Towards Guidelines and Practices for Teaching Data Visualization

Bernice Rogowitz, Bum Chul Kwon, and Emily Fuhrman

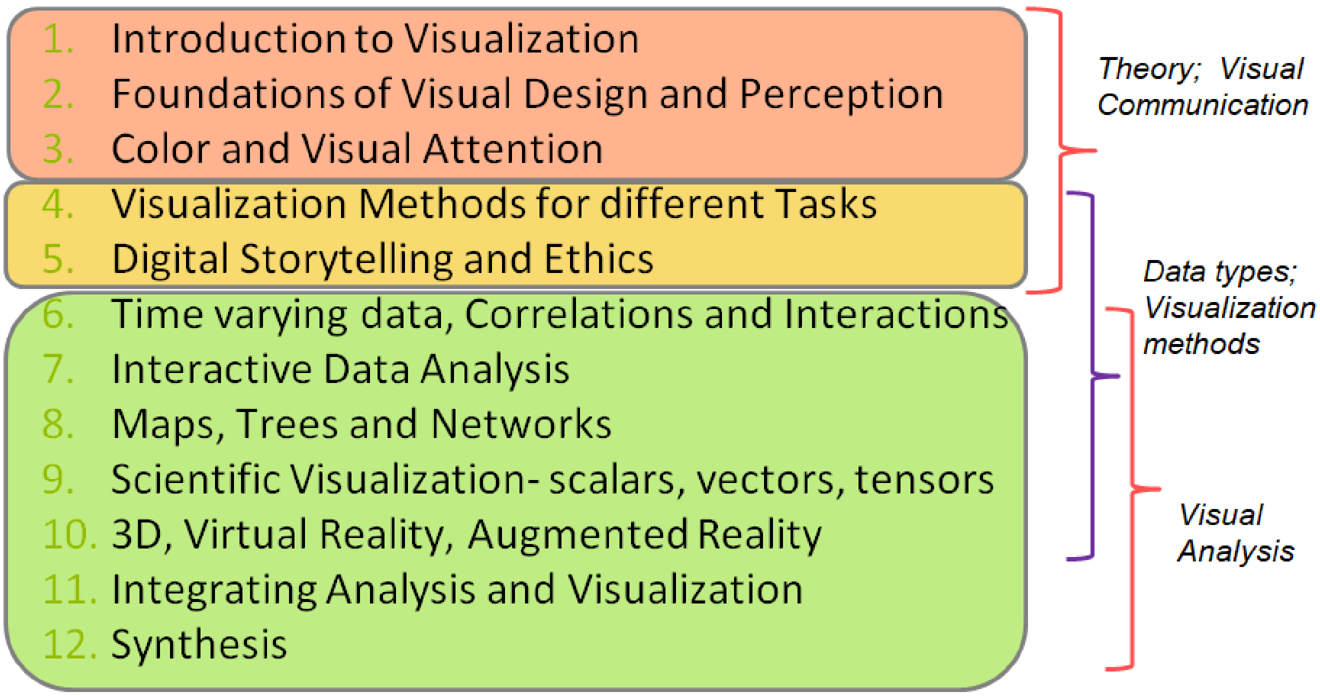

Fig. 1. Overview of Data Visualization Course

Abstract—The authors of this paper teach a masters-level course, called "Data Visualization and Design," at Columbia University. The goal of this paper is to share the components of the course, the principles and guidelines we are teaching to our students, and specific methods we have adopted. The course has three main themes: (1) Visual design, perception and cognition, (2) Fundamental insight into the mapping of different data types onto different visual geometries, depending on the task, and, perhaps most important, (3) Using visualization to discover patterns and features in data. The course includes readings in visualization, visual analytics, perception and design; hands-on homework assignments using current visualization software; and projects that challenge the students' skills in using visualization to solve real-world problems. We emphasize core skills that transcend specific choices of software, and which generalize across different data types and analytical methods. In teaching this course, we have looked at fundamental principles of data visualization science and practice through a pedagogical lens; we hope our experience will encourage further discussion in the visualization research community on how we teach, train, and assess students' mastery of visualization.

Index Terms—data visualization, pedagogy, education, visual analysis, perception, visualization literacy

1 Introduction

There has been a phenomenal growth in the past five years of Masters degree programs in Data Science. These programs attract students with undergraduate degrees in a wide range of disciplines, including business, economics, and computer science. These programs tend to focus on data acquisition, management, and analysis skills– few include visualization or visual design. Recently, however, data visualization has begun to be recognized as a core competency for Data Science. The program at Columbia University is in the vanguard of this movement. Starting in the Summer of 2017, it added a course in Data Visualization and Design to its core curriculum. The three authors of this paper were among the original instructors in this course, and the leading author was responsible for redesigning the course in the Fall of 2017, based on our collective experiences. This redesign strengthened the focus on principles from the visualization research community and increased the focus on the use of visualization for data analysis.

This course was specifically designed for students who have some technical background, but who are not necessarily programmers. Unlike academic visualization courses taught in Computer Science departments, this course does not assume that the students have pro-

- *Bernice Rogowitz is with Visual Perspectives. E-mail: bernice.e.rogowitz@gmail.com.*
- *Bum Chul Kwon is with IBM Research. E-mail: bumchul.kwon@us.ibm.com.*
- *Emily Fuhrman is with Two Sigma Investments, LP. E-mail: emily.fuhrman@twosigma.com.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/

gramming expertise, or even interest, and thus, this course focuses on providing guidelines and critical thinking skills that can be successfully applied without a formal understanding of computer science constructs.

Several questions drive the continued evolution of this course. What fundamental principles and guidelines will serve our students long-term needs? How do we teach visualization as a skill, in addition to an academic discipline? How do we judge what the students have really learned, and how well they will do in the real-world settings that demand their execution? And, which teaching methods are likely to be effective for a range of students and instructors? In this paper, we share our journey throughout multiple semesters and reflect upon our experiences. Our goal is to stimulate discussion in the visualization research community regarding guidelines for teaching visualizations to next generation data scientists.

2 The Design of the Data Visualization and Design Course

The Data Visualization and Design course exposes the students to a wide range of fundamental concepts from psychology, design, computer science, and data analysis, which they exercise using a variety of tools. The course materials include two texts, one from the academic visualization community (Tamara Munzners Visualization, Analysis and Design, 2014 [4]) and one from the world of visualization for newspapers, blogs and advocacy (Andy Kirks Data Visualization, 2016 [1]), plus selected primary source publications and visualization examples. In the real world, visualization is an integral part of a team effort that usually involves people from different disciplines and different goals. So, in addition to academic readings, the course builds collaboration and interaction skills through in-class design reviews and a group project. Throughout the course, there are hands-on exercises to build expertise with data visualization and design, culminating in a final project which requires the student to use exploratory visualization and analytic methods to discover a story in datasets they source, to demonstrate their ability to create a visual application that supports problem solving and decision-making. The students are told from the outset that they will be rewarded for being creative and ambitious.

This course provides an overview of modern data visualization and design theories, methods and techniques, and analytics. Throughout the course, they learn to use these principles to create visualization and visual applications that enhance comprehension, communication, and decision-making, through sections described in Fig. 1.

i **Perception and Design.** The first section of the course focuses on principles governing visual representations of data and analysis from graphic design, visual art, perceptual psychology, and cognitive science. The students learn that the semantic understanding of data is driven by how humans process the shapes, colors, and movements we map onto data, and that these representations not only influence detection and recognition, but also affect problem solving, decision-making, and interpretation. This section also includes a section on ethics, which complements the lesson on storytelling. A thin line separates advocacy from propaganda.

ii **Visualization Methods and Techniques.** This section provides a broad-based understanding of different data types, representation techniques, and tasks. They learn about different data types, data structures, and their representation. Although most of the course focuses on tabular data and maps, they gain experience with scalar fields, 3-D graphics and virtual/augmented reality, trees and networks. The lessons encourage the students to apply their learning about perception and design to the selection of visualization methods and their presentation.

iii **Analysis.** The third section of the course builds skills in using visualization for exploration, using interactive brushing, filtering, and highlighting. They integrate analytic methods, learning how to represent transformed data, fits to data, results of mathematical analyses, and models. The final project, which counts for 30% of their grade, demonstrates their ability to think visually, using perceptual and design principles, marshalling effective methods for exploring and representing phenomena in their data, and providing an effective visualization application.

Throughout the course, the curriculum is peppered with hands-on exercises, to give the students, for example, a hands-on feel for the three perceptual dimensions of color, or to feel how visual representations express semantic relationships. They are also introduced to many visualization software programs, which are exercised through homework assignments on shared data sources, and in class demonstrations. In the first few weeks, they draw by hand. For the first two projects, they use tools like Excel, Powerpoint and Infogram. For the second two projects, most go on to use the expressiveness of Tableau, Carto, R, D3.js, and Vega.js, to build all, or parts of their projects. Because they have studied fundamental principles of perception, design, and analysis, when they use these more advanced tools, they do so with some sophistication. That is, they are not simply putting data into tools. They are thinking about how the tools can help them discover and represent the features in the data, which can drive their analysis and give depth to the story these data contain.

3 How to teach technology in a world where technology always changes

While the basic principles of perception, design, analysis, and problem solving stay the same, the means through which to explore and apply them change. Visualization libraries, coding languages, and applications evolve over time. The rapid development of visualization technology tempers the allure of mastery with the threat of obsolescence. Students must learn to pick up new tools quickly, while developing practical competence in the ones that potential employees value. Over the past decade, there has been an explosion of software applications that provide easy access to visualization and analysis operations, and we anticipate that this trend will accelerate in the future. The tutorials for these tools, and the associated blogs, however, are often based on the particular experience of the designer, and often provide guidance that is not well-based in visualization research. Our goal is to teach our students to see these tools as vehicles for expression and exploration, and to use their own critical judgment in their application.

The course specifically addresses current, in-demand technologies, such as Tableau, R, and D3.js. Guided walkthroughs are included in the class resources, while individual assignments allow students to explore new tools more deeply on their own. Perhaps more importantly, the course teaches an attitude toward technology as a means to an end rather than an end in itself. This prepares the student for a professional setting, where available technologies might shift depending on the context.

When introducing new tools to students, reducing the barrier to entry is paramount. Interactive notebooks are useful for introducing syntax without the burden of setting up an environment, while posted tutorials enable participants to move at their own pace. To get students up and running with low friction accomplishes two ends: (1) deemphasizes the tools themselves by bringing into focus what they can facilitate, and (2) breaks through the aura of inaccessibility that surrounds advanced technology. In the visualization classroom, the ability to approach technology with confidence builds flexibility, depth, and intuition in the application of visualization to real-world communication and analysis.

With the advancement of web technologies and the growth of online courses throughout campus, we are in need of adopting new media to teach how to understand, use, and create data visualizations for a variety of analytic tasks. We have tried and found the following technologies promising to teach guidelines effectively. For these emerging media, we need discussion on how to effectively organize materials for different teaching objects and students. We launched and used interactive tutorials, TIVY [2] and VLAT [3] to teach how to read and use various kinds of visualizations. We also used a variety of interactive computing environments like Observable and iPython Notebook shared through Github repositories.

4 Visualization Teaching Methods

Our course focuses on fundamental principles from perception and design, provides guidance on selecting or creating the right visualization, and emphasizes the use of visual methods to explore, as well as communicate phenomena in the data.

As shown in Fig. 1, the course is divided into modules. Each module contains academic readings that expose the student to fundamental principles, which are exercised and reinforced through hands-on exercises. For example, the students read about the three fundamental perceptual dimensions of color (hue, luminance, and saturation), do an in-class exercise where they manipulate color chips to gain hands-on intuition into how these dimensions carry different visual information, and exercise color tools like Color Brewer, IWantHue, and color pickers to see a variety of methods for selecting and manipulating color. The lectures weave these academic and practical methods together to increase knowledge and skills, simultaneously.

During the course of the semester, several datasets are used and re-used to build different skills. For example, a dataset of economic variables is used in the early lessons to introduce different ways of mapping the same data onto different geometries, is used in a later lesson to introduce methods for exploring correlations between time-series variables, and how to use brushing to visually explore clusters in high-dimensional data, and is also used for the group project where they integrate individual visualization insights into a complete story.

We judge how well our students integrate the principles of visualization and design through their practical performance. There are no quizzes or exams. Each week they do homework assignments that exercise the academic concepts, but 75% of their grade depends on their performance on four visualization projects. The first project evaluates their ability to think visually, as they portray a verbal description of a banks mission into a visual infographic. The second reinforces the idea that different visual representations convey different meaning by having them produce a visual application that contrasts visual truths and lies drawn from the same dataset. Since visualization is usually part of a collaborative experience in the real world, the third project divides the class into groups of three to five to create a visualization application based on the now-familiar economy dataset. For this project, they not only learn visualization skills, but also learn to assign roles, divide work responsibilities, and collaborate on visual design and analysis. This group experience also allows them to teach each other as they build their collaborative design. In the fourth project, students source their own data and build a visualization application that demonstrates their ability to explore data, identify important insights, and to provide a meaningful navigation path for viewers. For this project, they also produce a storyboard that shows that they have given explicit thought to the way the analysis is unveiled, a description of the analyses they have performed, and a compendium of all sources, including dataset(s). These projects differ in terms of the tools the students use, the sophistication of the analyses they integrate, and the level of artistic expression. But, we have been stunned by the level of competence and professionalism they display.

We attribute a good part of this success to the design review process we use. All four projects are reviewed in class. The instructor and the other students critique each project, and the students have an additional week to submit their final versions. These interactive discussions reinforce good visualization and design thinking, and provide a vehicle for students to articulate and communicate the principles they are learning.

5 Visualization Guidance

5.1 Overarching Guidance

Our overarching guidance is that the students should focus on the task and the exploration process, and that the whole context be considered. Visualization, whether in the form of a static graphic or an interactive application, needs to be effective on a number of levels, including:

- Practical. How do I make the graphic/application? What tools do I need to use to do it?
- Analytical. Given this data/these data types, which visualization methods are available to me? What possibilities for transformation exist?
- Abstract. Does the the graphic/application cohere as both an airtight visual design and as a mode of communication?
- Critical. What are the implications of the graphic/application? What assumptions does it make? What does it communicate? Is it a faithful representation?
- Visual/Aesthetic. Does the graphic/application look good? Is it readable?

5.2 Visualization Guidelines

Here we introduce some guidelines we integrate into the visualization curriculum.

- **Representation**
  - Hue is a great marker for semantically calling out data. Use hue to highlight a region, segment regions, draw attention, or for brushing.
  - Be sure to have good luminance contrast for text, labels, and data points.
  - Know the data type for each variable and use appropriate representations and colors
  - Divergent color scales are excellent to show variations above and below a zero in the data, or a user-defined set point; sequential color scales. which vary monotonically in luminance, are excellent for representing monotonic variations in a continuous variable.
  - The user's task plays a dominating role in deciding how to represent data; the decision-making process, the context, and the problem being solved are key factors.
- **Analysis**
  - Patterns in data can provide e a ”hook” for motivating the discovery of important and interesting relationships: outliers, clusters, bimodal histograms, changes in slope, discontinuities, periodicities.
  - Brushing can be a very useful tool for understanding how other variables in the dataset behave before, during, or after the above patterns occur.
- **Storytelling**
  - Creating a storyboard for your visualization application can be a good way to understand the flow of your story, and to identify problems with the narrative flow.
  - There is a slippery slope when representing data. Creating a totally faithful story is good; highlighting and annotating can enhance the communication of the message; but, advocacy can lead to misrepresentation, propaganda and downright lying, so ethical considerations in data representation are important to understand.

6 Conclusion

Data visualization is a diverse field, embodying concepts from many domains, which are applied to every situation where data reaches people. There are many approaches to teaching, from the formal and mathematical to the intuitive and artistic. The fact that all visualization tools are designed to interface with humans, however, put constraints on the set of useful guidelines. A formal mathematical system for representing a design space becomes valuable, for example, when the constraints of that system guide the effective choice of mappings and operations, for the users task and domain. Our goal has been to teach guidelines and heuristics that generalize across a wide range of situations. Our main objective is to give the students a framework for asking questions about the selection of visualization methods and operations. Instead of asking, "what choice must I make here," we want them to ask, "what is the user trying to accomplish, and how do the choices I make depend on the data, the task, and human perception, cognition, and judgment."

This course was designed to give students a feel for their data, and the skills and intuition to explore and communicate the phenomena they contain. This course is currently offered as an elective, but it continues to draw record enrollment based on word-of-mouth from former students. In their critiques of the course, the students routinely tell us that they learned critical thinking skills that they use not only in data visualization, but also in data analysis, and even in technical writing.

REFERENCES

[1] A. Kirk. *Data visualisation: a handbook for data driven design*. Sage, 2016.

[2] B. C. Kwon and B. Lee. A comparative evaluation on online learning approaches using parallel coordinate visualization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 993–997. ACM, 2016.

[3] S. Lee, S.-H. Kim, and B. C. Kwon. Vlat: Development of a visualization literacy assessment test. *IEEE transactions on visualization and computer graphics*, 23(1):551–560, 2017.

[4] T. Munzner. *Visualization analysis and design*. AK Peters/CRC Press, 2014.