course in visualization and data analysis, which was recently offered to graduate and post-graduate students at our department and met with positive feedback from the course participants.

CCS Concepts

• Human-centered computing → Visual analytics; • Information systems → Data mining; • Social and professional topics → Computing education;

1. Introduction

Visual analytic systems [KKEM10, SSS¹¹⁴] bring data analysis closer to end users by effectively combining interactive visualization and complex algorithms, guided by the underlying analytical processes inherent to the data and the application at hand. The methods and specific techniques proposed within the information visualization (InfoVis) and visual analytics (VA) research communities have been attracting a lot of attention within academia, industry, and public sector, as they can facilitate the processes of representing and exploring complex data [FvWSN⁰⁸], deriving new insights and knowledge from the data [SSS¹¹⁴], externalizing the respective findings and models [ALA¹⁸], and raising the trustworthiness of the underlying analyses and models [CMJ²⁰].

However, implementation of VA solutions requires understanding of the concepts and mastery of the skills corresponding to computational methods, interactive visualization design [Mun⁰⁹], integration of both groups of approaches, and user-centered evaluation of the resulting solutions [Pur¹², IIC¹³, CE¹⁹]. This wide variety of topics and skills presents challenges for educational activities focusing on VA, in particular, higher education courses. Thus, further discussion of the design and outcomes of such courses can be helpful for the community of academic educators.
• Importance of data and visualization for answering analytical questions
• Selected examples of state-of-the-art VA systems
• Information visualization and visual analytics in applied projects
• Data analysis and processing in realistic projects
• Tools, services, and software libraries for data analysis and visualization
• Challenges and opportunities at the interfaces between the human analyst, computational models, and visual display
• Evaluation of visualizations in applied projects
• Types of bias in data, analysis, and visualization

Table 1: The PVDA course topics specified in the syllabus.

Inspired by the existing work discussing the design and implementation of courses in InfoVis and VA, we report on the design and first instantiation of a full term PVDA (Project in Visualization and Data Analysis) course, which was offered to graduate and post-graduate students at Linnaeus University during autumn term 2020 (see Figure 1) and met with positive feedback from the course participants.

The course is designed to cover multiple topics in VA, as listed in Table 1, and it focuses on a VA project with a given analytical problem and setting. The students are expected to work using agile processes [CH01, HC01] in teams. They are first introduced to the theoretical aspects and tools for VA as the initial preparation for the project work. Afterwards, the course participants are expected to create the conceptual design of the VA project, implement their designs as a VA software prototype, conduct an evaluation of their solution, and finally, present and reflect on the results.

In the rest of this education paper, we describe the design of our PVDA course in the context of the curriculum and additional circumstances regarding graduate and post-graduate education at our department. We describe the design of our course, the specific details of the first course offering that took place during autumn term 2020, and the respective outcomes.

2. Related work

The general pedagogical principles, guidelines, and techniques applicable for higher education have been discussed by, for instance, Anderson and Krathwohl [AK01], Collis and Moonen [CM12], Ramsden [Ram03], and Biggs and Tang [BT07]. Curricula for computer science and related fields [Joi13] have been proposed and widely discussed, too.

More specific publications relevant to our paper include the previous reports and discussions of courses in InfoVis and VA, including the works by Domik [Dom09], Hanrahan [Han05], Rusmeier et al. [RDDY07], Kerren et al. [KSD08], Owen et al. [ODE13], Kerren [Ker13], Rohrdantz et al. [RMNK14], Ahlers [Ahl15], and Spence [Spe16]. These contributions were very useful for the design of instructional materials and assessment activities for our course, which are discussed in the next section. Finally, we should also acknowledge the existing works addressing the challenges and opportunities of offering visualization courses outside (or on the fringe) of the scope of traditional computer science educational programs, including the contributions by Domik [Dom09, Dom12, Dom16], Elmqvist and Ebert [EE12], Jánicek [Ján20], Santos and Perer [SP20], and Burch and Melby [BM20], among others. These works are relevant to our course as we had anticipated the scenarios involving admission of graduate and post-graduate students without prior knowledge in visualization or even limited knowledge in computer science in general, as discussed in more detail below.

3. PVDA course design

In this section, we describe the design of our course, starting with top-level concerns and circumstances, and proceeding with the discussion of lecture materials and assessment methods.

3.1. Curriculum and intended participants

While the research and educational agenda of our department strongly focuses on the field of software engineering, our research group offers several courses in computer graphics (for bachelor’s students) and visualization (for master’s students). The PVDA course is a recent addition to the curriculum, and it is designed as a 10 ECTS (European Credit Transfer and Accumulation System) credits course with 33% activity rate lasting a complete term of 18 study weeks + 2 weeks of the winter break (see Figure 1). At the same time, most of the other courses offered at our department are designed for a single study period (quarter). This presents both challenges and opportunities with regard to scheduling, as this constraint also has to be reflected in the design of course prerequisites, as discussed below.

The PVDA course is designed for master’s students who are expected to be reasonably proficient in programming tasks as well as team work within software development projects [CH01, HC01]. However, no prior knowledge or skills in InfoVis or VA are expected from most students (including international and exchange students, among others). Therefore, the syllabus for the PVDA course does not include any formal prerequisites for visualization courses, and this project course is thus designed to be self-sufficient with regard to teaching material and tasks on the respective topics. The students are notified and strongly encouraged, however, to enroll in at least the first InfoVis course offered in the first study period of the same term (see Figure 1).

Besides master’s students, we have also allowed PhD students at our department to participate in the PVDA course; furthermore, as our department is part of an interdisciplinary knowledge center at our university, we have agreed to potentially admit the respective PhD students majoring in different disciplines at our university, as long as they have the sufficient set of skills in data analysis and programming. This decision has been motivated by the interest for visualization and visual analysis skills and techniques from researchers and practitioners from a variety of disciplines and fields. We had expected that PhD students would be interested in immediately applying the knowledge and skills from our course to their ongoing research tasks, or perhaps be inspired with new ideas related to their fields of study. Thus, our intention was to allow PhD students to suggest their own data sets and tasks in order to work individually within the scope of this course; in contrast, master’s students would be required to work in teams.
After completing the course the student shall be able to:

### Knowledge and Understanding

- Explain and motivate information visualization and visual analytics from a human-computer perspective as well as how these areas facilitate data analysis
- Identify characteristics of state-of-the-art VA systems

### Skills and Abilities

- Independently organize and carry out an agile project
- Independently learn to use tools, methods, and software libraries used within InfoVis and VA
- Identify analytical requirements/tasks and based on these decide which visualizations, data analyses, and interaction methods are most suitable (design goals and choices)
- Implement and deploy a VA system using reasonable technical solutions
- Based on the design goals, define and perform evaluations of a VA system

### Judgement and Approach

- Reflect on which types of questions and analysis processes are best supported by different combinations of visualization techniques and data analysis algorithms
- Reflect on how the choice of visualizations is affected by the groups of people that will use the system with respect to, e.g., used metaphors and common understanding
- Reflect on the bias that exists in the system based on collected data, data processing, analysis methods, and visualization techniques

### Table 2: Intended learning outcomes for the PVDA course.

The PVDA course syllabus specifies the topics listed in Table 1, and the intended learning outcomes included in the syllabus are designed appropriately to structure and facilitate the students’ learning process. In accordance with the educational guidelines at our department, the intended learning outcomes (see Table 2) are grouped in three categories motivated by the revised Bloom’s taxonomy of educational objectives [AK01], namely, knowledge and understanding, skills and abilities, and judgment and approach. The instructional materials and assessment activities designed for this course in order to achieve constructive alignment with the intended learning outcomes [BT07] are discussed in the following subsections.

#### 3.2. Instructional materials

While it would be possible to focus exclusively on the students’ project work within the scope of this course, we understand that it would be extremely challenging for the master’s and even PhD students who lack prior experiences with InfoVis techniques, and even more so, VA approaches. Thus, we designed the first part of the course (7 weeks) to contain a number of lectures focusing on theoretical and practical aspects of VA (with a strong focus on visualization), supplemented with several pre-recorded practical tutorial videos in order to provide the students with initial examples and tips. The list of our lecture and tutorial topics is provided in Table 3, and it was designed to address the course topics specified in the syllabus (see Table 1). We have tried to balance the more theoretical (or research-oriented) materials with the practical lectures and tutorials related to the existing software (e.g., Tableau) and programming toolkits (e.g., Plotly, Bokeh, D3.js, but also scikit-learn, for instance) that might be useful for the students both within and beyond the scope of this course.

After discussing the overall VA approach as well as its visual and computational aspects, we proceed with a discussion of several example state-of-the-art VA systems as an inspiration for the students’ project work. The next major topic is evaluation of VA approaches, which is discussed from several points of view and aimed at preparing the students for the evaluation of their own projects by the end of our course. Finally, one of the more practical aspects addressed by our tutorials was the deployment of the resulting VA applications, as we intended for most students to develop their projects as web applications (at least with regard to the interactive visualization components).

While we did not include any mandatory reading tasks in this course and also aimed to design the instructional materials to be self-sufficient, the course literature recommended to the students involved the books by Munzner [Mun14] (focusing on visualization), Keim et al. [KKEM10] (visual analytics), and Purchase [Pur12] (evaluation methods).

#### 3.3. Assessment and examination

The PVDA course was intended to focus on project work, including team work (for master’s students) and presentation activities. Our general expectations for a project topic include the presence of both computational and interactive visual components in the corresponding software implementations; the more concrete analytical requirements, user tasks, and designs would have to be proposed by the students for the chosen problem and the respective data set. The PhD students are also allowed—and in fact strongly encouraged—to propose their own research problems and data sets for the project.

All of the submissions and mandatory sessions are centered around the preliminary exploration of the provided data set, formulation of an initial design of a VA approach for the data set, and iterative project work according to the agile development principles, as demonstrated in Table 3. The active phase of such project work would start after the initial block of lectures and tutorials was completed (see the timeline in Figure 1). By the end of each project iteration (typically lasting two weeks), the participants are required to make a working release of their prototype (we use a local installation of Gitlab for these purposes), supplement it with a short report (essentially, notes about the progress and plans for the next iteration), and demonstrate this updated prototype during a seminar afterwards. During these seminars (as well as optional tutoring sessions), the participants would then receive feedback from the instructors as well as other students.
After several project iterations, the participants are expected to conduct evaluations of their prototype solutions with several fellow students as the study participants. These sessions are expected to provide mainly qualitative feedback on the usability of the prototypes (either in general or in relation to the particular user tasks, depending on the study design). To find a common denominator, we have also decided to make it mandatory to use the ICE-T questionnaire [WAM*19] while providing the feedback.

The final steps after the evaluation session include reporting the corresponding evaluation results, working on the final project release that should also take the feedback from the evaluation session into account to some degree, and conducting the final presentation of the project. Besides the software implementation and presentation, the students are also required to submit two reports: (1) a group report on the project results, titled “Human-centered aspects of visual analytics”; and (2) an individual report titled “How to manage an agile project”, reflecting on the project process instead.

4. PVDA course offering results

Our PVDA course was first offered during autumn term 2020; thus, the course planning phase and the various course activities including the team work all had to take place amid the COVID-19 related events. This has put additional stress on both the instructors and the course participants, but fortunately, the course offering proceeded without major issues overall. Most of the teaching tasks were divided between the instructors with regard to the lectures and tutorials. Seminars (including project presentations) were typically attended by several instructors, and we were thus able to track the progress of the participants and provide them with extensive feedback. Additionally, we provided tutoring sessions during the weeks free from seminars.

The course participants this year included one team of three master’s students in computer science, who also attended the data mining course and one or both visualization courses during the same term (see Figure 1). To assign them with a data set and tasks for their project, we decided to make use of the recent data from the VAST Challenge [CGW14], which has been successfully used through the years by other colleagues for teaching [RMNK14] and undergraduate student research [AC19] purposes. After suggesting several options, the students chose the VAST Challenge 2019 Mini-Challenge 2 data set [CCH*19], which focuses on spatiotemporal data analysis [HMC*19, WST*19, WLS*21, XLQ*19]. As the result, they developed a VA approach supporting clustering and classification of time-series data, which can be investigated and controlled by the user from a web-based interactive visualization interface, as demonstrated in Figure 2. In general, the reports and verbal feedback from the students suggest that they were interested by the subject and appreciated the ability to represent and interact with the results of computational analyses as part of their VA solutions; the evaluation experience was also valuable.

The course also included four PhD students, two of whom specialize in InfoVis and VA. All of the PhD students worked on the problems within the scope of their respective research interests, and implemented web-based VA tools. The InfoVis PhD students are now both working on publications based on their project work. Another participant, who is conducting his postgraduate education as an industrial PhD student affiliated with a large corporation, has implemented a prototype tool addressing an existing data analysis problem at his main affiliation, and intends to continue the development in order to make use of this solution alongside his colleagues (i.e., industrial practitioners).

5. Discussion

While our department is not large with regard to the number of master’s students enrolled, the design of our PVDA course is scalable regarding the number of teams and, to some extent, the number of members in each team. Thus, while our first course offering involved only seven active participants (as opposed to, e.g., the experiences reported by Burch and Melby [BM20]), in the future it could accommodate a larger participant number [Ker13] without issues. Furthermore, in such cases we could consider mixing master’s and PhD students within teams, and also assigning the roles related to specific agile methodologies such as Scrum [HC01]. This time, the master’s student team organized itself according to the compe-

<table>
<thead>
<tr>
<th>Instructional Materials</th>
<th>Mandatory Activities and Seminars</th>
<th>Deliverables</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Introduction and Motivation for Visual Analytics</td>
<td>• Project Data Set Presentation</td>
<td>• Project Proposal Draft</td>
</tr>
<tr>
<td>• Information Visualization Lecture I</td>
<td>• Project Proposal Discussion</td>
<td>• Final Project Proposal</td>
</tr>
<tr>
<td>• Information Visualization Lecture II</td>
<td>• Project Implementation Kick-off</td>
<td>• Release + Short Progress Report (1)</td>
</tr>
<tr>
<td>• Practical Lecture I: High-Level Visualization Tools</td>
<td>• Demo + Planning (1)</td>
<td>• Release + Short Progress Report (2)</td>
</tr>
<tr>
<td>• Video Tutorial I: Initial Data Preprocessing and Exploration</td>
<td>• Demo + Planning (2)</td>
<td>• Release + Short Progress Report (3)</td>
</tr>
<tr>
<td>• Visual Analytics Lecture I</td>
<td>• Demo + Planning (3)</td>
<td>• Evaluation Results Report</td>
</tr>
<tr>
<td>• Practical Lecture II: Visualization Toolkits and Libraries</td>
<td>• Project Evaluation Session</td>
<td>• Final Project Release</td>
</tr>
<tr>
<td>• Video Tutorial II: Initial Design</td>
<td>• Demo + Planning (4)</td>
<td>• Project Results Group Report</td>
</tr>
<tr>
<td>• Practical Lecture III: Computational Toolkits</td>
<td>• Final Project Presentation + Demo</td>
<td>• Project Process Individual Report</td>
</tr>
</tbody>
</table>

Table 3: Overview of the PVDA course contents.
tency and interests, which aligns well with the general agile principles [CH01]. The team work and communication were somewhat hindered by the necessity for remote work due to COVID-19, which was mentioned by the students in their reflection reports. We also witnessed several complications during the VA prototype evaluation sessions, which also had to be conducted remotely via Zoom; hopefully, during the future course offerings this issue will not be actual anymore. Finally, one additional idea for a future course improvement is to slightly adjust the schedule in the first half of the course, so that the active project work would start earlier and several further project iterations could be included into the time plan.

6. Conclusions
In this paper, we have described the design of a full term project course on visualization and data analysis that we had recently developed. Both graduate and post-graduate student participants were overall positive about the first course offering that took place during autumn term 2020. Our analysis of their activity and results in comparison to the intended learning outcomes also indicates the positive results, despite the unexpected circumstances caused by COVID-19. We have also discussed the lessons learned and the plans for future adjustments of the course contents and schedule. We hope this report will be useful to other educators who are designing or planning to update their courses in visual analytics.

Acknowledgements
This work was partially supported through the ELLIIT environment for strategic research in Sweden. The authors would also like to thank the participants of our course given at LNU during autumn term 2020.

References