Rethinking Visualization: A High-Level Taxonomy

Melanie Tory Torsten Möller

Graphics, Usability, and Visualization Lab
School of Computing Science, Simon Fraser University

ABSTRACT

We present a novel high-level visualization taxonomy. Our taxonomy classifies visualization *algorithms* rather than *data*. Algorithms are categorized based on the assumptions they make about the data being visualized; we call this set of assumptions the *design model*. Because our taxonomy is based on design models, it is more flexible than existing taxonomies and considers the user's conceptual model, emphasizing the human aspect of visualization. Design models are classified according to whether they are discrete or continuous and by how much the algorithm designer chooses display attributes such as spatialization, timing, colour, and transparency. This novel approach provides an alternative view of the visualization field that helps explain how traditional divisions (e.g., information and scientific visualization) relate and overlap, and that may inspire research ideas in hybrid visualization areas.

ACM categories: H.5.m Information Interfaces and Presentation (Miscellaneous), H.1.1 Systems and Information Theory

Keywords: visualization, taxonomy, classification, design model, user model, conceptual model

1. INTRODUCTION

Visualization was historically categorized into two major areas: "scientific visualization" and "information visualization". Card, Mackinlay, and Shneiderman [4] define visualization as "....the use of computer-supported, interactive, visual representations of data to amplify cognition....". While this definition and objective of the field as a whole are generally agreed upon, the subcategories have been differentiated in many different ways, primarily:

- Whether the application area is scientific (scientific visualization) or non-scientific (information visualization) [4].
- Whether the data is physically based (scientific visualization) or abstract (information visualization) [4].
- Whether the spatialization is given (scientific visualization) or chosen (information visualization) [Tamara Munzner's statement in [12]].

In other words, "scientific visualization" typically involves scientific data with an inherent spatial component (e.g., wind tunnel vector data or three-dimensional (3D) medical images), and "information visualization" typically involves abstract, non-spatial data (e.g., financial data or document collections).

The various definitions of scientific and information visualization often contradict each other or contradict visualization experts' intuition about what belongs in each category. Because of this ambiguity, it can be difficult to decide where some data sets and application areas belong. For example, abstract mathematical functions (e.g., $f(x,y,z,w) = x^2 + y^2 + z^2 + w^2$) are scientific (meaning that they belong to "scientific visualization") but often the spatialization is not given and the equations are not physically based (so they also belong to "information visualization"). Similarly, air traffic control systems are physically based and have a given spatialization, but they are not necessarily scientific. If we agree that visualization is important in these domains, is it "information visualization" or "scientific visualization"? A clear understanding of how these areas differ is needed.

Furthermore, the classifications of scientific and information visualization force a division in the visualization field. Although this separation has practical utility, it also has the downfall that research bridging the two fields may be discouraged. Furthermore, other interesting areas may be left out altogether. In this paper, we present a taxonomy that provides a new perspective of the visualization field as a whole. Our taxonomy illuminates the fundamental differences between scientific and information visualization and helps us understand what areas overlap and what areas may be missing. We hope this taxonomy will encourage research in hybrid and novel areas by allowing researchers in various fields to see their common interests.

Although problems with the "information visualization" and "scientific visualization" terms may seem unimportant at first glance, we believe carefully constructed definitions are crucial to future visualization research. Visualization taxonomies can serve two major purposes:

1. Guide users. People outside the visualization community may have trouble finding visualization ideas in the literature if they are not categorized in a meaningful way. To achieve this goal, our taxonomy could be used as the basis for a new literature classification scheme.
2. Guide research. Researchers need to know where their research fits into a larger context and find people doing similar work. Also, research can sometimes become more focussed or progress more rapidly when we increase our appreciation and comprehension for the field as a whole. In this context, a more meaningful organization of current research may help us identify areas for future investigation. We believe our taxonomy will motivate researchers to think of visualization in a different way and therefore generate novel research ideas and discussion.

1.1 Overview and Objectives

In this paper, we temporarily set aside the terms "information visualization" and "scientific visualization" and the historical baggage they carry. We ask the reader to follow us in this thought experiment and think about the visualization field as a whole and how it might be categorized. Later in the paper, we use our taxonomy to illustrate how the two traditional areas relate and overlap.

We propose a classification scheme that organizes visualization techniques in a new way. The proposed new taxonomy is based on characteristics of models of the data rather than on characteristics

e-mail: {mktory, torsten}@cs.sfu.ca

IEEE Symposium on Information Visualization 2004
October 10-12, Austin, Texas, USA

of the data itself; therefore, we call it a model-based visualization taxonomy.

Our taxonomy provides a fresh perspective on visualization as a whole. We hope this perspective will stimulate new research ideas, particularly in hybrid and novel areas. We intend our taxonomy to inspire ideas and generate discussion on the topic.

1.2 Outline

Related research on classification schemes for visualization is presented in Section 2. In Section 3 we explain and motivate our model-based design. The taxonomy itself is presented in Section 4. Finally, we describe our conclusions and future work.

2. Related Work

Most visualization taxonomies are based on the type of data involved. Data has many characteristics, including [4, 19]:

- Number of independent variables.
- Number of dependent variables.
- Type of each variable:
  - Scalar, vector (2D, 3D, or nD), tensor, or more complicated structure.
  - Discrete (e.g., number of people in a city) or continuous (e.g., population density across a metropolitan area). In many cases, data is discrete but sampled from a continuous source (e.g., medical images); this data may be considered continuous since we can interpolate between sampled points to guess at the values in between.
  - Nominal (values have no natural order), ordinal (values have a natural order), interval (there is a meaningful distance metric between any two values, so arithmetic can be done on them), or ratio (an interval variable with a meaningful zero). For example: (1) The set {cherry, apple, banana} has no natural order and is nominal. (2) The sets {small, medium, large} and the alphabetically ordered {apple, banana, cherry} are ordinal. (3) Temperature is commonly measured using interval scales such as Celsius and Fahrenheit. (4) The Kelvin scale for temperature is a ratio scale because it has an absolute zero (i.e. 0 K is the lowest possible temperature).

Tweedie [17] describes several forms of data that may be represented visually: data values (described above), data structure (file hierarchies, rectilinear vs. curvilinear grids, etc.), and meta-data. Meta-data is derived from data values or structure (it is data about data). Derived data can consist of derived values and/or derived structure.

"Scientific visualization" is typically categorized by dimensionality of the data values (number of independent variables), and whether the data is scalar, vector, tensor, or multivariate (having more than one dependent variable). Examples of this basic classification scheme may be found in Brodlie *et al.* [2, pp. 40-43] and Schroeder *et al*. [15].

"Information visualization" can be similarly organized by data type. Common categories are multi-dimensional databases (often containing more than three dimensions), text, graphs, and trees [8, 16]. In addition to data type, some taxonomies have organized visualization systems by display style (e.g., table, information landscape, etc.) [3, 5] or include generic tasks performed by users of the system (e.g., gaining an overview, drilling down on details, filtering, etc.) [7, p. 38, 16].

Although we agree in general with the structure provided by these organizational schemes, we disagree with two components. First, we believe the high-level division into "scientific visualization" and "information visualization" could be more clearly defined. Second, we consider the dependence of existing classification schemes on data type to be problematic since data undergoes significant interpretation during the visualization process, by both programmers / designers of visualization algorithms and end users. Our proposed taxonomy attempts to address these problems while still maintaining as much of the common organizational schemes as possible.

Another area of related research is visualization design. In this area, the objective is to match data for specific applications to the most appropriate visualization techniques. Bertin [1] provides a thorough analysis of the match between data characteristics, graphic variables, and human perception. Mackinlay [9] describes an automated method based on similar ideas, in which the system chooses the "optimal" visualization techniques. Similarly, Wilkinson [21] describes a set of grammatical rules for defining graphics. Other visualization design systems address usability issues in more detail by including user objectives (and sometimes user input) in the design process. Robertson [13] includes the context of use and allows users to choose from (and modify) several "optimal" displays. Similarly, Zhang [22] describes a taxonomy to map data variables to display dimensions while considering the user's task. Espinosa *et al*. [6] describe a methodology for designers to consider user needs when developing visualization systems. Wehrend and Lewis [20] develop a "catalog" for users to easily look up and share visualization techniques for specified data types and tasks. Finally, users of SageBrush [14] can provide design directives and sketched prototypes to partially specify a graphic representation that will then be created by the SAGE automated system.

Visualization design requires a taxonomy of techniques to guide automated algorithms, programmers, or users. These taxonomies are similar to ours since they consider more than just data characteristics (e.g., most also consider user interests). Nonetheless, data type still plays a major role in these categorizations, whereas our taxonomy reduces this dependence. In addition, many visualization design taxonomies include only a small subset of techniques (e.g., 2D graphs [9]). Often these subsets fall in the "information visualization domain", neglecting the "scientific visualization" area. Our taxonomy provides a higher-level view of visualization and includes a wide range of techniques. Furthermore, we focus on visualization as a research field; hence, our main objective is to provide insight into how different research areas relate, not to provide guidelines for visualization design.

3. Model-Based Taxonomy Design

Before defining the new taxonomy, we define some terminology (see also Fig. 1):

Object of study: An object of study is "something mental or physical toward which thought, feeling, or action is directed" [10]. In visual data analysis, the object of study is the idea or physical object being investigated.

Data: Because the object of study cannot usually be studied directly, it is typically analyzed indirectly through a set of discrete samples, the data.

Design model: Visualization designers make assumptions about the data, and build those assumptions into the visualization algorithm they are designing. For example, a designer might assume that data is ordinal, that data can be interpolated, or that triplets of consecutive numbers represent 3D spatial directions. This information is often not directly represented by data values. We call this set of assumptions the design model. This definition is based on Norman's design model [11] (i.e., the conceptualization of a system that the designer has in mind); however, we refer to assumptions about data whereas Norman refers to assumptions about a system's function.

User model: Users have pre-conceived ideas about the object of study and interpretations of the data that affect their understanding of what the data represents. This conceptual model of the object of study is what determines which visualization techniques are chosen by the viewer. Interacting with a visualization can help users to refine and update this conceptual model. We call the user's set of assumptions about the data the user model. Like the design model, the user model is also based on a definition by Norman [11].

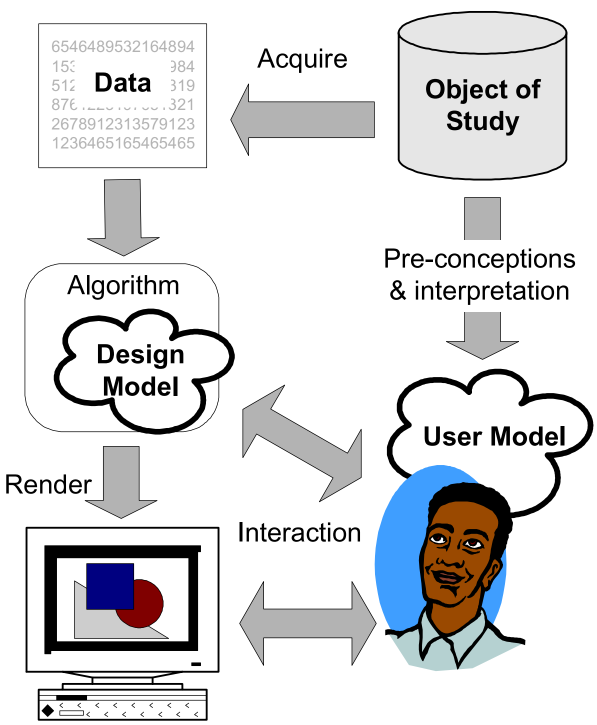

Figure 1: Relationships between object of study, data, visualization algorithm, design model, and user model.

3.1 Example

Consider a group of physicians viewing medical images to make a diagnosis of brain cancer. The *object of study* is the patient who has shown worrisome symptoms. Since the physicians cannot directly see inside the patient's brain (without traumatic surgery), they order an imaging scan such as computed tomography (CT) or Magnetic Resonance Imaging (MRI). After the scanner has acquired images of the patient's brain, it stores them digitally as a set of discrete numbers (the *data*).

Physicians can then visualize the data in a number of ways. How physicians think about the data (the *user model*) will determine which *visualization algorithms* they choose. Furthermore, while diagnosing the patient and explaining the diagnosis to other medical professionals, physicians may use several different visualization algorithms. For instance, they may use the original set of 2D slices to determine that a tumour is present, and then explain the position of the tumour to a surgeon using a 3D visualization. Designers of each visualization technique make assumptions about the data that will be visualized, and build these assumptions into the algorithm (e.g., a common assumption for medical imaging data is that interpolating the discrete data values would be meaningful). These assumptions are encoded in the algorithm but not the image pixel values themselves. These assumptions comprise the *design model*.

3.2 Object of Study

Because the object of study is the idea being investigated, it varies depending on users and their interests. For instance, primary care physicians in the example above may study a particular patient, whereas research physicians may study an illness. Furthermore, the object of study can be an abstract idea rather than an object in the real world. For example, an automobile executive may be interested only in marketing success for the 2004 Mazda Tribute, making sales for that specific vehicle the object of study. Alternatively, the executive may be interested in how sales of the Tribute compared to sales of other sport utility vehicles (SUVs). Here sales of all SUVs is the object of study, and Tribute sales may be a sub-object of special interest.

3.3 User and Design Models

The difference between the data and the models is best illustrated by example. As previously mentioned, although digital image data is discrete, the user model (i.e. the user's expectations and understanding) of medical images is usually continuous since the data was sampled from a continuous source. For this reason, users looking at medical image data often choose visualization algorithms with a continuous design model (e.g., direct volume rendering algorithms). These algorithms interpolate between data points, so the relationship between data points is represented as a continuous function by the design model (within the visualization algorithm).

User models and design models are closely linked. When users visualize their data, they select display techniques whose design model matches (or is expected to match) their own user model. A radiologist who expects a dataset to be a medical image would display the data using an image viewer (whose design model assumes the data represents a 2D array of grayscale values). On the other hand, an engineer building an MRI scanner might look at the output with a text viewer (whose design model assumes the data represents a linear array of characters) in order to carefully debug the machine for possible failures. The difference between design models and user models is that design models are explicitly encoded by designers into visualization algorithms whereas user models are in the mind of the user and are therefore more difficult to explicitly describe. To avoid the ambiguity and changeability of user models, we base our classification on design models.

In the medical image example, most users likely have similar conceptual models of the data. This may not always be the case. A single data set may be interpreted quite differently by different people, or by the same person at different times, affecting the type of visualization chosen. For example, the set of all buildings in a city may be thought of as a list of building names, populations, and coordinates (visualized as discrete points on a map) or as a spatial distribution of population density (visualized as a map with a continuous colour scale). Similarly, a set of documents may be thought of as unconnected (visualized in table form or organized by a measure of document content) or structured by references or hyperlinks (and therefore visualized as a graph). Since a single data set can be conceptualized in several different ways, it makes

Table 1: High-level visualization taxonomy, illustrated by examples. Design models are classified based on whether they are discrete or continuous and by how much the algorithm designer chooses display attributes (spatialization, timing, colour, and transparency). Examples show different constraints on spatialization.

	Display Attributes		
	Given	**Constrained**	**Chosen**
Continuous	Images (e.g., medical) Fluid / gas flow, pressure distributions Molecular structures (distributions of mass, charge, etc.) Globe – distribution data (e.g., elevation levels)	Distortions of given / continuous ideas (e.g., flattened medical structures, 2D geographic maps, fish-eye lens views) Arrangement of numeric variable values	Continuous (high-dimensional) mathematical functions Continuous time-varying data, when time is mapped to a spatial dimension Regression analyses
Discrete	Classified data / images (e.g., segmented medical images) Air traffic positions Molecular structures (exact positions of components) Globe – discrete entity data (e.g., city locations)	Distortions of given / discrete ideas (e.g., 2D geographic maps, fish-eye lens views) Arrangement of ordinal or numeric variable values	Discrete time-varying data, when time is mapped to a spatial dimension Arbitrary entity-relationship data (e.g., file structures) Arbitrary multi-dimensional data (e.g., employment statistics)

more sense to classify visualization techniques based on the design models they use, rather than on the data itself.

3.4 Constructing User Models

Users sometimes have clear preconceived ideas about the object of study. For example, a weather forecaster who views the same types of data day after day probably has specific expectations about the data structure and may even have initial expectations about the data values based on the previous day's weather.

In other cases, user models may be sketchy because little is known about the object of study. Constructing a user model is a complex process that may include making assumptions about the data and the display algorithm, developing hypotheses, searching for evidence to support or contradict hypotheses, and refining the model [18]. User models are developed and refined by interacting with data via visualization tools. Interactions are guided by questions/hypotheses posed by users as they go through the process. Users iteratively utilize their conceptual models to ask questions and choose visualization techniques and then refine their models as more information becomes available. For more details about conceptual model development in general, see [18].

4. Proposed Taxonomy

We categorize visualization techniques based on their design model. This approach differs from traditional classifications based on data type. To our knowledge, design models have not been used as the basis of a visualization taxonomy. Furthermore, user models are closely related to design models because users will choose visualization techniques that match their ideas and intentions; thus, our taxonomy emphasizes the human side of visualization. Hence, characterizing the visualization field based on models rather than the data itself is our first main contribution.

4.1 High Level Taxonomy Structure

The high-level structure of our proposed taxonomy is outlined in Table 1. We classify design models according to two criteria:

1. Whether the model assumes the object of study is discrete or continuous.
2. How much the visualization designer chooses display attributes (spatialization, timing, colour, etc.).

4.1.1 Discrete / Continuous Classification

Visualization algorithms are classified according to whether their design model is discrete or continuous (see the rows in Table 1). We believe this provides a simple, clear division that puts similar techniques together. Although classification of data based on whether it is continuous or discrete is well known, we offer two novel ideas: (1) we use the continuous/discrete division at the top-level of a taxonomy of visualization techniques, and (2) we characterize design models (not data) as continuous or discrete. This division based on continuous/discrete design models is the second major contribution of our paper.

Continuous models assume that data can be interpolated, whereas discrete models assume data cannot be interpolated. Interval and ratio data can be interpolated, but users could choose not to do so; thus, interval and ratio data can be visualized by either continuous or discrete model techniques as desired by the user. Nominal and ordinal data can often only be visualized by discrete model techniques since interpolation is not meaningful. For example, there is no meaningful value between male and female or between apple and banana.

Converting from a continuous model to a discrete model is a matter of leaving data points as discrete entities (not interpolating), sampling a continuous function, or aggregating data points into bins or categories. The reverse process, converting from a discrete model to a continuous one, requires parameterizing the model or embedding it into a continuous space. For instance, the list of buildings and their populations (mentioned in Section 3.3) can be embedded in physical space by using the buildings' locations. The population can then be thought of as a distribution over that continuous space.

4.1.2 Display Constraints

A traditional way of categorizing visualizations is whether the spatialization is given or chosen [Tamara Munzner's statement in [12]]. We believe this division is meaningful but incomplete because it neglects the idea that spatialization is sometimes partially given and partially chosen; that is, it is constrained. The level of such constraints falls along a continuum from completely given to completely chosen. For example, spatialization is mostly given for geographic data shown on a globe; however, several choices have to be made when the data is displayed on a flat map (e.g., where to cut the world open and whether to display lines of longitude in parallel).

Furthermore, other display attributes (besides spatialization) can also have varied levels of constraints. For a visual display, these other attributes are colour, transparency, and time. Multimedia displays could include additional attributes such as smell and haptic effects.

We address these two issues and incorporate the concept of constraints on display attributes into our taxonomy (see columns in Table 1; notice that the columns are not divided by lines, indicating a continuum). These ideas are the third major contribution of our paper. Note that we agree spatialization may be the most important display attribute in terms of differences between visualization categories. We therefore focus on space; nevertheless, our taxonomy easily extends to all display attributes mentioned in the previous paragraph.

4.2 Relationship to Information and Scientific Visualization

Observing Table 1, we noticed that scientific visualization tends to occupy the top left area and information visualization tends to occupy the bottom right. Middle areas are ambiguous and belong to both categories or neither. We believe the fact that our taxonomy places scientific and information visualization at opposite extremes of the table with ambiguous areas in between confirms the taxonomy's meaningfulness. More importantly, the categorization helps us see where the two fields overlap and how they relate. This idea extends to other traditional visualization categories (e.g., math visualization tends to occupy the top right).

Notice that we avoid classifying data based on whether it is scientific or physically based. We therefore avoid the problems that these criteria created, namely that some scientific data were visualized using "information visualization" techniques (e.g., bioinformatics data), and some data visualized with "scientific visualization" techniques were not physically based (e.g., mathematical functions). Hence, we believe our new taxonomy is less ambiguous and better describes the visualization techniques in each new major category. We encourage researchers to set aside the intrinsic meaning of the terms "scientific" and "information" visualization and use them primarily for their historical value. In this context, our high-level taxonomy can help define these areas and illuminate their differences and similarities.

Furthermore, our taxonomy illustrates similarities between fields that appear quite different on the surface (e.g., air traffic control and molecular structure visualization). This may encourage discussion between researchers in different application areas, leading to research advances.

4.3 Low Level Taxonomy Structure

At lower levels of the hierarchy, we classify continuous and discrete design models in similar ways to previous taxonomies, using the number and type of variables in the design model and whether the design model consists of structure or values. We believe these schemes do a reasonable job of characterizing each subarea. Tables 2 and 3 classify continuous and discrete design models respectively, with examples of the types of visualization techniques that could be used for each design model.

Many of the sample visualization techniques in Tables 2 and 3 can be used with either given or chosen display attributes. For example, 3D scatterplots can be used to display airplane positions for air traffic control (a given spatialization) and income broken down by gender and educational level (a chosen spatialization). However, techniques that *assign* display attributes such as spatialization (e.g., parallel coordinates) will only be used when those attributes are not given.

4.3.1 Continuous model visualization

Continuous model visualization can be broken down according to the number of independent and dependent variables, and the type of the dependent variables, as shown in Table 2. This produces three major categories: scalar, vector, and tensor (matrix) visualization, with 1D, 2D, 3D, and nD versions of each. Multivariate visualization addresses design models containing more than one dependent variable. Examples of scalar visualization techniques include line graphs for 1D design models, colour gradients and isolines (contours) for 2D models, and direct volume rendering and isosurfaces for 3D models. Vectors can be visualized using glyphs (arrows that point in the direction of flow), particle traces (where imaginary particles are placed in the flow field and tracked over time to trace a line of motion), and line integral convolution (where a white noise texture is warped along the direction of flow). Tensors can be visualized using ellipsoid-shaped glyphs, where the principle axes of each ellipsoid represent the eigenvectors of the matrix and the principle radii represent the eigenvalues. See [15] for a more detailed review of continuous model visualization techniques.

Table 2: Low-level taxonomy of continuous models.

# Independent Variables	Data Structure: Scalar	Vector	Tensor	Multi-variate
1D	- Line graph			Combine scalar, vector, & tensor methods
2D	- Colour map - Isolines	- LIC - Particle traces - Glyphs		
3D	- Volume rendering - Isosurfaces		- Tensor ellipsoids	
nD	Multiple 1D, 2D, or 3D views			

4.3.2 Discrete Model Visualization

Discrete model visualization is first broken down according to whether data structure (e.g., hyperlinks connecting documents) or data values (e.g., document size and file type) are visualized. This structure/value division is based on a taxonomy by Tweedie [17].

We then categorize "Value" visualization techniques according to the number of dimensions the visualization supports (2D, 3D, or nD). This breakdown is shown in Table 3.

Structural visualization techniques include node/link diagrams, hierarchical approaches, and space-filling mosaics (e.g., tree levels can be represented as consecutive rows in an image with tree items represented by coloured rectangles). Structural data is considered discrete because a structure is composed of nodes and relationships (discrete entities), even if those nodes and relationships are present in a continuous space. For example, a mesh in computer graphics can be displayed as a continuous surface. However, the topology of a mesh consists of discrete points and connections and can be visualized as a node-link diagram. Mesh structure (topology) is typically studied in the field of graph theory, whereas displaying a mesh as a surface in space or parameterizing it (e.g., by embedding it on the surface of a sphere) is studied in computer graphics. Similarly, classification of continuous data could be viewed as a decision tree; however, by doing this, the data is segregated into discrete categories.

2D and 3D data values can be visualized using scatter plots, bar charts, etc. Techniques to visualize higher dimensional data include multiple views, glyphs, parallel coordinates, and lower dimension techniques with the addition of visual attributes such as colour (e.g., bar charts with each bar coloured according to some attribute). For a review of discrete visualization techniques, see [4, 8, 19].

Table 3: Low-level taxonomy of discrete models.

Structure

Graph & Tree Visualizations:

- Node - link diagrams (2D and 3D)	- Hierarchical graphs
	- Space-filling mosaics

Values

Number of Variables	Variable Types	Example Techniques
2D	1 Dep. + 1 Indep. variable	- Scatter plot - Bar chart
3D	1 Dep. + 2 Indep. or vice versa	- 3D scatter plot - 3D bar chart
nD	Any number of Dep. and Indep. variables	- Charts + colour - Multiple views - Glyphs - Parallel coordinates

Like in continuous model visualization, variables can be dependent or independent, and this is reflected in our taxonomy. Notice, however, that users do not necessarily need to know in advance which variables are dependent or independent. In multi-dimensional databases, it can be uncertain which variables are dependent, since dependency relationships might be precisely what the user is trying to discover through visual data analysis. To select a visualization technique, users simply need a hypothesis about the dependencies. This is also true with continuous models, although the situation may be less common in that case.

Discrete models do not require that all variables be discrete. Discrete models can include continuous variables, as long as at least one variable is discrete. For example, axes in scatter plots and parallel coordinate displays can be scaled continuously, but plotted points in scatter plots and lines in parallel coordinate displays are discrete entities. Plotting data "continuously" requires a perceptibly infinite number of data points. Plotting this many lines in a parallel coordinates display would produce a complicated entanglement that would be difficult or impossible to interpret. Finding a continuous embedding and then interpolating between points or lines generates continuous model techniques (e.g., line graphs or colour maps).

4.4 Visualization Tasks

Some tasks users perform are common to many visualization areas, while others differ. Shneiderman [16] incorporated tasks into a visualization taxonomy, but did not consider how tasks vary in different visualization areas. Our taxonomy can be used to illustrate relationships between types of tasks and to examine what types of tasks can be performed with each design model.

Fig. 2 illustrates how different design models enable users to perform different tasks with a visual representation:

- When the spatialization is largely given, spatial relationships such as above/below, right/left, and inside/outside can be studied. Furthermore, spatial regions of interest can be specified, extracted, and/or examined in detail. (Blue region in Fig. 2.)
- Discrete structural models allow analysis of connectivity relationships such as parent/child, linkages, and discrete path planning. For example: What is connected to X? What is the child of Y? What links do you follow to get from A to B most efficiently? (Orange)
- Discrete value models allow pattern analysis. Examples include identifying outliers and clusters of data points. Discrete data points are necessary for these types of relationships to exist. (Magenta)
- Discrete models allow users to study details of discrete items and to filter data sets (i.e. exclude items). (Yellow).
- Continuous models (and discrete value models when the data is ordinal) enable users to study numeric trends, such as increasing / decreasing. (Green)

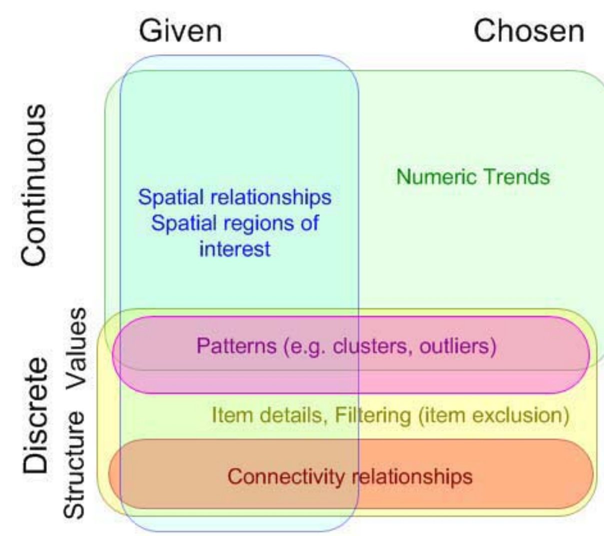

Figure 2: Classification of visualization tasks. The classification is broken down according to how much the spatialization is constrained and whether the design model is continuous or discrete (with or without structure). Colours match figure text to outlined / shaded areas.

4.5 Additional Examples

To help explain our taxonomy, we now give examples from several application domains. These are real world examples obtained by reviewing a large number of recent application and case study papers (e.g., from the IEEE Visualization conference and the IEEE Symposium on Information Visualization).

4.5.1 Temporal Data

Consider a medical record example. A patient visits his or her physician every few weeks, and information is added to the patient's medical record at each visit. This temporal data consists of three major types:

- Quantitative values recorded by measurements at each visit (e.g., cholesterol levels, blood pressure).
- Complaints or events that are noted at specific visits but do not go on continuously (e.g., migraine headaches, immunizations).
- Long-term events that have start and end points but go on for an extended period of time (e.g., persistent pain, drug treatments, depression).

Users visualizing the medical record would likely use techniques from both discrete and continuous sides of the taxonomy. Complaints could be visualized with discrete model techniques (e.g., points on a timeline). Long-term events could be visualized similarly, as bars on a timeline. However, data from individual long-term events (e.g., pain levels over time) could be interpolated and viewed as line graphs. Similarly, derived data such as complaint frequency could be shown with continuous techniques. Line graphs could also visualize ongoing measurements (e.g., cholesterol levels) to see trends, but it is also possible to display individual measurements as points on a scatter plot or heights on a bar chart, to see variations in the actual measured values.

This example highlights some important points. First, not all data sets contain only one data type. More importantly, each type of data in this example could be conceptualized in more than one way (e.g., as either discrete or continuous). How the data is conceptualized determines which visualization technique is most appropriate. This illustrates that the design model is more important than the original data type when choosing and classifying visualization techniques.

Display attributes in this example would be mostly chosen rather than given, but with a few constraints (e.g., time-dependent events would probably be constrained to chronological order). Thus, the example falls on the right side of Table 1 and belongs more to information visualization than scientific visualization. At the same time, we can see how it uses ideas from scientific visualization (e.g., interpolation of continuous variables).

4.5.2 Geographic Information Systems

Geographic Information Systems (GIS) can be used to analyze the spatial distribution of census data such as population, income levels, occupations, etc. GIS data has an inherent spatial component, but this is often distorted during visualization (e.g., to create a flat map or a map that only roughly approximates reality). Thus, the spatialization is constrained, but not entirely given. GIS data such as population could be considered either discrete or continuous, depending on the conceptual model. If we think of the data as a list of cities or locations and their populations, then the data is discrete and can be visualized with methods such as a bar chart or a map with glyphs indicating population. On the other hand, if we think of the data as a continuous spatial distribution, we might visualize it using a map with a colour gradient to indicate population density. As a third alternative, population movement might be conceptualized as a graph showing connections (e.g., roads or flight routes) between cities, with edge weights indicating the amount of traffic flow on each route.

As another example, meteorological visualizations display atmospheric conditions such as temperature, pressure, and wind direction. Continuous models of meteorological data can be visualized using 2D and 3D scalar and vector techniques. Alternatively, weather conditions in major cities can be visualized using icons on a map (a discrete visualization).

Because GIS visualization can be both discrete and continuous and has varying levels of constraints on spatialization, it is near the middle of Table 1 and uses substantial ideas from both scientific and information visualization. This emphasizes the overlap between the two fields and the need to involve ideas from both to create effective visual representations.

4.5.3 Bioinformatics / Cheminformatics

Our first bioinformatics example involves search results from genetic sequence databases. Scientists who find the sequence of a new protein or gene can search databases to find other compounds with similar sequences. Most users would consider genetic sequence data discrete, since each compound in the database is a discrete entity. Beyond this level, how we visualize the data depends on our goals. If we are interested only in finding compounds most similar to the new one, we might create a bar chart showing percent similarity between the target and the top matches. Alternatively, if we are interested in specific overlap areas, we may think of the data as a structure, where overlapping areas between compounds are drawn as connections. In both these cases, the spatialization is largely chosen, although the sequence data would be constrained to its original order.

Another example is protein structure visualization. Protein structure data could be thought of as a set of discrete connected components such as atoms or secondary structures (visualized as a graph), or as a continuous distribution of particles such as protons and electrons (visualized using direct volume rendering or isosurfaces). In both cases, the spatialization is given; however, if the structure was deformed (e.g., by flatting the structure to a 2D visualization or by using a fish-eye lens to highlight a part of interest), then the spatialization would be constrained but partially chosen.

This example illustrates how a particular application area can utilize a broad spectrum of visualization techniques. Bio- and cheminformatics illustrate an awkward division in the traditional taxonomy: visualization of protein structure is usually considered "scientific visualization" whereas visualization of protein sequences is most often considered "information visualization". The division is awkward since a protein's sequence is directly related to its structure. Bioinformatics visualization spans almost all parts of Table 1, illustrating how it uses ideas from both scientific and information visualization, and how these two fields relate. We hope this example will convince the reader that the boundaries between information and scientific visualization are inherently blurred, and that successful work in areas such as bioinformatics will require researchers to consider both fields. We expect that our high-level classification will clarify how the traditional areas relate, making this process easier.

4.6 Limitations and Future Work

We believe our classification scheme provides valuable insight, but is by no means complete. Lower levels of the hierarchy could be described in greater detail, and examples of more techniques could be included. Furthermore, the taxonomy could be applied to compare complete visualization systems.

In addition, the current taxonomy defines "dimensionality" of a design model differently for continuous and discrete models. For continuous models, "3D" means that there are three independent variables and one or more dependent variables. By contrast, "3D" for a discrete model means that there are three dimensions in total (regardless of whether the variables are dependent or independent). Part of the reason for this discrepancy is that users are not always sure which variables are dependent, as described in Section 4.3; this occurs more commonly in discrete model visualization. Future work may allow us to find a better continuity between the definitions of "dimensionality" in discrete and continuous models.

5. Conclusion

We described a high-level taxonomy of the visualization field that provides new insight into how different visualization areas relate and overlap. Our taxonomy is novel because it is based on models of data and categorizes these models based on whether they are continuous or discrete and according to how much they constrain display attributes. Our taxonomy also highlights the role of users and their conceptual models. We believe the taxonomy is a valuable framework for organizing literature and ideas in visualization and will facilitate research in new and hybrid areas.

6. Acknowledgements

We thank Tamara Munzner, Ben Shneiderman, Colin Swindells, and John Dill for valuable discussions, and the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding.

References

[1] Bertin, J., *Semiology of Graphics*, University of Wisconsin Press, Madison, 1983.

[2] Brodlie, K.W., Carpenter, L.A., Earnshaw, R.A., Gallop, J.R., Hubbold, R.J., Mumford, A.M., Osland, C.D., and Quarendon, P., *Scientific Visualization: Techniques and Applications*, Springer-Verlag, Berlin, 1992.

[3] Card, S.K., and Mackinlay, J., "The Structure of the Information Visualization Design Space", *Proc. IEEE Symposium on Information Visualization*, 1997.

[4] Card, S.K., Mackinlay, J.D., and Shneiderman, B., *Readings in Information Visualization: Using Vision to Think*., Morgan Kaufmann Publishers, San Francisco, 1999.

[5] Chi, E.H., "A Taxonomy of Visualization Techniques using the Data State Reference Model", *Proc. IEEE Symposium on Information Visualization*, 2000, pp. 69-75.

[6] Espinosa, O.J., Hendrickson, C., and Garrett, J.H., "Domain Analysis: A Technique to Design A User-Centered Visualization Framework", *Proc. IEEE Symposium on Information Visualization*, 1999, pp. 44-52.

[7] Fayyad, U., Grinstein, G.G., and Wierse, A., *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann Publishers, San Francisco, 2002.

[8] Keim, D.A., "Visual Exploration of Large Data Sets", *Communications of the ACM*, 44, 8, 2001, pp. 38-44.

[9] Mackinlay, J., "Automating the Design of Graphical Presentations of Relational Information", *ACM Transaction on Graphics*, 5, 2, 1986, pp. 110-141.

[10] Merriam-Webster dictionary, http://www.m-w.com/home.htm.

[11] Norman, D.A., *The Design of Everyday Things*, Doubleday, New York, 1990.

[12] Rhyne, T.-M. (Chair), Tory, M., Munzner, T., Ward, M., Johnson, C., Laidlaw, D.H. (Panelists), "Panel: Information and Scientific Visualization: Separate but Equal or Happy Together at Last", *IEEE Visualization*, 2003, pp. 611-614.

[13] Robertson, P.K., "A Methodology for Choosing Data Representations", *IEEE Computer Graphics and Applications*, 11, 3, 1991, pp. 56-67.

[14] Roth, S.F., Kolojejchick, J., Mattis, J., and Goldstein, J., "Interactive graphic design using automatic presentation knowledge", Proc. CHI, 1994, pp. 112-117.

[15] Schroeder, W., Martin, K., and Lorensen, B., *The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*, 2nd ed., Prentice Hall, Upper Saddle River, 1998.

[16] Shneiderman, B., "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations", *Proc. IEEE Symposium on Visual Languages*, 1996, pp. 336-343.

[17] Tweedie, L., "Characterizing Interactive Externalizations", *Proc. Conf. Human Factors in Computing Systems*, 1997, pp. 375-382.

[18] Waern, Y., "On the Dynamics of Mental Models", In *Mental Models and Human-Computer Interaction 1*, Eds. D. Ackermann and M.J. Tauber, Elsevier Science Publishing Company, New York, 1990, pp. 73-93.

[19] Ware, C., *Information Visualization: Perception for Design*, Morgan Kaufmann Publishers, San Francisco, 2000.

[20] Wehrend, S. and Lewis, C., "A Problem-oriented Classification of Visualization Techniques", *Proc. IEEE Visualization*, 1990, pp. 139-143.

[21] Wilkinson, L., *The Grammar of Graphics*, Springer-Verlag, New York, 1999.

[22] Zhang, J., "A representational analysis of relational information displays", *Int. J. Human-Computer Studies*, 45, 1996, pp. 59-74.