Conceptual Curriculum Design for Data Visualization Program

Roberto Pelayo
University of Hawaii at Hilo

Jason Leigh
University of Hawaii at Manoa

Maththew Platz
University of Hawaii at Hilo

Francis Cristobal
University of Hawaii at Hilo

Robert Chi
University of Hawaii at Hilo

We can easily enjoy a finished gourmet dish, but we don't exactly know how the gourmet dish is prepared. Like a gourmet dish, Data Visualization is a gourmet subject. "What is this gourmet subject made of?" and "How is a digital gourmet object created?" are still largely unknown questions for many people. In the information age today, we are experiencing a cloud of data from a variety of sources and in a variety of formats. How educational programming for Data Visualization can be adequately developed to meet such challenges is still one of the widely discussed topics today for both online and in-classroom instruction. This paper describes a newly established concept model of data visualization. The domains, scopes, and strategies of educational programming for Data Visualization are described and discussed.

INTRODUCTION

Like a gourmet dish, Data Visualization is a gourmet subject. "What is this gourmet subject made of?" and "How is a digital gourmet object created?" are still largely unknown questions for many people. In the information age today, we are experiencing a cloud of data from a variety of sources and in a variety of formats, which have been challenging the preparation of data visualization. Consequently, comprehensive educational curricula that could address these challenges appear critical for the new profession of Data Visualization.

In order to meet these challenges, data visualization educational programming must have unique curriculum that distinguishes its graduates. The solution is to educate students in a comparable manner across the domain of Data Visualization. This paper describes a newly established concept model based of data visualization. This model illustrates the domains and sub-domains of educational programming for Data Visualization.

BACKGROUND

Existing educational curricula, both online and in-classroom instruction, for Data Visualization are not comparable: some are digital media oriented while others are statistical data manipulation in nature. Consequently, the knowledge and skills of graduates from these programs are difficult to assess and comparable. Due to this reason, a majority of these programs are limited to two-year community college programs with a narrow study scope for students, which confines the professional advancement and marketability of the graduates. The challenge is to promote academics that develop graduates for prospective employers and to fulfill the criteria and the evolving needs of various industries and educational institutes, and simultaneously compete with other scientific programs such as computer science and information technology.

There are many different definitions for data visualization. Generally speaking, data visualization was defined as a 'method to transform abstract scientific data into intuitive graphics that represents the data' by Wikipedia. Dr. Friendly suggested two main scopes of data visualization: "statistical graphics" and "thematic cartography" (Friendly, 2008). In his article, "Data Visualization: Modern Approaches", Friedman however provided a different point view and highlighted seven subjects of data visualization. They are: Mindmaps, Displaying News, Displaying Data, Display Connections, Displaying Web-sites, Articles & Resources, and Tools and Services. Friedman also provided a 'periodic table' for visualization methods. In this 'periodic table', the six types of visualization were described as Data Visualization, Information Visualization, Concept Visualization, Strategy Visualization, Metaphor Visualization, and Compound Visualization (Friedman, 2007). However, these definitions are difficult to be implemented in academic curriculum because "What is the required knowledge base for students?" and "What are the skill sets required in order to enable graduates to perform data visualization tasks?" could not be addressed.

As a form of "governing knowledge", data visualization has been described as an emerging scientific study that have become a key focus in studies of educational policy (Williamson, 2016) and needs to be mathematically or theoretically described in order to have a better understanding of its domains and scopes of study. Is it important to define a model for Data Visualization? The answer is worth pursuing. A well-defined model will not only reveal the characteristics of Data Visualization, but it will also provide guidance for the development of new, or modify existing, online or in- classroom educational program. By doing so, the data visualization profession will also stand to benefit in three-fold: first, by clarifying the vagueness of the definition of data visualization, secondly, by identifying the scope and content for educational programs, and, thirdly, by defining career opportunities for its graduates.

Since its emergence as a distinct discipline more than two decades ago, Data Visualization has been broadly and narrowly included in many college's curricula. This interdisciplinary field is expanding in response to the increasing demands from the various industries and research institutes for skilled personnel in digital media, data manipulation, special digital effects, data management, and digital movie and simulation visualizations. Thus, the development of dedicated academic programs for data visualization education has become a major discussion topic among many colleges and research institutes.

"How is the scope of a curriculum for data visualization defined?" "What are the core requirements for degree programs in data visualization?" "How to set up standards for accrediting such academic programs?" and "What are the career paths for data visualization graduates?" are the some of the most frequently asked questions.

Comprehension of such a diverse discipline like Data Visualization requires a paradigm shift when attempting to define its domains and attributes. A simple approach would be to tolerate a reasonable amount of imprecision and uncertainty in defining a contemporary model of Data Visualization. The results may not be perfect; nevertheless, a reasonable concept model may lead to a theoretical resolution for the understanding of Data Visualization. With a clear described model, academic programming may also easily adopt the evolving new technologies and trends into the curriculum for applying new data visualization technologies into a broad sphere of the real world reality.

In order to appreciate what data visualization is, it would be beneficial to first take a look at the cellular structures and the quantum attributes of data visualization. By doing so, we will have a better perception for the scopes of data visualization, which will help envision the upcoming challenges and prospective outcomes for the research and education adventures.

Although digital data are from a variety of sources and in a variety of formats, they are comparable at their root level since they are all in binary form. Generally speaking, there are four cellular data groups that data visualization normally deals with. They are:

- Numerical Data
- Theoretical Data
- Textual Data
- Logical Data

These four data groups have five possible quantum attributes in terms of visualization. They are:

- One-dimensional visualization, DV = |x|
- Two-dimensional visualization, DV= (x y), or f (x)
- Three-dimensional visualization, DV = (x y z)
- Four-dimensional visualization, DV = (x y z) + t (t=time)
- Multi-dimensional visualization, DV = f (x) + (x y) + (x y z) + t...)

The tasks for visualization of numerical data are to identify patterns, trends, structures, or movements, and to display the numbers collected from instruments, sensors, or surveys into some visual forms.

The tasks for visualization of theoretical data are to illustrate ideas, techniques, or theories, and to animate them into visible patterns, or simulations.

The tasks for visualization of textual data are to create illustrations, simulations, and movies. Textual data is usually a set of linear manuscripts that convey information. The principal for visualizing textual data is to portray the meanings of manuscripts and create visual effects for such meanings. Those dazzling and astonishing Disney and Hollywood digital movies are the examples that show the popularity and prosperity for the textual data visualization.

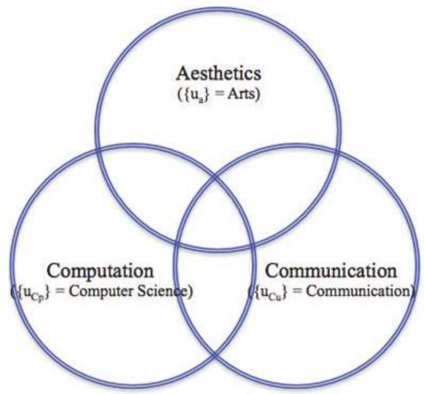
The tasks for visualization of logical data are to discover existing or potential relationships, profiles, or behaviors through software techniques such as 'data mining'. Logical Data is an algorithm that can be used to reveal existing or potential relationships among almost any kind of data. This logical data can be used for profiling images, texts, or behaviors from random collection of data. When a criterion of logical data is executed, multiple sets of relevant information should be generated instantaneously from the pool of data collection.

CONCEPT MODEL OF DATA VISUALIZATION

Data Visualization is a system (DV) that should consist of three major domains: Aesthetics domain, Computation domain, and Communication domain, as shown in Figure 1: i.e. DV∈({uA}{uCp}{uCu})

The processes of teaching and learning activities through some defined computational means will transform knowledge from these three domains into a unique interdisciplinary study. The unique set of skills developed from this study will not only enable students to shape scientific data into artistic display, but it will also make possible to transform complex data into an intuitive information for the purpose of better understanding the nature of data.

FIGURE 1
THREE DOMAINS OF DATA VISUALIZATION

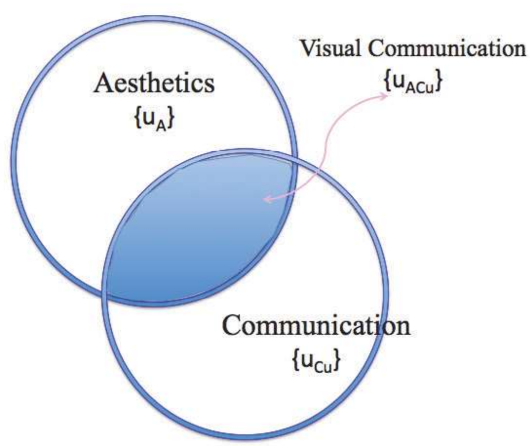


These three domains are the foundation for the study of data visualization, which should be the ingredients for the academic curriculum.

These three domains are independent. Each domain has its own distinct discipline and profession, such as Aesthetic {uA} is the study of arts, Computation {uCp} refers the study of computer science, and the Communication {uCu} is a study of its self. These three domains are also interrelated. This interrelationship structures the unique characteristics of Data Visualization, i.e. DV∈ ({uA}{uCp}{uCu}).

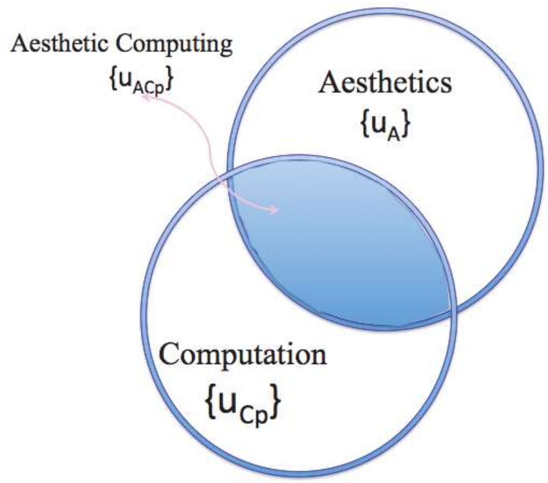
When the Aesthetics domain intersects the Computation domain, the Aesthetic Computing sub-domain is formed, i.e. ({uA}∩{uCp}→{uACp) as shown in Figure 2. This sub-domain reveals the first unique content for the study of Data Visualization, which is to apply art theory and practice into computing and transform scientific data into beautiful and attractive art displays. Samples of learning objectives are: "Student will be able to create digital arts, illustrations, models, and artistically orientated animations or simulations."

FIGURE 2
AESTHETIC COMPUTING SUB-DOMAIN



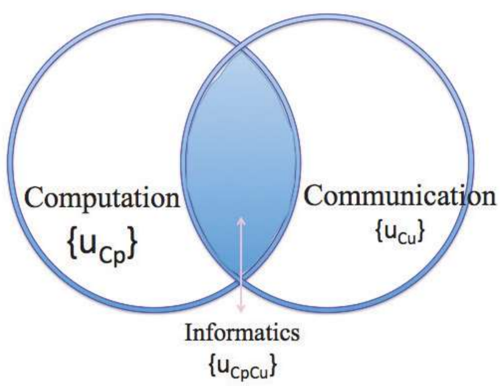
When the Aesthetics domain intersects the Communication domain, the Visual Communication sub-domain is formed, i.e. ({uA}∩{uCu}→{uACu}) as shown in Figure 3. This sub-domain discloses the second unique scope for the study of Data Visualization, which is to apply communication theories through digital means to convert scientific, theoretical, textual, and logical data into easily understood information. Samples of learning objectives are: "Student will be able to create digital illustrations, movies, and digital marketing materials."

FIGURE 3
VISUAL COMMUNICATION SUB-DOMAIN



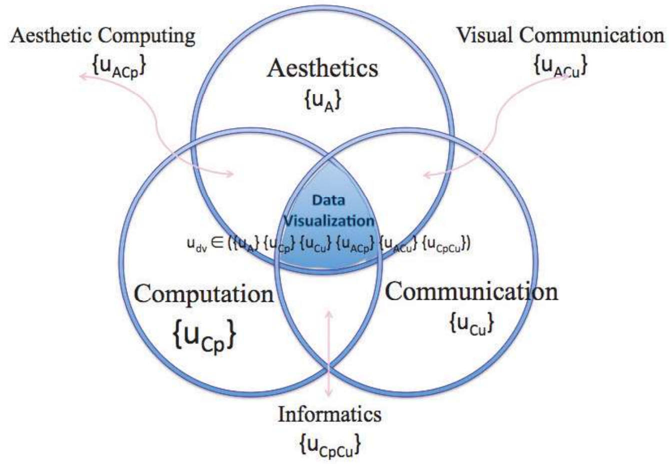
When the Computation domain intersects the Communication domain, the Informatics sub-domain is formed, i.e. ({uCp}∩{uCu}→{uCpCu}) as shown in Figure 4. This sub-domain reveals the third scope for the study of Data Visualization, which is to apply computation and communication theories into the methodologies for data storage, retrieval, and management, and optimize the acquisition and use of information.

FIGURE 4
INFORMATICS SUB-DOMAIN



The interconnections among the above three domains and three sub-domains make up the core for the study of Data Visualization: i.e. DV∈ ({uA}∩{uCp}∩{uCu}∩{uACp}∩{uACu}∩{uCpCu}) as shown in Figure 5.

FIGURE 5
THE CONCEPT MODEL OF DATA VISUALIZATION



This concept model not only illustrates the didactic curriculum scopes for Data Visualization, it also shows the unique knowledge base and skill sets for the graduates.

CURRICULUM

Although this Venn diagram in Figure 5 only illustrates the equal balanced knowledge and skill sets for the graduates, actually, this model is a flexible curriculum model that allows students to select subjects or courses of study and build their own combinations of knowledge and skill sets, which will help students develop unique leadership in a specific area of Data Visualization. They could become a Digital Artist, a Digital Movie Producer, a Scientific Data Modeler, a Visual Marketing Specialist, or a Digital Simulation Expert.

Applying F-expert system (if-then) rules, a well-known concept in the field of knowledge-based systems, to this concept model, the proportion of the knowledge from the three domains of Data Visualization could be skewed to construct different scenarios for each student by selecting different courses among the academic curriculum. For example, if a student explores the Digital Artist, this student could select more courses in Atheistic Computing sub-domain and more training in digital arts and digital media, which will help student to be specialized as a digital artist.

Following is a curriculum example: in order to develop a Masters' degree program, the curriculum should have at last 36 credit hours. The program could have the four following content areas:

Content Areas	*Credit*
1. Fundamentals of Data Visualization	3
2. Aesthetic Computing	10
3. Visual Communication	10
4. Informatics	10
Master's Terminal Project Total	3
	36 credit hours

The above content areas are examples of an ideal systematic designed academic curriculum, which could become the standard for programming assessment and accreditation.

Since the concept model of Data Visualization is a flexible model, the coursework layout should allow students to select their interested content areas and skew their knowledge and skill sets toward different leadership positions. This flexibility may post some challenges for coursework offered. An ideal solution for this challenge would be to offer more courses.

The coursework for Fundamentals of Data Visualization area is a uniform knowledge base that each student must take in order to have a comprehensive understanding of the theories, pedagogies, and research methods in Data Visualization. Samples of the coursework are:

- Introduction to Data Visualization (1 credit)
- Research Design and Method in Data Visualization (2 credits)
- Advanced Human Psychology (Optional 2 credits)

The coursework for Aesthetic Computing area is a selective knowledge base, from which students can choose and build up their unique expertise and comprehensive understanding of digital art theories, principles and acquisition of practical skills of digital art and media relevant to the situations for education and digital media industry. Samples of the coursework are:

- Introduction to Aesthetic Computing (1credit)
- Digital Art Design and Illustration (3 credit)
- Digital Graphics (3 credits)
- Digital Movie and Animation (3 credits)
- Technology (3 credits)
- Computer Instructional Design (3 credits)

The coursework for Visual Communication area is also a selective knowledge base, from which students can choose and build up their unique expertise and comprehensive understanding of data and communication theories and acquire practical skills through illustration designing, modeling, rendering, and methodologies of visual communication. Samples of coursework are:

- Introduction to Visual Communication (1 credit)
- Communication Theories in Data Communication (3 credits)
- Data Visualization Methods (3 credits)
- Interface Design for Visual Communication (3 credits)
- Network and Communication Technology (3 credits)
- Multidimensional Projection System (3 credits)

The coursework for Informatics area is also a selective knowledge base, from which students can choose and build up their unique expertise and comprehensive understanding how to apply computational and communicational theories into the methodologies for data storage, retrieval, and management, and optimize the acquisition and use of information, through which they acquire practical skills accordingly. Samples of coursework are:

- Introduction to Informatics (1 credit)
- Data Governance and Management (3 credits)
- Copyrights, Fair Use Guidelines, and Related Issues (3 credits)
- Methodologies of Data Mining (3 credits)
- Artificial Intelligence (3 credits)
- Practical Training Projects in Data Visualization (3 credits)

The master's terminal project is a 3-credit coursework. Students will be able to demonstrate their independent hands- on skill for design and development of a completed data visualization project, which could be an outstanding digital art and illustration, an excellent modeling, animation, or a digital movie, and/or a written thesis in the area of Data Visualization.

ONLINE DELIVERY CHALLENGES

The major challenge for delivering courses of data visualization online is how to teach and facilitate student's hands- on learning activities. For most of the lectures, however, narrated PowerPoint or Podcast (asynchronous) is the simplest and most effective strategies, even though the interactive component is lacking. Why? Because Data Visualization is an application study that often answers "How to" questions, which could be carefully programmed throughout narrated PowerPoint lectures. For example, in order for students to understand how to create a 'Theme' for a simulation, the color theory could be delivered by giving examples of color contrast between background and foreground with instructor's voice and illustrations.

Screen cast (asynchronous) is a very effective strategy for showing student how to use computer applications, such as SAGE2, AVS Express, Photoshop, iMovie, Cinema 4D, etc. This strategy not only provides a step-by-step instruction for the topics, but it also allows student to review the instructions and demonstrations 24 hour a day and seven days a week, which supports the pedagogy of learning reinforcement with hands-on activities.

Of course, the best online learning strategy is to provide students with interactive tools that allows instructor and students to work on the same subject at the same time (synchronized), such as using Cyber-CANEO, Adobe Connect®, Elluminate®, etc. Most of the synchronized delivery tools also provide recording option. Recording synchronized learning, such as a screen cast lecture, is highly recommended, which will provide student with a flexible learning modality.

Both synchronized and asynchronous learning modalities for online instruction are often hindered by the speed of students' Internet connection. The Hybrid learning strategy appears useful and effective, especially to the 'traditional students', for them the face-to-face instruction is necessary in comparing to the 'non-traditional students' who most likely have more independent learning experiences than those 'traditional students'.

With streaming technology, a student can see and hear the instructor as if in a live classroom. The synchronous modality that parallels asynchronous delivery system will enrich the student learning experience. With the unknown conditions of student's Internet connection speed and the configuration of computer hardware, the Hybrid learning strategy would accommodate both synchronous and asynchronous modalities of the delivery system for the Data Visualization courses and ensure the quality of the learning.

The subject nature of Data Visualization features its learning assessment strategy as project oriented. The learning objectives should attribute this feature and provide more performance oriented behavioral assessment, such as student will be able 'to do', 'to demonstrate', 'to create', 'to design', 'to configure', 'to develop', etc.

One of the challenges for the online learning programming is student's retention and education quality assurance. The program should complement its academic standards with the accreditation requirements. One of the effective programming strategies is the cohort strategy. Like a medical school, students are grouped into a cohort toward a year of graduation. With overlapping of academic courses offered throughout curriculum, students will be able to complete their subject study and training in the scheduled time frame in a collaborative way.

CONCLUSION

Based on this newly established concept model, Data Visualization as a scientific discipline that studies the integration of digital and information technology across all aspects of research, media industries, and movie enterprises can now be logically defined as the study of methodologies to transform data, either numerical, theoretical, textual, or logical data into visual oriented information that carries the meaning of such data.

Colleges and universities, media and marketing industries, and movie enterprises are passionately seeking individuals to fill the roles of digital media and special effect experts, animators, data managers, and data visualization experts who possess both knowledge and skills in digital arts, computational modeling, and visual communication. Well-designed data visualization education programs are the portal for producing the prospective candidates for today's and tomorrow's data visualization experts, digital artists, media specialists, and digital movie developers.

What are the successful factors for developing such an exciting educational program? Faculty development, administrative support, and a lot of creativity and hard work are the essence for the development of excellent programs for the gourmet subject of Data Visualization.

ACKNOWLEDGEMENT

The work in this manuscript was supported by the Hawai`i EPSCoR 'Ike Wai grant, OIA #1557349.

REFERENCES

B. Williamson (2016). Digital Education governance: Data Visualization, Predictive Analytics, and "Real-time" Policy instruments. Journal of Education Policy, Volume 31, No.2, 123 – 141.

D. Bergeron, W. Cody, W. Hibbard, D. Kao, K. Miceli, L. Treinish, and S. Walther (1993). Database Issues for Data Visualization: Developing a Data Model. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 3 – 15.

D. Hand, H. Mannila, P. Smyth (2001). Principles of Data Mining. MIT Press, Cambridge, MA. ISBN 0-262- 08290-X.

D. Kao, R. Bergeron, and T. Sparr (1993). An Extended Schema Model for Scientific Data. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 69 – 82.

D. Keim, and H. Kriegel (1993). Using Visualization to Support Data Mining of Large Existing Databases. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 210 – 229.

J. Boyle, S. Eick, M. Hemmje, D. Keim, J. Lee, and E. Sumner, Jr. (1993). Database Issues for Data Visualization: Interaction, User Interfaces, and Presentation. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 25 - 34.

J. P. Lee and G. Grinstein (1994). Database Issues for Data Visualization: IEEE Visualization '93 Workshop, San Diego.

K. Rya (1993). Data Integration for Visualization Systems. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 83 – 93.

M. Arya, N. Grady, G. Grinstein, P. Kochevar, D. Swanberg, V. Vasudevan, L.Wanger, A. Wierse, and M. Woyn (1993). Database Issues for Data Visualization: System Integration Issues. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 16 - 24.

M. Friendly (2008). Milestones In The History Of Thematic Cartography, Statistical Graphics, And Data Visualization. Journal of Computational and Graphical Statistics, Volume 16, Number 2, Pages 421-444.

M. Hemmje (1993). A 3D Based User Interface for Information Retrieval Systems. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 194 – 209.

M. Woyna, D. Zawada, K. Simunich, and J. Christiansen (1993). A Hyperspectral Image Data Exploration Workbench for Environmental Science Applications. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 160 – 169.

N. A. Gorban, B. Kégl, and A. Zinovyev (2007). Principal Manifolds for Data Visualization and Dimension Reduction. LNCSE 58. Springer.

P. Kochevar (1993). Database Management for Data Visualization. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 109 – 117.

S. Walther (1993). Inherent Logical Structure of Computational Data: Its Role in Storage and Retrieval Strategies to Support User Queries. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 94 – 106.

T. Fenwick, E. Mangez & J. Ozga (2013). Governing Knowledge: Comparison, Knowledge-Based Technologies and Expertise in the Regulation of Education. World Yearbook of Education 2014. London: Routledge.

V. Friedman (2007). Data Visualization: Modern Approaches. Smashing Magazine (www.smashingmagazine.com), an online magazine dedicated to designers and developers. August 2nd, 2007.

V. Vasudevan (1993). Database Requirements for Supporting End-User Visualizations. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 138 – 147.

W. Hibbard, C. Dyer, and B. Paul (1993). The VIS-AD Data Model: Integrating Metadata and Polymorphic Display with a Scientific Programming Language. Proceedings of the IEEE Visualization '93 Workshop on Database Issues for Data Visualization, Pages: 37 – 68.