# PUC

ISSN 0103-9741

Monografias em Ciência da Computação

n° 16/12

# Semantic Characterization of Visualization Mechanisms

**Taissa Abdalla Filgueiras de Sousa**

**Simone Diniz Junqueira Barbosa**

Departamento de Informática

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO**

**RUA MARQUÊS DE SÃO VICENTE, 225 - CEP 22451-900**

**RIO DE JANEIRO - BRASIL**

Monografias em Ciência da Computação, No. 16/12 ISSN: 0103-9741
Editor: Prof. Carlos José Pereira de Lucena November, 2012

# Semantic Characterization of Visualization Mechanisms

Taissa Abdalla Filgueiras de Sousa
tsousa@inf.puc-rio.br
Simone Diniz Junqueira Barbosa
simone@inf.puc-rio.br

**Abstract.** Several researchers have shown the need to create formal meanings of different types of data visualizations. This concern includes not only scientific visualization, but also more general information visualization. Lately, some projects have been conducted towards developing ontologies, creating semantic for visualization and making these representations amenable to computer processing. In this research, we investigated two recent ontologies - VisKo and UK National e-Science Centre - aiming to bring new contributions to the field of information visualization. The goal of our research is to develop an ontology to define relationships between data model, visualizations and user tasks represented by questions. Our ontology involves some usual and unusual data structures that include Cartesian graphs, maps and hierarchical relationships and we defined questions that could be answered by these visualizations.

**Keywords**: Ontology visualization, semantic web, information visualization.

* Research work sponsored by CNPq, FAPERJ, VRAc/PUC-Rio, and IBGE.

**In charge of publications**:

Rosane Teles Lins Castilho
Assessoria de Biblioteca, Documentação e Informação
PUC-Rio Departamento de Informática
Rua Marquês de São Vicente, 225 - Gávea
22453-900 Rio de Janeiro RJ Brasil
Tel. +55 21 3527-1516 Fax: +55 21 3527-1530
E-mail: bib-di@inf.puc-rio.br

# 1 Introduction

Knowledge about visualization involves psychology, cognition and semiotics. The goal of visualizations is to aid understanding of data by leveraging the ability of the human visual system to identify patterns, detect trends and discrepancies [15]. Graphical representations are effective means of communication, since they take advantage of cognitive mechanisms, particularly perception [20] and, if well designed, they can replace cognitive calculations with simple perceptual inferences [15]. Adequate visualizations can aid in comprehension, memory and decision making, while inadequate visualizations may hinder the reader's interpretation of the depicted data.

Researchers have shown concern about the need to create formal meanings for visualizations for several reasons [23][11][12][6][18]. They have conducted surveys with the goal of making visualizations more understandable on the Web, aggregating semantics to visual representations to improve communication between users and visualization systems at all stages of the communication cycle. To achieve this, both the user and the system must share a common visualization language. Duke et al. [12] state that these meanings can be expressed in different forms and list three relevant concepts: terminology, taxonomy, and ontology. However, they also affirm that only ontologies are formal and can be processed by computer. Ontologies describe concepts through sets of constructors with predefined meanings, for instance, using a set of possible relationships between primitives.

Regarding ontologies, Allemang [1] claims that the Semantic Web should not be confused with an effort to create a single ontology with which all agree, but it should achieve some degree of interoperability. Thus, our effort has been towards studying solutions to create an ontology to define relationships between data and visualization models to help answer users' questions via efficient visualizations.

Data visualization can either represent scientific data with a spatial component (3D) or more abstract and non-spatial data [23][14]. Our ontology includes only a set of representations for the latter. The selected visualization models include common types of representations and other less common but still useful. Some of them are part of Heer's research [15]: a) Different compositions of 2D Cartesian charts to represent multiple time series, scatter graphs, maps with or without charts composition and hierarchical structures.

Some factors must be considered when designing efficient visualizations. According to Chang [7], to design multidimensional visualizations we must analyze, for instance: (1) What visual features should we use for each data attribute? (2) Is there visual interference between different visual features? (3) What is the importance of a data attribute in the overview? (4) Is the domain of a data attribute continuous or discrete? (5) Is the spatial frequency of certain attributes high or low? Throughout this research we will answer these questions when designing and justifying our ontology.

This paper is organized as follows. Section 2 presents and compares existing solutions and studies on the development of visualization ontologies. In Section 3 we define and detail our proposed ontology. Sections. Finally, Section 4 concludes with a brief discussion about future work.

# 2 Related Work

In this section we present two projects that use data visualization ontologies: Visko [10] and the ontology of the UK National e-Science Centre.

VisKo focuses on the capture and management of visualization knowledge required to construct pipelines. The collective knowledge of visualization is used to provide answers to visualization queries that require input data to be transformed into their equivalent graphical representation. The project is composed of four components and one of them is the ontology to encode knowledge. VisKo's ontology defines a formal visualization language that allows describing different visualization abstractions, toolkit operators to generate those abstractions, and web services to implement the toolkit operations. To distinguish between these different visualization concerns, the ontology is defined by three sub-ontologies [24]. Among them, VisKo-View and VisKo-Operator are of most interest to this study.

The View-VisKo ontology (Figure 1) defines different types of graphical data or information, such as contour maps, volumes, graphs, and networks. The ontology classifies these visual representations according to their construction from geometrical primitives such as points, lines, polygons, and grids or high-level graph-like boxes or other figures, and connectors commonly found in information visualization [24].

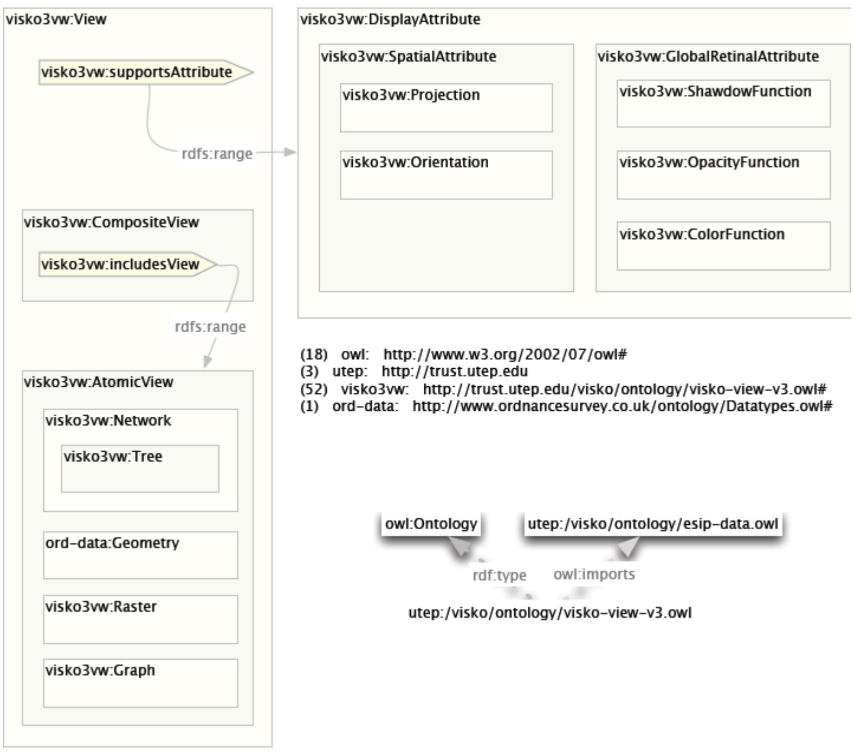



**Figure 1: VisKo-View Ontology**

The VisKo-Operator Ontology (Figure 2) defines different classes of operators: viewers, mappers, and transformers. It also declares a common property, which states that data formats are processed by an operator [24].

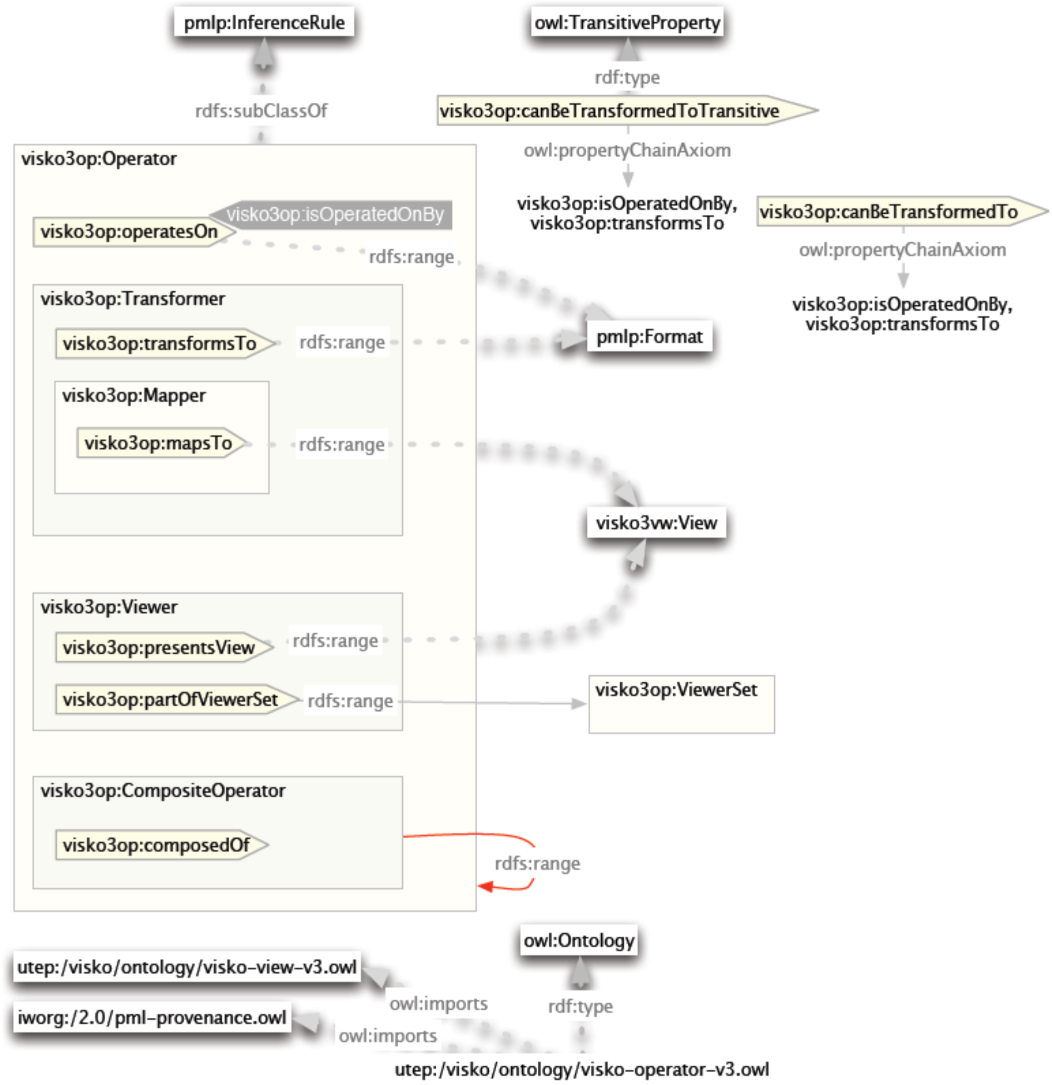



**Figure 2: VisKo-Operator Ontology**

The second project we investigated is the ontology proposed by Program UK National e-Science Centre. It was developed because activities in the program identified a need to establish a visualization ontology. They were motivated by the definition of Web services and semantic grids to support collaborative work, curating, and sustaining research and education on visualization. Thus, members of the UK visualization community have developed a structure for the ontology [11][18].

Figure 3 presents the top-level components of the visualization ontology.

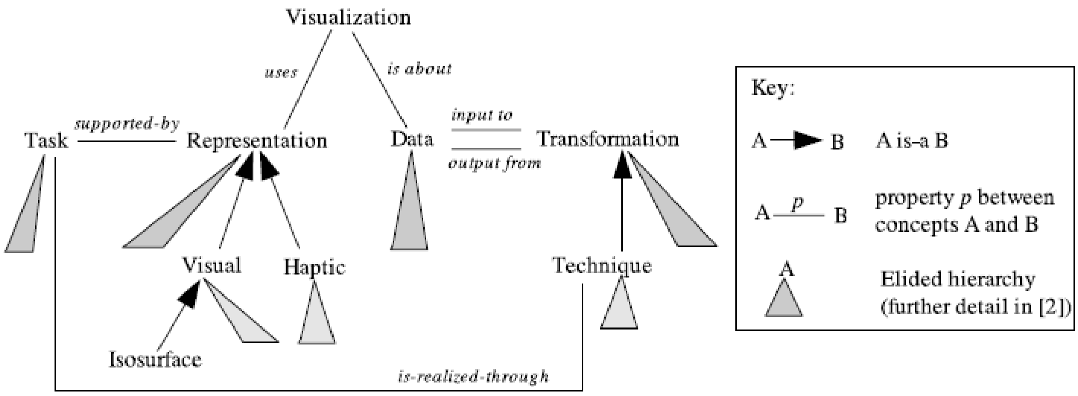



**Figure 3: Top-level components of the UK visualization ontology**

According to Duke et al. [11], based on this structure (Figure 3), members of the UK visualization community suggested it could be classified into four groups of concepts:

1. Tasks and use: Its features includes why data is being visualized and what is done with the generated visualization.
2. Representation: It is related to how data is presented.
3. Process: This concept aims to organize how the components of transformed data will be expressed in primitives.
4. Data: How are the data processed through transformations expressed and organized?

In the next section, we draw on the related work to define our visualization ontology.

## 3 A Visualization Ontology

Based on the two investigated ontologies, we observed that some classes were common to both: Visualization (View in VisKo), Data and Transformation. Visko still has a class Display Attributes to define the visualization features. The concept of task appears only in the UK's ontology.

We concluded that we needed five classes to create our ontology. The three common classes, display attributes and task to classify questions the user want to answer. Then, our ontology was organized in the following top-level classes:

1. Data: define the data attributes.
2. Display Attribute: visualization techniques and their relation to features of data to be displayed.
3. Visualization: graphical representations. In this research we selected a limited set of graphical representations.
4. Task: the goal of the user viewing the data.
5. Transformation: the process of separating or classifying the data components into display attributes.

Each of these classes will be detailed next.

### 3.1 Data

In order to create efficient and appropriate visualizations to the data, Heer et al. [15] pose as challenge the fact of for each dataset, the number of visual encoding (design of recognized as a challenge the large number of available visual representations. It is necessary to know the data properties to select effective visual encodings and to map data values onto graphical features, such as position, size, color and shape.

Based on the work of Tory et al. [23], Ignatius and Senay [16], and Freitas et al. [14], we set up Table 1 to facilitate the understanding of the data classifications.

**Table 1: Data classification [23] , [14] and [16]**

| | Criteria | Class | Group |
|---|---|---|---|
| Dimensions<br>(1D, 2D, 3D, ... nD) | Number of variables | Dependent<br>Independent | |
| Variable type | Data structure | Scalar<br>Vector<br>Tensor | Quantitative |
| | Domain Nature | Discrete<br>Continuous | Quantitative |
| | Type of Value | Nominal<br>Ordinal<br>Interval<br>Ratio | Qualitative<br><br>Quantitative |

Table 1 shows the criteria used for classifying data. The number of dimensions is given by the number of variables. Independent variables are those that are manipulated whereas dependent variables are only measured or registered

]. Variables also need to be classified according to their structure. Structures of scalar data have magnitude but no directional information besides the signal. These structures are therefore defined by a single number. Vectors have direction and magnitude. Tensors are specified by the dimensionality of the coordinate system and the order of the tensor [16]. In this paper, we work only with scalar data.

Another classification criterion is the nature of the domain. Discrete data can take finite numbers or infinite countable numbers of distinct values. Continuous data are the ones that can take values of a real interval [5]. Along the other dimension, quantitative data can be classified as interval, ratio, and ordinal, while qualitative data as nominal and ordinal. In this research, ordinal data will be always classified as qualitative.

Interval scales preserve the real differences between quantitative values (such as Fahrenheit degrees), but do not have a natural zero point. Ratio scales are interval scales with a natural zero and can be defined in terms of arbitrary units. For example, U$200 is twice as much as U$100 [16]. Nominal scales have no natural order (*e.g.*, red, blue, green), while ordinal scales do (*e.g.*, sets of small, medium, and large) [23].

We investigated ontologies that describe statistical data, such as Scovo and Data Cube Vocabulary [9]. Scovo uses three basic concepts ]:

1. Dataset: the container of some data, such as a table with some data in its cells;
2. Item: a single piece of data (*e.g.*, a cell of a table);
3. Dimension: some kind of unity of a single piece of data. (*e.g.* time, location)

Data Cube Vocabulary is focused only on publishing multidimensional data on the web and uses the Dataset class of the Scovo vocabulary.

he following diagram (Figure 4) depicts the proposed data ontology, representing the data, its attributes, its properties, and the relationships between its elements.

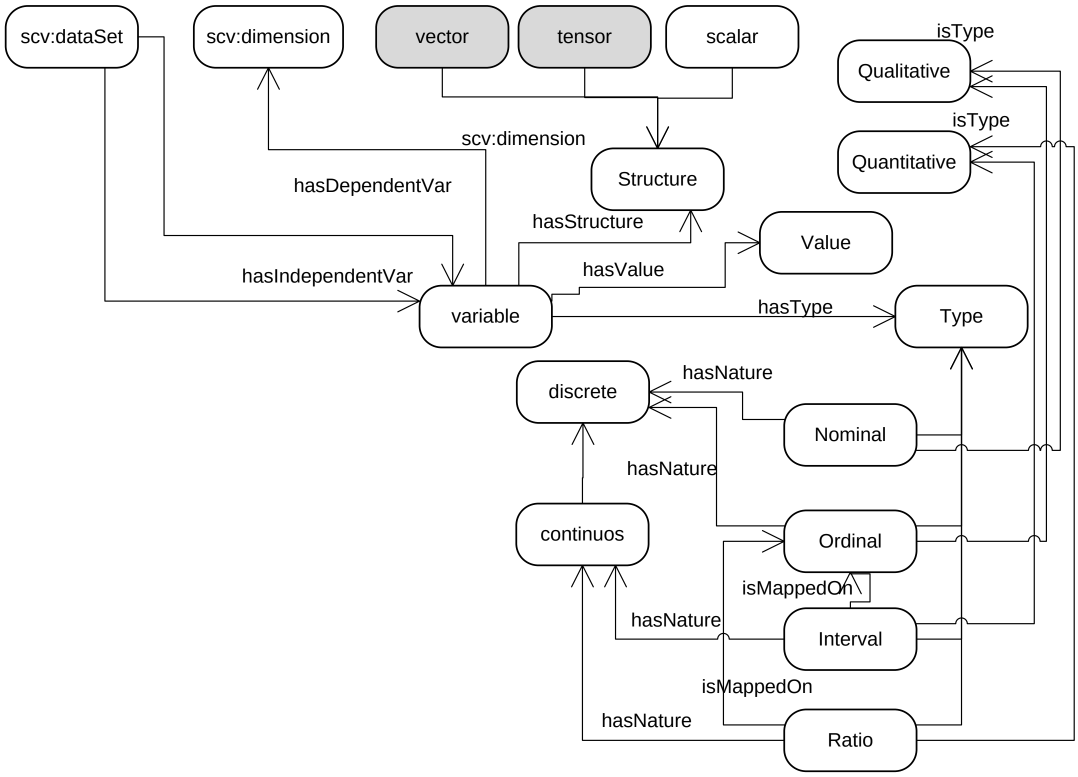



**Figure 4: Data Ontology**

In addition to the relationships defined in Table 1, to create Figure 4 we used the dataset concept from the Scovo vocabulary to define the data class. We also used the property named dimension and the homonymous class from Scovo to determine the variable dimensions. We created relationships (isMappedOn) between the quantitative interval and ratio data and the qualitative ordinal categories. We propose to allow transformations from quantitative scales to qualitative categories and also to include other dimensions in the representations. The shaded classes are not part of our universe of study.

## 3.2 Display attributes

To understand the Display Attributes, some concepts are important: **visualization vocabulary, mark, compound mark, types of perception,** and **composition rules.**

According to Ignatius and Senay [16], a **visualization vocabulary** identifies basic construction blocks of techniques of scientific visualization. Moody [17] called it a visual alphabet and defined it as a set of visual variables. According to Moody [17], visual variables define a set of atomic building blocks that can be used to build any visual representation, setting the dimensions of the graphical design space.

The concept of **mark** defined by Ignatius and Senay [16] is the most primitive building block that can encode some useful information in data visualization. They explain that each mark can be classified as simple or compound. There are four types of simple mark:

1. Points – marks with a single center that could indicate a significant position.
2. Lines – marks with a significant length or a significant connection.
3. Areas – marks with a significant interior that could indicate a region or 2D space
4. Volume – marks with a significant interior that could indicate 3D space

**A compound mark** is a collection of simple marks that form a perceptual unit.

According to Bertin [4], marks can vary in position on a plane. Therefore, the mark can express a correspondence between two series constituted by two planar, horizontal or vertical dimensions.

Marks have properties. These properties are classified into positional, temporal, and retinal. The data are usually encoded by varying the properties of the marks in a visualization technique. Positional encoding of information is a variation of the positions of the marks in the image. Temporal encoding of information is a variation of the properties of the mark over time. Retinal encoding of information is any variation in the properties of the marks that the eye retina can perceive, besides position. The retinal properties are: size, texture, orientation, shape, and the three features of color: hue, saturation, and brightness.

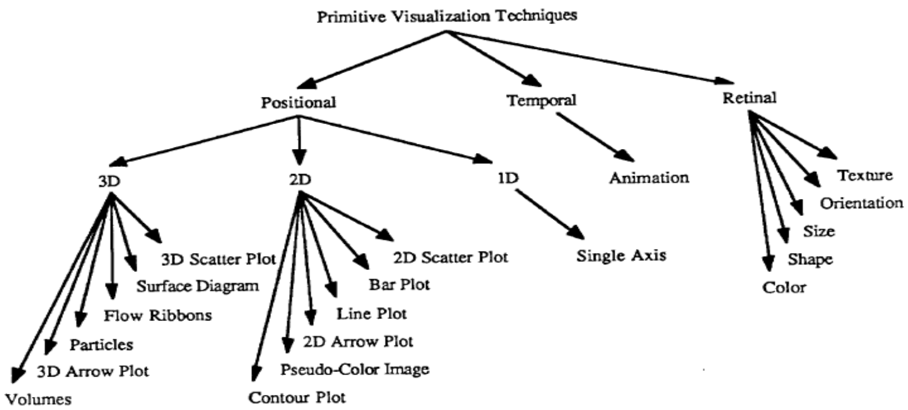



**Figure 5: Primitive Visualization Techniques supportesd by project VISTA.**

Primitive Visualization Techniques are also defined by Ignatius and Senay [16] as those that encode one dependent variable and up to four independent variables (Figure 5). Additional variables (dependent or independent) that can exist in a given data set can also be encoded using Primitive Visualization Techniques or by combining two or more of these techniques in a single project. Each technique can usually be classified into one of three categories: positional, temporal, or retinal — it depends on the property of the mark. Positional techniques can be one, two, or three dimensions; for instance: single-axis, graph contouring, and surface diagram, respectively. There is only one temporal technique: animation. Retinal techniques correspond to the set of properties of retinal marks such as shape, size, orientation, color and texture.

In Moody's research [17], color is presented decomposed into hue and brightness, its set of graphical techniques does not consider the variation of saturation. Bertin [4] alerts to the use of saturation because it is not of constant value. Saturation tone varies in value according to the hue, and can thus cause perception and interpretation problems.

As the selected representations have only two positional dimensions, to include a third or a fourth dimension and show multidimensional data, we need to use retinal or temporal techniques. The use of efficient retinal marks could help the interpretation, so it should be selected in line to the types of perception. According to Bertin [4], there are four **types of perception**:

1. Associative: Useful when we want to equalize a variation or to create clusters with all marks of each category. Variations in shape, orientation, hue and texture are associative. Variations in shape and orientation are used to reveal similar and different elements.
2. Selective: Used to answer a question: "Where is the category?" The eye should be able to isolate all elements from this category, disregard all other signals and perceive the image through the given category. Size, brightness, hue, texture, and orientation are perceived immediately. However, Bertin [4] affirms that selective perception is very limited and does not advise more than five steps to define a category in size and seven in brightness (in grayscale).
3. Ordered: Used to compare two or more orders. Texture, brightness and size establish orders that are universal and immediately noticeable. You cannot reorder brightness.
4. Quantitative: Used when you need to numerically define the relation between two signs or to search signs homogeneously. Only size is quantitative.

These types of perception are related to the concept of the **composition rules** defined by Ignatius and Senay [16] to display multidimensional data. According to them, five rules describe a large amount of composite visualizations:

1. Mark composition: merges marks of the component visualization techniques by pairing each mark of one technique with a compatible set of marks of the other.
2. Composition by superimposition: merges marks of the component visualization techniques by superimposing one mark set onto the other.
3. Composition by union: combines marks of a pair of component visualization techniques using set union.
4. Composition by transparence: combines a pair of visualization techniques by manipulating the opacity values of marks belonging to either or both visualization techniques.
5. Composition by intersection: combines a pair of visualization techniques by first computing their intersection, then superimposing the intersection onto one of the components.

We can now understand the Display Attribute class used in the VisKo-View ontology. It is subdivided into "Global Retinal Attribute" and "Spatial Attribute". Table 2 illustrates these subclasses.

**Table 2: Classification of the attributes of the Visko-View ontology**

| **Class** | **Subclass** | |
|---|---|---|
| Display Attribute | Global Retinal Attribute | Color<br>Opacity |
| | Spatial Attribute | Orientation<br>Projection |

Color was classified here as a Global Retinal Attribute, as in Bertin's semiology. Although opacity is not described in these sources, it makes sense to also classify it as a Global Retinal Attribute.

Orientation, as seen in Figure 5, is also one of the retinal attributes. It is related to the mark's property of assuming an infinite number of different orientations without changing the position of its center [4]. In Visko-View, it is classified as Spatial Attribute, which can represent the two planar dimensions: horizontal and vertical.

Projection is the transposition of the sphere on a plane, with bending or tearing. The planar representation of the spherical surface is always conducted with deformation of the arrangement of surface elements [4]. The most common projections used in cartography are: Mercator, Miller and Berhmann, and Robinson, to represent the world map [2].

Figure 6 illustrates the display attribute class of our visualization ontology, created based on the previous analysis.

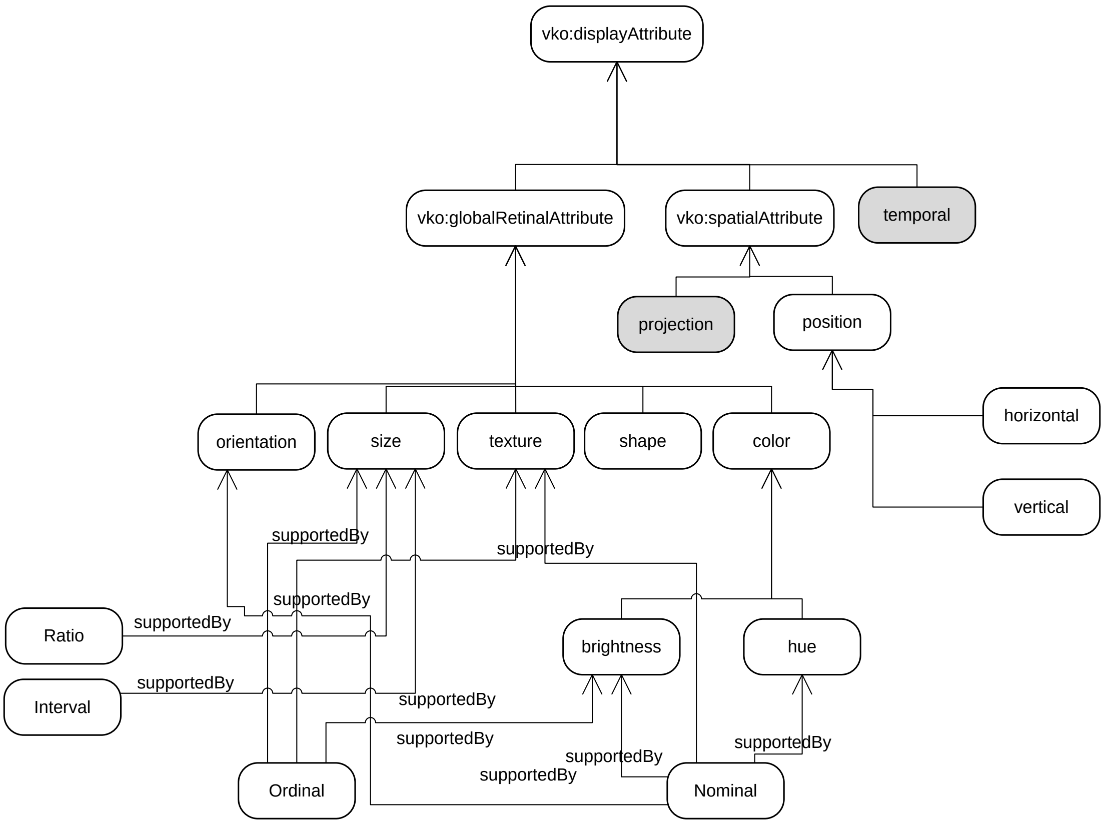



**Figure 6: Display Attribute Class**

We used the top-level classes of Visko-View to start our display attribute class, including more options of retinal attributes and the orientation subclass.

As stated, the orientation, size, texture, shape and color classes define the set of retinal attributes. Among the subdivisions in color, we will use only brightness and hue, because saturation can cause problems due to the difference in brightness across hues. We have also seen that the size variable is ordered and selective, so we can encode quantitative (interval and ratio) data  and qualitative (ordinal) data up to 5 steps. Likewise, brightness and texture also codify ordinal data. Unordered qualitative data (nominal) may also be encoded in the variables: hue, orientation and texture, of selective perception; and brightness, of ordered perception.

In this ontology, we will not define temporal attributes, because they depends on animation and we aim to define only static visualizations. We are not working either with projection variables, because maps projections for these visualizations will always be one of the common projections used in cartography.

## 3.3 Visualization

We defined our visualization class based on Tory et al. [23], Ignatius and Senay [16], and Freitas [14], as depicted in Table 3.

**Table 3: Visualization class.**

| Class | Subclass | | Example |
|---|---|---|---|
| Visualization | Geometric | Point | - |
| | | Curve (1D) | - |
| | | Surface (2D) | Point charts (2D Scatter-plot, qqplot)<br>Bar Chart<br>Series Chart<br>Maps |
| | | Volume (3D) | - |
| | Network | Tree | Radial<br>Cartesian<br>Indented<br>Icicle<br>SunBurst<br>Polar<br>TreeMap |

Tory and Möller's taxonomy [23] considers both scientific and information visualization areas. Figure 7 shows the taxonomy of discrete data (on the left) and of continuous data (on the right). As we can see in Figure 7, the maximum number of dependent and independent variables is determined by the number of dimensions.

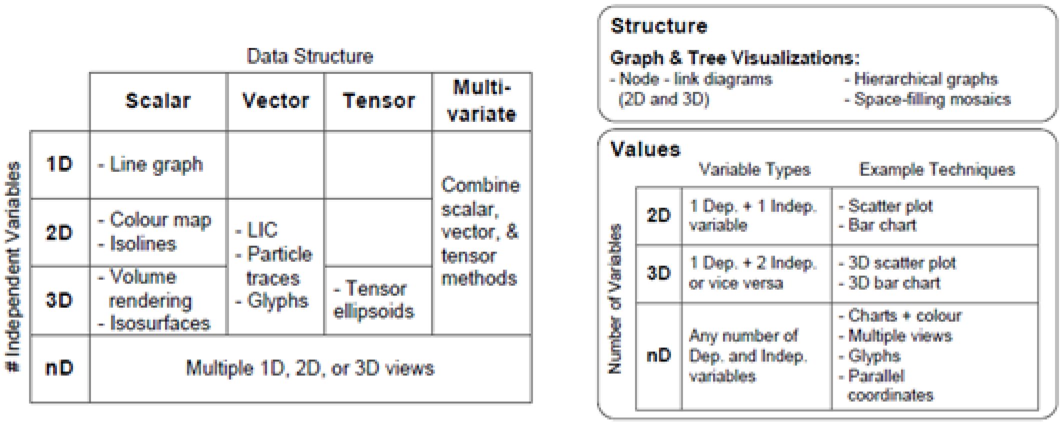



**Figure 7: On the left, taxonomy of discrete models.On the right, taxonomy of continuous models.**

Tory and Möller [1] considered the use of 3D graphics for data visualization in three dimensions. However, according to Few [13], 3D visualizations should be avoided in the information visualization. According to him, there are two ways to use 3D charts. The first is by adding a third depth dimension at the objects that are used to encode quantitative values, without the addition of a third axis. The second is by adding a third dimension of depth graph for the third line with a global scale. None of these forms is efficient for different reasons. Heer [7] also does not use 3D graphics in his study. Some of the views below were studied by Heer [7]. Most of them are charts composed of up to 3 dimensions.

## Time Series Charts

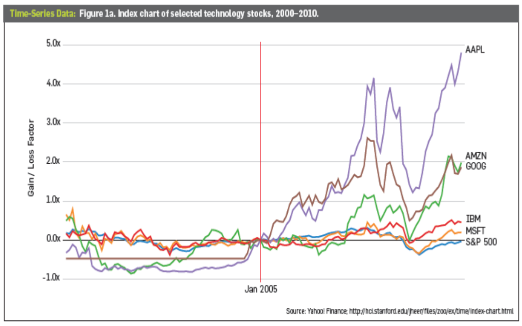



**Figure 8:** **Index chart**

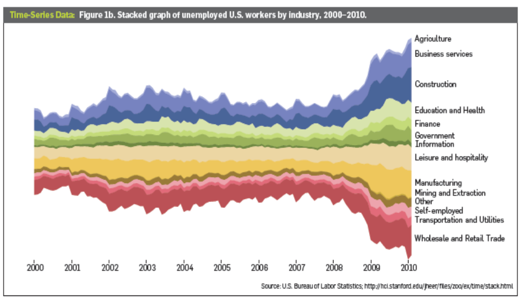



**Figure 9: Stacked series**

Time series charts shows quantitative values in relation to sequential points in time [13]. The four graphs presented in Figure 8 (Index chart), Figure 9 (Stacked series), Figure 10 (Small multiple series) and Figure 11 (Horizon series) correspond to time series. All of them express a sense of continuity that is required to express time [13]. They are all formed by compositions. Figure 8 shows the composition of multiple stocks simultaneously. Although with only two axes, there is a third dimension represented by hue, which is a display attribute. The chart in Figure 9 is a stacked chart and, according to Heer et al. [15], it does not allow negative numbers. The fact is that it allows only numbers with the same sign (all positive or all negative) but we can limit it to positive numbers once we can know the

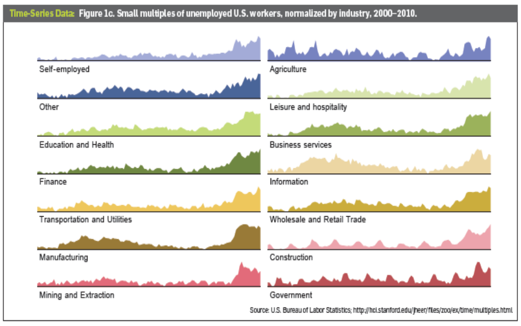



**Figure 10: Small multiple series**

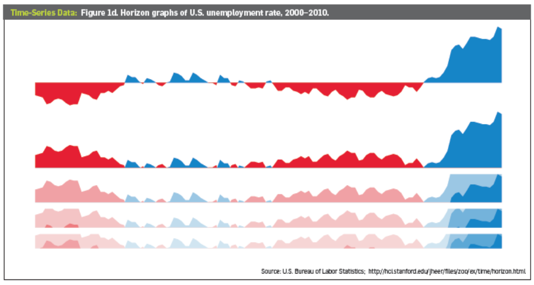


**Figure 11: Horizon series**

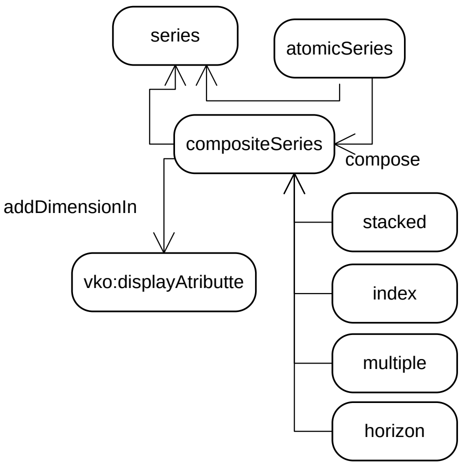



**Figure 12: Series presented in triples**

unit of the data. For instance, the number of people will never be negative, unless it is an indicator, so this chart cannot be used. Figure 9 also has hue as a third dimension. It defines the third dimension, but rather the position, created from new axes.

Figure 10 has a different composition. Color is not the variable that defines the third dimension. It is defined by other axes.

Each sample of Figure 11, shows only two dimensions, and color is used to differentiate between positive and negative numbers because, from the second sample, negatives values are positioned above the horizontal axis as positive values. Thus, brightness is used to reduce the designed area through an overlapping shapes technique.

We set the diagram in Figure 12 to represent the series visualization. Series is an abstract class, which represents both primitives as their containers. AtomicSeries is the primitive object and compositeSeries is a composition of series. All series charts shown are subclasses of compositeSeries. The composition still requires a property addDimensionIn related to a component of this class which can be a retinal mark or a new axis.

## Hierarchical Visualizations: Tree

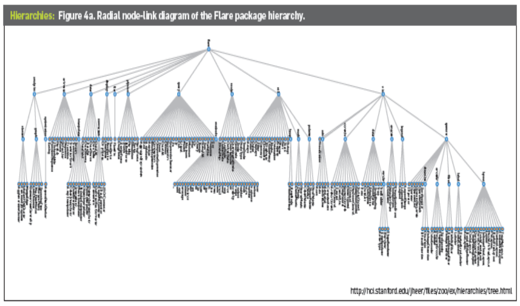



**Figure 13: Radial Tree**

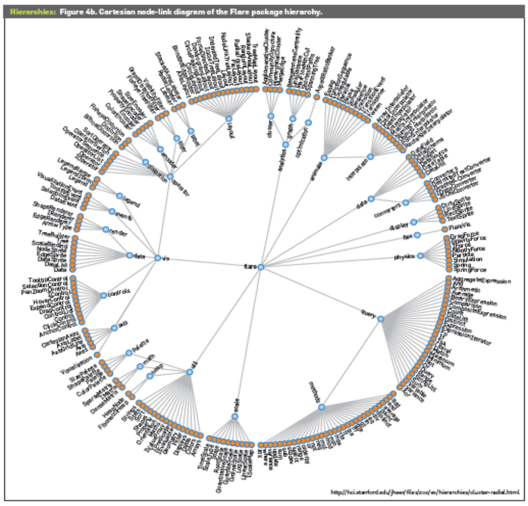



**Figure 14: Cartesian Tree**

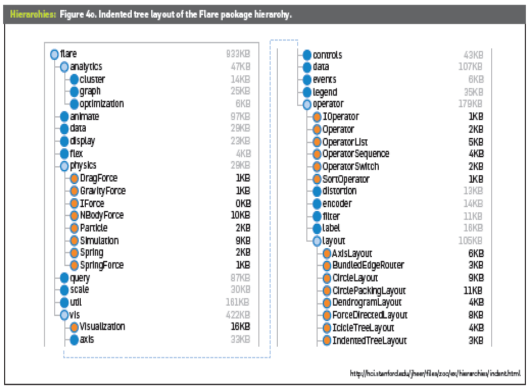



**Figure 15: Indented**

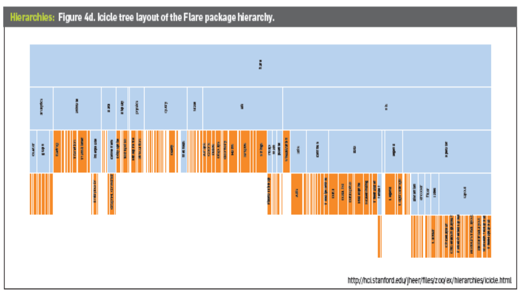



**Figure 16: Icicle**

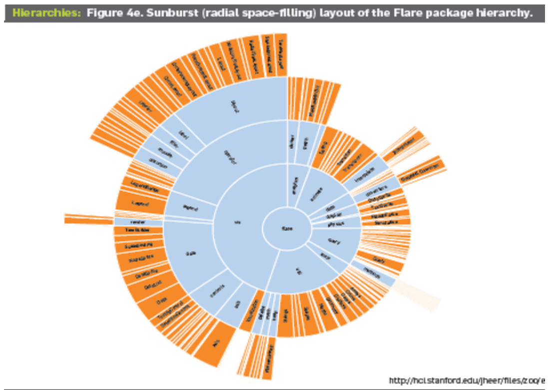



**Figure 17: Sunburst**

Some data can be sorted into natural hierarchies through trees. Trees occupy the two planar dimensions. The elements of this set have the following properties: they are transitive, asymmetric and not reflective. Figure 13 shows a radial tree. It is the most common disposal to represent trees. The color is used to differentiate leaves of intermediate nodes. Figure 13 highlights the leaves, displaying them all at the same level. Figure 15 is a hierarchical structure also often used: an indented tree. This type of display does not facilitate multi-scalar inferences [15]. Figure 16, an Icicle, is similar to a radial tree. It places the root node at the top and the children at the bottom. Icicle, however, includes a new dimension with the size retinal mark. It is possible to identify the size of the class, since top-level classes cover the entire area used by children. Figure 17, Sunburst is also similar to Icicle. It also uses the size as a new dimension.

Figure 19 and Figure 20 depict treemap and nested circles, respectively. They also fill space with a dimension, but using an external container to set the upper nodes. It is easy to verify the limitations of each container, but is not that easy to identify the levels where the leaves are the treemap. That is a little easier on the nested circles.

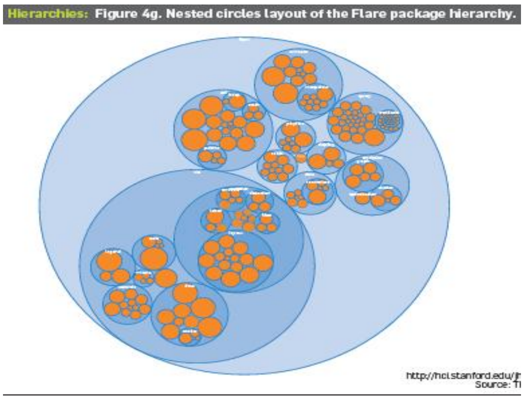



**Figure 18: Nested Circles**

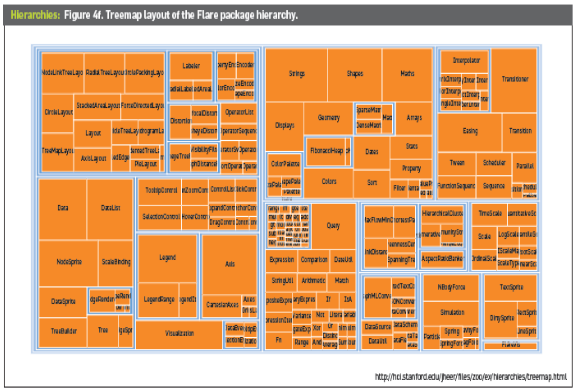



**Figure 19: TreeMap**

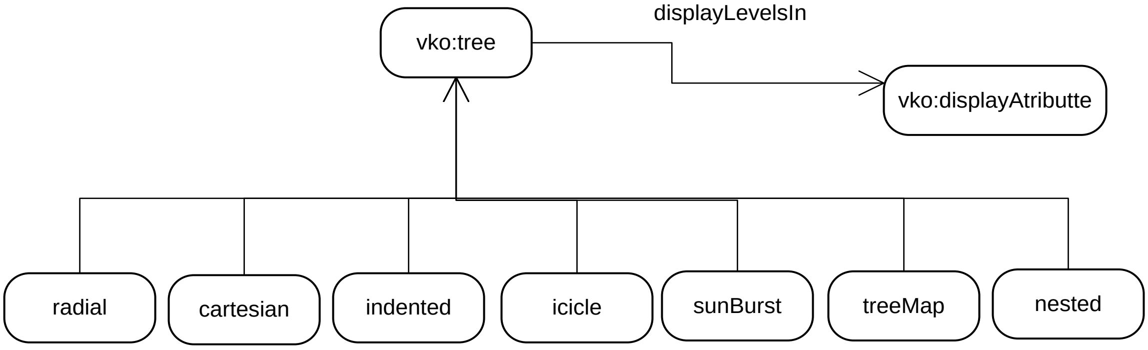



**Figure 20: Series presented in triples**

Figure 20 shows the part of the visualization class that covers the hierarchical visualizations. All presented samples are subclasses of the top-level tree and they also need a property to create dimensions on display attributes.

## Point charts

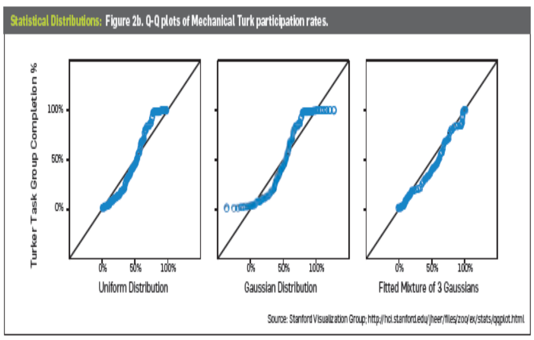



**Figure 21: Q-Q plot**

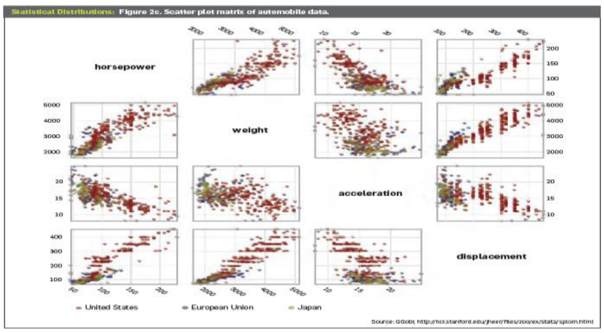



**Figure 22: Scatterplot**

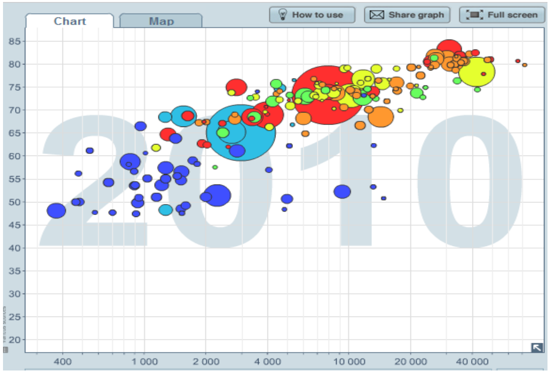



**Figure 23: Bubble. Life expectation /GBD capita.**

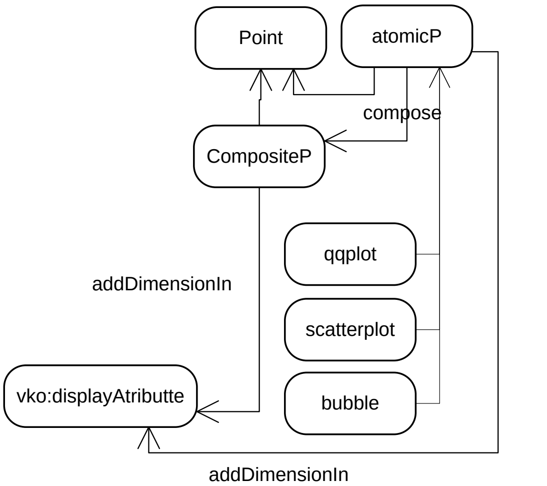



**Figure 24: Point charts in triples**

Dot plots, such as Q-Q plots (Figure 21), scatter plots (Figure 22) and bubble charts, (Figure 23) are used to show correlations between pairs of figures [13]. When distributions are compared, the goal is to understand how the distributions move from one dataset to the next [8]. The difference between the bubble and the scatter plots is that the latter may have another dimension quantitatively by size. In all, the color is an essential element to represent the class of the pairs or triplets of the data.

Figure 24 represents the diagram of the point class. It is an abstract class that represents both primitives as their containers. We also have AtomicP, the primitive object, and compositeP, the composite of point charts. As all charts shown are subclasses of compositeP. The composition still requires a property addDimensionIn related to a component of this class which can be a retinal mark.

## Maps

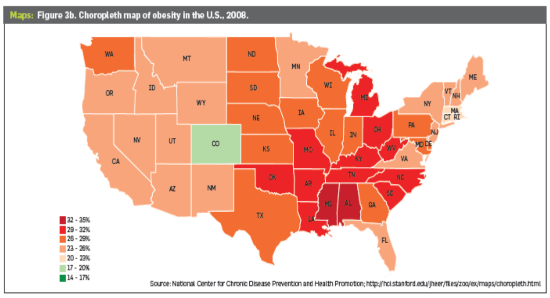



**Figure 25: Choropleth**

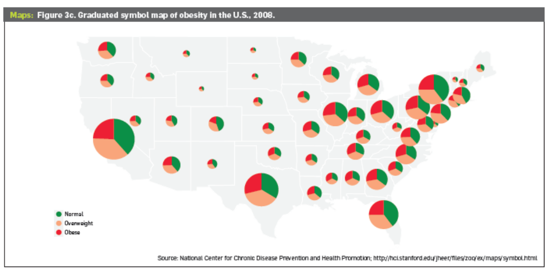



**Figure 26: Graduated Symbol**

In cartography, geographic component uses the two planar dimensions [4]. The quantitative or qualitative information, however, must be represented by one or more retinal marks, or through compositing with charts. Following the pattern we have an atomic map that uses the two planar dimensions and a third dimension must be added through retinal attributes to represent a quantitative variable. Figure 25, used brightness and hue together to represent percentage of obese people in the states of USA. Although, the quantitative data must have to be transformed in classes to allow that gradations were perceptible to the retina of human eyes. In this map, Heer merge hue and brightness to create a chromatic scale.

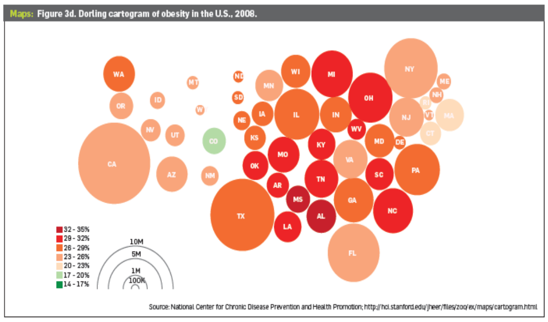



**Figure 27: Cartogram**

According to Bertin [4], a scale with variation only in brightness of a single hue would be more efficient because brightness is ordered and hue is selective. Figure 26 is a composite of map and chart to represent part of whole, the pieChart. In these charts the unit of measure is perceptual [13]. Figure 27 keeps only the center of the geographic component. The shape was changed to represent a quantitative dimension. In this case, data is not transformed, because size is being used to show a quantitative value and not a class.

In the next page, Figure 28 presents the whole diagram of the class Visualization of our ontology. It was divided between subclasses: network, geometry, and map. The first includes the hierarchical relations defined by trees presented before. The second has all types of Cartesian charts. Visualizations 3D were not detailed because as mentioned before, they are not efficient to information visualization. The third subclass defines the maps we have just presented. It also includes bar charts, which we did not detail in this section because histograms or bar charts are more common representations. We also include part–to-whole charts represented by pie chart because it was used on the map of Figure 26.

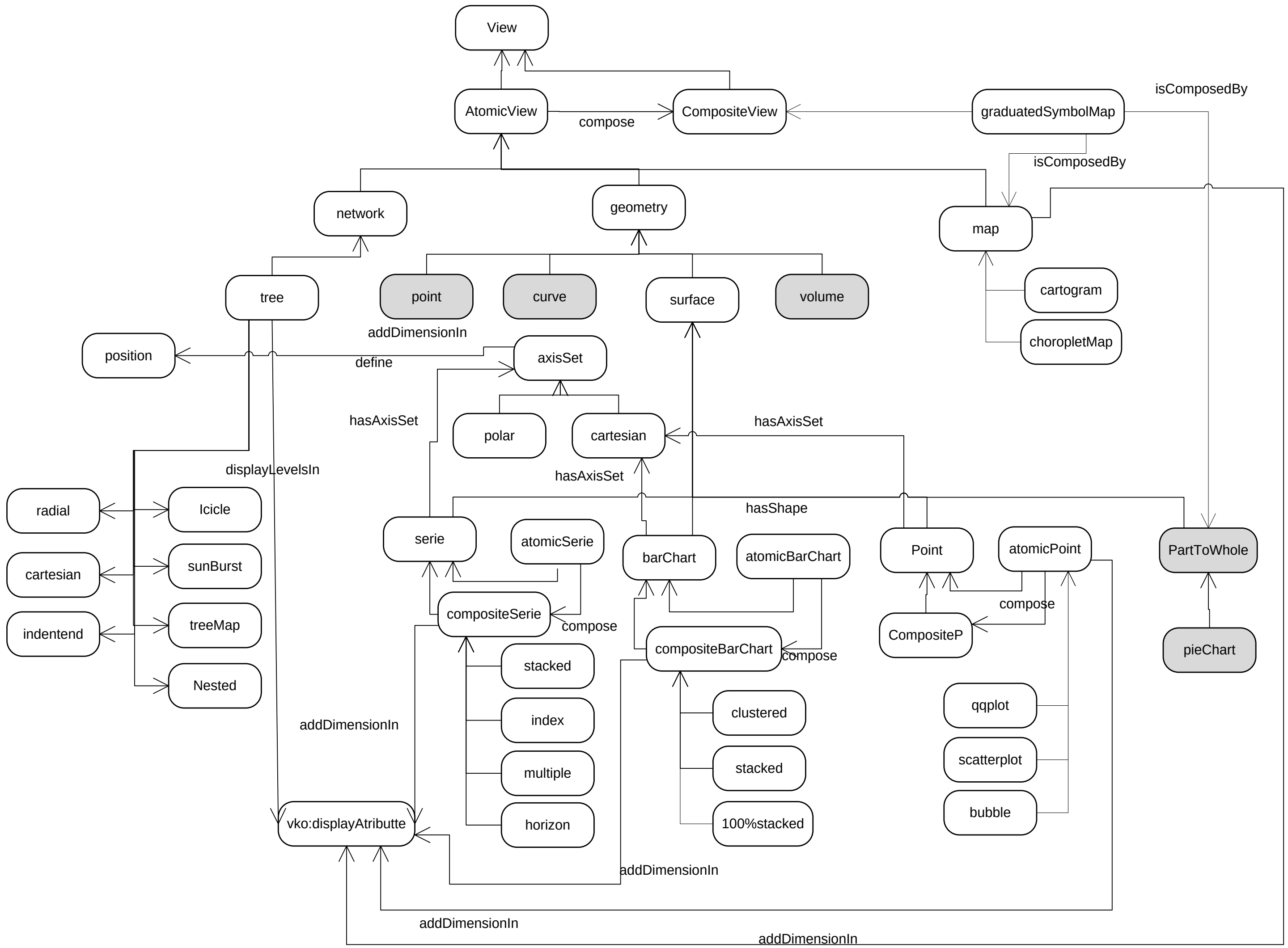



**Figure 28: Complete diagram of the visualization class**

## 3.4 Task Class

A task is related to the goal that the user wants to achieve with the data visualization and what he or she needs to do with it.

One criterion for efficient visualization is the formulation of questions [22]. The task is related to the question that user wants to answer and with the efficiency of the representation.

According to Bertin [4], the efficiency of a representation is defined by the following proposition: "If, in order to obtain a correct and complete answer to a given question, all other things being equal, one construction requires a shorter observation time than another construction, we can say that it is more efficient for this question" [4 p.139]. Bertin continues: "the most efficient constructions are those in which any question, whatever its type and level, can be answered in a single instant of perception, that is, in a single image." [4, p. 146].

Regarding the number of questions answered by each visual representation, Bertin [4] says there are as many questions as the number of dimensions. Each dimension produces a type of question.

Thus, we created the table below to illustrate the relationship between questions, visual representation and tasks, where:

- N is an indefinite number.
- M is a finite number. It is limited by the features of the attribute of the Display Attribute class. Example: If we define a variable in a brightness scale, as we have seen before, it can have 7 values. In this case, M = 7. In case we use size, M = 5.
- V is a set of variables {v1,v2,v3,...,vn}
- T is a time interval (t1,t2,t3,...,tn)
- C is a set of qualitative variables {c1,c2,c3...cm} or a set of quantitative variables {c1,c2,c3...cn}

**Table 4: Relating questions, tasks, and visual representations**

| **Question** | **relation** | **Visualization** | **relation** | **Task** |
|---|---|---|---|---|
| Is there any correlation between $\mathbf{v_1}$ and $\mathbf{v_2}$? | supportedBy | Scatterplot | relatedBy | Correlate **N pairs of quantitative** variables. |
| Is there any correlation between $\mathbf{v_1}$, $\mathbf{v_2}$and $\mathbf{v_3}$? | supportedBy | Bubble chart | relatedBy | Correlate **N triples of quantitative** variables. |
| What element **N** had the highest (or the lowest) value of a category **M?** | supportedBy | Clusttered Column Chart | relatedBy | Find extreme between **N quantitative** variables of **M categories** |
| Which category **M** had the highest (or lowest) value in time **t** | supportedBy | Index timeseries | relatedBy | |
| Which category **N** (place) had the highest (or lowest) value? | supportedBy | Cartogram | relatedBy | Find extreme between **N quantitative** values |

| Question | relation | Visualization | relation | Task |
|---|---|---|---|---|
| What element has the highest (or the lowest) perceptual value of a category**?** | supportedBy | 100% stacked column chart | relatedBy | Find extreme in perceptual value between **N quantitative** variables of **M categories** |
| What element has the highest (or the lowest) **sum** of its **M** categories? | supportedBy | Stacked column Chart | relatedBy | Derived value of sum from the **M** quantitative values of categories and find extreme between **N** qualitative variables. |
| What is the value of a variable $\mathbf{v_1}$ of class $\mathbf{m_1}$ in time **t**? | supportedBy | Index time series | relatedBy | Retrieve a quantitative value |
| When did variable $\mathbf{v_1}$ of $\mathbf{M_1}$ class have value **X**? | supportedBy | Index time series | relatedBy | Retrieve index of **N** values |
| | supportedBy only if (M==1) | Horizon time series | relatedBy | |
| Which categories had the highest (or lowest) variation of a category in the interval $\mathbf{t_1}$ to $\mathbf{t_n}$? | supportedBy (if has positives and negatives values) | Index time series Multiple time series | relatedBy | Visualize trends, changes, increase, fluctuation, growth, decline of a quantitative variable V of M classes |
| | supportedBy (if only positive values) | Stacked time series | relatedBy | |
| Which places belong to a category **M**? | supportedBy | choroplethMap | relatedBy | Compare **M** categories of a **qualitative** variable between **N** categories of a **geographic component** |
| In which category, place **X** is classified? | supportedBy | choroplethMap | relatedBy | |
| What is the perceptual of the category X in the place Y? | supportedBy | graduated Symbol Map | relatedBy | Visualize part of a whole of **M quantitative** values of **N** categories that is a **geographyc component** |
| How many level has the hierarchy? | supportedBy | Radial Icicle Sunburst | relatedBy | Find steps |
| Who are the children of category **N**? | supportedBy | Radial Cartesian Indented Icicle Sunburst Treemap Nested | relatedBy | Find subclasses |
| What are the leaves of the hierarchy? | supportedby | Cartesian Treemap | relatedBy | Find leaves |

| Question | relation | Visualization | relation | Task |
|---|---|---|---|---|
| What is the most general class? | supportedby | Radial<br>Cartesian<br>Icicle<br>Sunburst | relatedBy | Find the root |
| Which subclass has the highest (or lowest) value of variable X? | supportedby | Treemap<br>Nested | relatedBy | Find extreme of a quantitative variable N between N categories (leaves) |

## 3.5 Transformation Class

Functions of transformation class define the dimensions that can be designed in each display attribute to create an efficient visualization to answer the corresponding question.

The property isMappedOn maps quantitative data into qualitative. This transformation will happen whenever you need to display such data as a category. The transformation equally distributes the data into categories, called quartiles [19]. Qualitative classes must also be mapped onto display attributes following some priorities. For instance, time values must be displayed in horizontal bars. Time series charts always use the horizontal axis for the time scale and the vertical axis for the quantitative scales [13]. Figure 29 shows some transformations in triples.

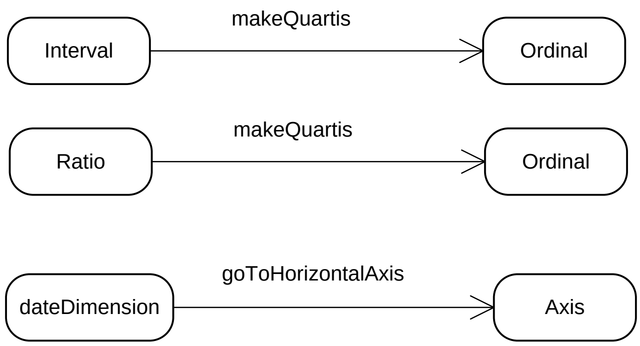
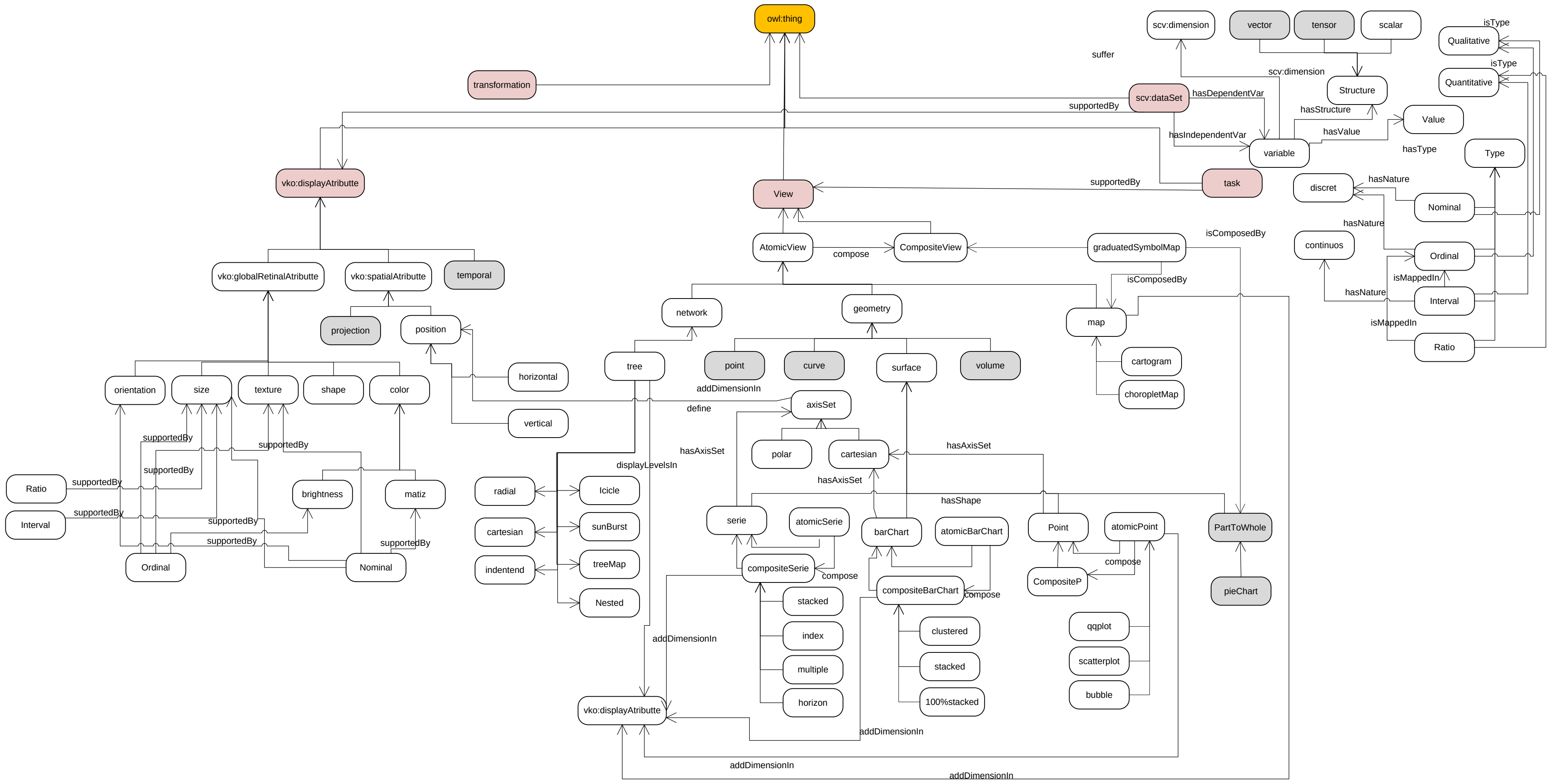



**Figure 29: Sample Transformations**

## 3.6 Ontology Overview

The diagram of Figure 30 depicts the data, display attribute, and visualization classes, and their relations. It shows an overview of the proposed ontology . The task and transformation classes are not detailed because they are defined at a level of abstraction too low to represent in this diagram.




## Figure 30: Proposed Ontology

## 4 Conclusion

Two main obstacles hinder the creation of ontologies for visualizations. First, understanding a visual representation depends on cognitive factors. Second, there are numerous possibilities to represent data using the eight variables that graphics offer (the two planar dimensions, color, brightness, texture, orientation, size and shape).

Using a limited number of graphical representations, we searched components that must be included in a visualization ontology. Although an ontology provides a vocabulary by which users and systems can communicate between themselves, it is only useful if it reflects the consensus of a community. Therefore, the proposed ontology should be tested with new visual representations to verify the breadth of its usefulness and its degree of interoperability.

Our next step is to identify questions that have not been covered in this work. For future work, we also propose the use of other visual representations to extend this ontology. We intend also to add features for interaction and animation to formalize the way people interact with the proposed visualization, to help them communicate with information systems.

## Acknowledgments

We thank Daniel Schwabe for valuable discussions that contributed to this work, CNPq (313031/2009-6 e 557.128/2009-9), FAPERJ (E-26/170028/2008 - Programa INC&T - Project "Instituto Brasileiro de Pesquisa em Ciência da Web"), VRAc/PUC-Rio and IBGE for the provided support.

## References

[1] Allemang Dean, Hendler James. What is the Semantic Web? In: Semantic Web for the working ontologist: effective modeling in RDFS and OWL. Second Edition. Elsevier 2011. pp. 1–12.

[2] Atlas Geográfico Escolar./IBGE. 4ª Edição. Rio de Janeiro, Brasil. IBGE 2007. pp. 21-24

[3] Ayers, Danny; Feigenbaum, Lee; Halb, Wolfgang; Hausenblas, Michael; Heath, Tom; Raimond, Yves. The Statistical Core Vocabulary (scovo). Available at http://sw.joanneum.at/scovo/schema.html [last access on 07/Dec/2011]

[4] Bertin, Jacques. Semiology of Graphics: Diagrams, Networks, Maps. 1918. ESRI first edition 2011, USA

[5] Brandão, Rita M. Conceitos básicos em estatística. In: Introdução à Estatística. Departamento de Matemática da Universidade de Açores. Available at http://www.pgarrao.uac.pt/IntEstatistica_08_09/AcetatosCap1_1parte.pdf [last access on 15/Nov/2011]

[6] Brodlie, Ken W.; Noor, Nurul M. Visualization Notations, Models and Taxonimies. In: Lim, I.S., Duce, D. (eds.) EG UK Theory and Practice of Computer Graphics, pp. 207–212 (2007)

[7] Chang, Jiae. Perceptual Visualization Assistant for multi-dimensional data visualizations. Master's Thesis in Computer Science, North Carolina State University (2001)

[8] Cleveland, William S. Visualizing Data. AT&T Bell Laboratories, Murray Hill, New Jersey (1993).

[9] Cyganiak, Richard; Reynolds, Dave; Tennison, Dani. The RDF Data Cube vocabulary. W3C Working Draft (2012). Available at: http://publishing-statistical-data.googlecode.com/svn/trunk/specs/src/main/html/cube.html [last access on 07/Dec/2011]

[10] Del Rio, Nicholas; da Silva, Paulo P. Visualization Queries. Computer Engineering Commons (2010). Available at http://trust.utep.edu/visko/docs/docs/visquery.pdf [last access on 12/Nov/2011]

[11] Duke, David J., Brodlie, Ken W. and Duce, David A. Building an Ontology of Visualization, Proceedings of the conference on Visualization '04, IEEE Computer Society, 2004

[12] Duke David J., Brodlie, Ken, Duce David A., Herman, Ivan: Do you see what I mean? IEEE Computer Graphics and Applications 25, 3 (2005), 6–9.

[13] Few, Stephen. Show me the numbers. Designing tables and graphs to enlighten. Analytics Press, Oakland, California. 2004

[14] Freitas, Carla M. D. S.; Chubachi, Olinda M.; Luzzardi, Paulo R. G.; Cava, Ricardo A. Introdução à Visualização de Informações. Revista de informática teórica e aplicada. Porto Alegre. Vol. 8, n. 2 (out. 2001), pp. 143-158

[15] Heer, Jeffrey; Bostock, Michael; Ogievetsky, Vadim. A Tour Through the Visualization Zoo. Communications of the ACM Volume 53, Number 6 (2010), pp. 59-67

[16] Ignatius, Eve; Senay, Hikmet. A Knowledge Based System for Scientific Data Visualization. (1992) Technical Report TR-92-79. CESDIS, Goddard Space Flight Center, Code 930.5, Greenbelt, MD 20771.

[17] Moody, Daniel L. The "Physics" of Notations: Towards a Scientific Basis for Constructing Visual Notations in Software Engineering. IEEE Transactions On Software Engineering, Vol. 35, No. 5, November-December 2009

[18] Brodlie, Ken W.; Duce, David A. ; Duke, David J. and al. Visualization Ontologies: Report of a Workshop held at the National e-Science Centre. Technical report, e-Science Institute, 2004. Available at http://www.nesc.ac.uk/talks/393/vis_ontology_report.pdf [last access on 13/Nov/2011]

[19] Piegel, Murray R. Estatística. McGraw-Hill do Brasil, São Paulo, 1985.

[20] Pinker, Steven. A theory of graph comprehension. In R. Freedle (Ed.), Artificial intelligence and the future testing. Hillsdale, NJ: Erlbaum. (1990).

[21] Reis, Marcelo M. Conceitos elementares de estatística. Available at http://www.inf.ufsc.br/~marcelo/intro.html [last access on 15/Nov/2011]

[22] Steele Julie; Iliinsky, Noah. Beautiful Visualization. Looking at Data through the Eye of Experts. O'Reilly, 2010.

[23] Tory, Melanie; Möller, Torsten. Rethinking visualization: A high-level taxonomy. In Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on, pages 151–158, 2004.

[24] Visko Visualization Knowledge. Available at http://trust.utep.edu/visko/dl/ [last access on 13/Nov/2011]