EG UK Theory and Practice of Computer Graphics (2007)
Ik Soo Lim, David Duce (Editors)

Visualization Notations, Models and Taxonomies

Ken Brodlie and Nurul Mohd Noor

School of Computing, University of Leeds, Leeds LS2 9JT, UK

Abstract

Visualization taxonomies are an important means of imposing some structure on a rather diverse field. We review some earlier work in this area, particularly work based on the use of a notation to label classes of visualization techniques that are appropriate to particular entities. We propose a new notation introducing it in the context of a new visualization reference model, one we hope will lead eventually to a means of describing visualizations in a clear and unambiguous way.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Methodology and Techniques

1. Introduction

There is a continuing interest in the development of taxonomies for visualization. A well-organised taxonomy can provide an important structuring of the field, grouping visualization techniques into classes according to some criteria. Moreover such a taxonomy can act as the basis for a much needed visualization ontology.

A good notation can help in the development of taxonomies, by providing a concise yet expressive description of the equivalence classes of visualization techniques. In this work-in-progress paper we chart the development of notations and taxonomies for visualization over the past twenty years, and relate this to work on reference models for visualization which provide an overall framework. We put forward a new reference model as an evolution of previous models, and suggest a new style of notation that may help develop a visualization taxonomy. Our long-term vision is an ability to describe visualizations in a clear and unambiguous manner, with a formal description that can be stored as metadata with a visualization for archival purposes. This description would allow the recreation of the essential features of a visualization – perhaps long after the software which created it in the first instance became defunct.

2. Evolution of visualization taxonomies, notations, models and ontologies

One of the first proposals for a visualization taxonomy was put forward by Wehrend and Lewis [WL90]. They placed visualization problems and the associated techniques as the elements of a matrix whose rows are objects (such as scalars, directions, positions) and whose columns are operations (such as identify, locate, distinguish). Some 600 examples were classified by Wehrend in his MSc thesis according to this matrix scheme. Keller and Keller [KK92] in their book "Visual Cues", present a large variety of visualization examples, each classified using Wehrend's scheme. In addition, they distinguish the role of dependent and independent variables in the dataset to be visualized. However the classification is performed simply on examples, rather than being a systematic organisation of visualization techniques.

Brodlie [Bro92] takes the view that visualization is concerned with understanding the underlying field (or entity) from which a dataset has been sampled. Thus he aims to provide a taxonomy of entities expressed as a multivariate function of several independent variables. A notation, called the *E-notation*, is proposed, using a superscript to indicate the type of dependent variable and subscript to indicate the type of independent variable. Thus the superscript can take values S (scalar), V (vector), T (tensor), and the subscript is a number indicating the dimension – but using [] to indicate aggregation and {} to indicate an enumerated set. For example, a volumetric entity such as a temperature field in a cavity would be labelled as E_3^S. This proved a useful means of classifying techniques – that is, the notation acted as a label for equivalence classes of techniques which could be applied to visualize entities of that type. Thus the E_3^S notation acted as a label for volume visualization techniques such as isosurfac-

© The Eurographics Association 2007.

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

ing and volume rendering, but did not attempt to distinguish between the two, quite different, approaches. Similar efforts were made in Brodlie [Bro92], [Bro93] and Gallop [Gal94], with further work by Wright [Wri07] in a new textbook on scientific visualization. While this approach does provide some structuring of techniques, and while Brodlie [Bro93] did try to label display techniques as well as entities, generally the approach lacks the 'problem-oriented' theme of Wehrend and Lewis, in which the user's goal is incorporated into the classification process.

This early work tended to focus on scientific visualization. Tory and Moller [TM04] provide a taxonomy with much broader scope, unifying the fields of scientific and information visualization. Their work is also notable for its underpinning by a reference model for visualization. In this model, they distinguish the data from the underlying *object of study* (similar to the entity in [Bro92]), and show how visualizations are then derived on the basis of two models: the *design model*, in which the algorithm developer incorporates specific choices such as interpolation method, and the *user model*, in which the user selects the technique on the basis of their own mental model. Another important contribution is a clear definition of the datatypes involved in visualization, and we shall make use of this work later in this paper.

Work on reference models for visualization has proceeded in parallel to the efforts on taxonomies. A landmark contribution was the dataflow reference model proposed by Upson et al [UFK*89] and Haber and McNabb [HM90]. This expressed the visualization process as a pipeline of elementary processes, each of type *Filter* (to refine the data), *Map* (to select a visualization idiom, turning numbers into geometrical representation) or *Render* (to display the geometry using computer graphics techniques). This model has served the community well: it acted as the basis for the development of Modular Visualization Environments (MVEs) in the late 1980s and early 1990s, and was extended to support collaborative visualization in the work of Wood et al [WWB97] and Duce et al [DGJ*98].

A different approach, first suggested by Duce, was proposed in Brodlie et al [BDG*04]. This saw visualization as a three layer process: a *conceptual layer*, in which a high level view of the required visualization is conceived; a *logical layer*, in which this view is refined in terms of specific pipelines of processes; and a *physical layer*, in which the pipeline is realised in terms of specific software and hardware resources. This can be seen in partnership with the Upson/Haber–McNabb model; and the conceptual layer has some affinity with the user model in the Tory–Moller work.

Ontologies provide a much broader structure to a subject than a simple classification or taxonomy. A visualization ontology will define the concepts involved in the subject and the relationship between them. A discussion of visualization ontologies is provided by Duke et al [DBDH05]. Work on notation and taxonomies can be a useful contribution to ontology research, and indeed Shu et al [SAR06] derive a visualization ontology starting from the E-notation mentioned earlier.

The work in this paper begins with a new look at reference models for visualization, since getting the framework correct is an important basis for subsequent work. This is covered in section 3. We then propose in section 4 a new notation – a development of the earlier E–notation. This is based on the underlying entity, and therefore will correspond to the earlier taxonomies of Brodlie and Tory and Moller. Section 5 presents conclusions and the direction of our future work.

3. A Reference Model for Visualization

It is useful we believe to present any taxonomy in the context of a clear reference model. In this section we describe the model we are using; it is an evolution from the previous efforts described above. The model is shown in Figure 1. We distinguish data elements from processes that operate on data.

We start from the *entity*, or underlying field, we are trying to visualize. Our taxonomy will provide a notation to describe this in terms of its dependent and independent variables. The *data* to be visualized arises from a sampling by measurement of this field, or by a simulation process. We attach to this *metadata* which will influence the visualization technique that is applied: for example, we might have metadata which specified that the underlying scalar field is everywhere positive, and so this can be incorporated in the interpolation process which recreates a model of the field for visualization.

The *analytical task* describes the goal of the user in carrying out the visualization. This is similar to the *operation* of Wehrend and the *user model* of Tory and Moller. It is used to distinguish between different techniques that are applicable to a given entity. Essentially the description of the entity, the data, the metadata and the analytical task drive the route through the rest of the model-which itself is a variant on the Upson/Haber–McNabb model. The idea of including the analytical task is also motivated by the work of Tufte who has long argued that the design of a graphic must be driven by the mental task it is meant to support [ZT04].

The first process in the pipeline, *Data Operations*, applies necessary transformations to the data. This may include subsetting, or in the case of multivariate datasets, a splitting of the data into individual components. Thus the pipeline can fan out at this point.

Each pipeline then proceeds by applying a visualization *Technique*, perhaps informed by a modelling step which recreates an estimate of the underlying field. The output from this step is an abstract, or logical, version of the visualization: by this we mean it comprises the essential elements that distinguish the visualization, without reference to the particular way it is displayed on any given device. This abstract version (an AVE, or Abstract Visualization Element) then passes to a *Representation* step, in which particular rendering attributes are added to create an RVE, or Realised Visualization Element. This separation of meaning

© The Eurographics Association 2007.

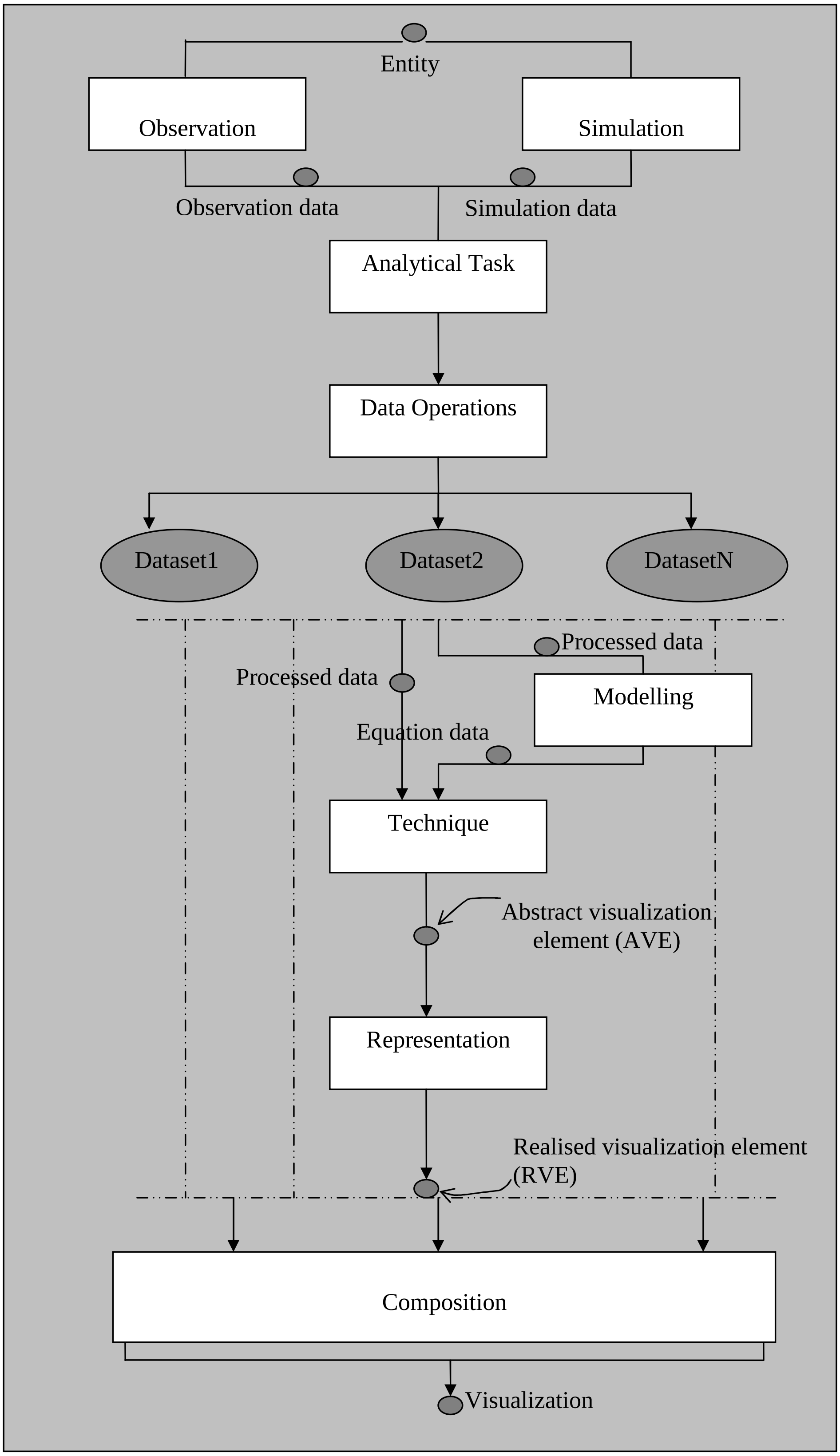


Figure 1: *Reference Model.*

© The Eurographics Association 2007.

from representation has its roots in the graphics standards work of the last century, when GKS for example, was defined in terms of a logical, workstation-independent, picture, and its subsequent representation on different workstations [HDGS86]. Our thinking here is influenced also by Duce's reference model for visualization, where the conceptual layer expresses a high level description of a visualization. Finally we are also motivated by the potential use of the model for provenance and archiving of visualizations-we may wish to recreate the logical version exactly, but be unconcerned whether we reproduce the precise representation.

Finally we have a *Composition* step, in which the RVEs from the set of pipelines are combined in an appropriate way (often depending on the analytical task).

4. Domino Notation

We now discuss a notation for describing the underlying field, as a labelling for the taxonomy we wish to introduce. We shall start from the earlier E-notation, but we shall aim to incorporate some of the improvements we can glean from the work of Tory and Moller, and others. Our guiding principles are:

- We wish to retain the separation into dependent and independent variables.
- We wish to incorporate a better description of datatypes.
- We wish to avoid redundancy in the notation (a criticism from Hopgood, [Hop93], of the E-notation was that the largest symbol, that is, E, is redundant!).
- We wish to find a distinctive and memorable idiom that will appeal to a wide community.

4.1. Datatypes for Dependent and Independent Variables

We shall simplify the earlier notation by using the same datatypes for dependent and independent variables – that is, the range and the domain. We shall consider four basic types: real (R); integer (Z) – with possibility of positive integers (Z^+); and two categorical types, ordinal (O) where there is an ordering, and nominal (N) where there is not.

We see a need to distinguish the case where there is a single value, or a set of values – for example, an interval of real numbers rather than a single real number. We shall use the {} notation to indicate a set of values. There is also increasing awareness in visualization that uncertainty may not just be expressed as a simple interval of values, but rather as a Probability Density Function (PDF) (see Love et al, [LPK05]). At present we shall allow the {} notation to cover this case as well but it may be in the future that the two cases should be distinguished.

4.2. Dimensionality

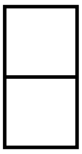
We shall indicate dimension as a *multiplicative* factor. Thus a three-dimensional volumetric domain would be designated RxRxR, or R^3. This allows us also to have mixed type domains or ranges, for example, RxZ^+. Multivariate data is indicated as an *additive* factor. Thus a CFD dataset of velocity, pressure and temperature would be designated as $R^3 + R + R$.

4.3. Redundancy

To make the notation as compact as possible, we remove any symbol such as E that does not contribute identification information. Furthermore, we expect that in most instances we shall be dealing with a single value, rather than a set of values or a PDF, and so we only include symbols (i.e. '{}') if we wish to indicate a set of values.

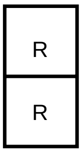
4.4. Metaphor

To make the notation memorable, we shall use a *domino* as the metaphor. The top indicates the dependent variable; the bottom the independent variable. Thus we have:

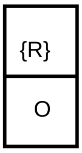


4.5. Simple Examples

A simple example is data representing values of a quantity measured at various points along an axis-for example, length of a bar measured at different temperatures and typically visualized as a 1-D graph. This would be represented as:



Another simple example but showing different datatypes is the following. Figure 2 (taken from http://labwrite.ncsu.edu/res/gh/gh-bargraph.html) shows a visualization of the range of temperature observed at Raleigh-Durham airport each day in December 1996. The entity here has a range that is an interval of real numbers, and a domain which is an ordinal variable- hence designated:



© The Eurographics Association 2007.

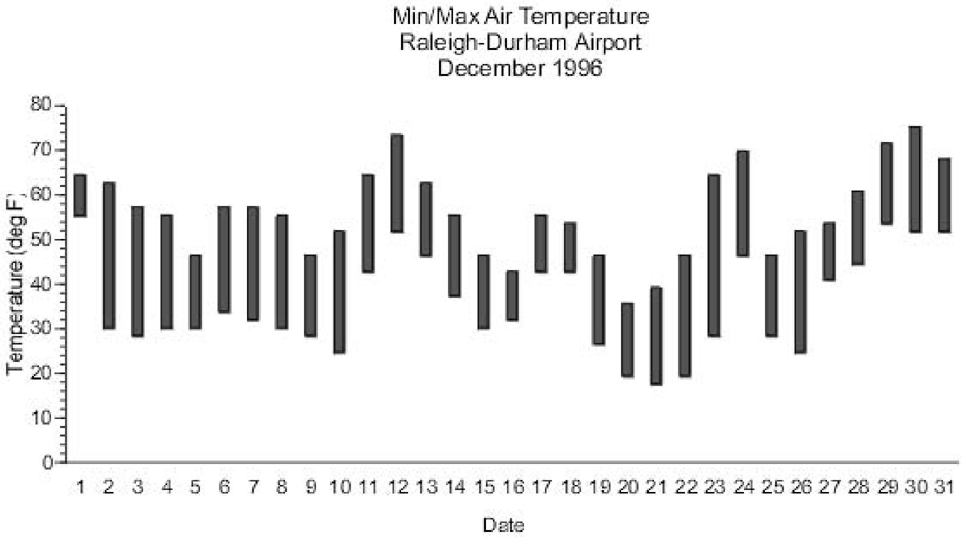
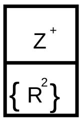


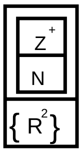
Figure 2: *Interval range and ordinal domain.*

4.6. A more complex visualization

The notation can also be used in more complex cases. For example, consider the situation where the visualization is of the population of the countries in the world. This would be an integer dependent variable, with a 2D 'region' independent variable, hence the notation:

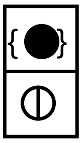


However suppose the entity was the male-female split of population: then the dependent variable is another entity in itself, namely an integer dependent variable and a nominal independent variable (male, female). We can then nest the dominos as follows:



4.7. Another domino representation

We are also exploring different representations of the dominos. For example, rather than use letters as symbols, can we take the domino metaphor further and use dots. Figure 3 provides a suggestion; the data from Figure 2 would be designated as:



5. Taxonomies

The notation of section 4 will act as a basis for a taxonomy of visualization techniques, similar to that of Brodlie [Bro93] and Tory and Moller [TM04].

In order to build up examples to populate this taxonomy we are creating a wiki, to which visualization scientists can add interesting examples. The wiki is being established at: `http://dominonotation.wikispaces.com`

Datatypes	Symbols
Real Numbers(R)	●
Integer (Z, Z^+)	○
Ordinal (O)	⦶
Nominal (N)	⊕

Figure 3: *Table showing a possible graphical representation of the notation.*

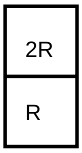
6. Conclusions and Future Work

We are re-visiting the work on notations, models and taxonomies for visualization. Our focus has initially concentrated on notation where we have reflected on

© The Eurographics Association 2007.

the E-notation from Brodlie [Bro92] and aimed to both simplify it and make it more expressive- proposing the domino notation. We have also suggested a variation on the Haber-McNabb reference model for visualization, in which we incorporate the concept of analytical task.

Our next step is to develop a language in which the analytical task can be expressed. To give a flavour of our thinking, consider a simple two variable scalar 1D dataset – that is, in our notation:



The analytical task might be to **display** the two variables, with an aim to **compare by juxtaposing** the graphs. This would then drive the subsequent path through the reference model: 'compare' drives the Data Operations process to separate the two variables so that each can be processed by a separate pipeline; 'display' selects the default choice of Modelling and Technique for this type of data; and finally 'juxtapose' drives the Composition stage to arrange the two graphs side-by-side.

Longer term we see our work contributing to a means of describing visualizations in a more rigorous and precise way than is possible at present. This will have benefits in long term archiving of visualizations – so they can be recreated long after the software that generated them has expired- and in applications such as asynchronous collaborative visualization where a precise description is needed in order to exchange descriptions between collaborators.

References

[BDG*04] Brodlie K., Duce D., Gallop J., Sagar M., Walton J., Wood J.: Visualization in grid computing environments. In *Proceedings of IEEE Visualization 2004* (2004), Rushmeier H., Turk G., van Wijk J. J., (Eds.), IEEE Computer Society, pp. 155–162.

[Bro92] Brodlie K.: Visualization techniques. In *Scientific Visualization - Techniques and Applications*, Brodlie K., Carpenter L., Earnshaw R., Gallop J., Hubbold R., Mumford A., Osland C., Quarendon P., (Eds.). Springer-Verlag, 1992, ch. 3, pp. 37–86.

[Bro93] Brodlie K.: A classification scheme for scientific visualization. In *Animation and Scientific Visualization*, Earnshaw R. A., Watson D., (Eds.). Academic Press, 1993, pp. 125–140.

[DBDH05] Duke D., Brodlie K., Duce D., Herman I.: Do you see what I mean? *IEEE Computer Graphics and Applications 25*, 3 (2005), 6–9.

[DGJ*98] Duce D., Gallop J., Johnson I., Robinson K., Seelig C., Cooper C.: Distributed cooperative visualization - the MANICORAL approach. In *Proceedings of the 1998 Eurographics UK Conference* (1998), pp. 69–85.

[Gal94] Gallop J.: State of the art in visualization software. In *Visualization in Geographical Information Systems*, Hearnshaw H. M., Unwin D. J., (Eds.). Wiley, 1994, pp. 42–47.

[HDGS86] Hopgood F. R. A., Duce D. A., Gallop J. R., Sutcliffe D. C.: *Introduction to the Graphical Kernel System (GKS) (2nd ed. revised for international standard)*. Academic Press Ltd., London, UK, UK, 1986.

[HM90] Haber R., McNabb D.: Visualization idioms : A conceptual model for scientific visualization systems. In *Visualization in Scientific Computing*, G.M. Nielson B. S., Rosenblum L., (Eds.). IEEE, 1990, pp. 74–93.

[Hop93] Hopgood F.:, 1993. Personal communication.

[KK92] Keller P., Keller M.: *Visual Cues*. IEEE Computer Society Press, 1992.

[LPK05] Love A., Pang A., Kao D.: Visualizing spatial multivalue data. *IEEE Computer Graphics and Applications 25*, 3 (2005), 69–79.

[SAR06] Shu G., Avis N., Rana O.: Investigating visualization ontologies. In *Proceedings of UK E-Science All Hands Meeting* (2006).

[TM04] Tory M., Moller T.: Rethinking visualization: A high level taxonomy. In *Proceedings of Information Visualization* (2004), IEEE Computer Society Press, pp. 151–158.

[UFK*89] Upson C., Faulhaber T., Kamins D., Laidlaw D., Schlegel D., Vroom J., Gurwitz R., van Dam A.: The Application Visualization System : A computational environment for scientific visualization. *IEEE Computer Graphics and Applications 9*, 4 (1989), 30–42.

[WL90] Wehrend S., Lewis C.: A problem-oriented classification of visualization techniques. In *Proceedings of IEEE Visualization 90* (1990), IEEE Computer Society Press, pp. 139–143.

[Wri07] Wright H.: *Introduction to Scientific Visualization*. Springer, 2007.

[WWB97] Wood J., Wright H., Brodlie K.: Collaborative visualization. In *Proceedings of IEEE Visualization 97* (1997), ACM Press, pp. 253–259.

[ZT04] Zachry M., Thralls C.: An interview with Edward Tufte. *Technical Communication Quarterly 13*, 4 (2004), 447–462.