Hyperglyphs: Exploring the Limits of Glyph Structure to Improve Visual Analytics Methodologies

Draft for Submission to PEARC 2018

The concept of a glyph in scientific visualization is well known and has found numerous applications over the years. However, the limits to the level of complexity of glyph structure have only begun to be fully explored. At the same time, a growing percentage of the big data torrent consists of semi-structured, unstructured, and non-traditional data, presenting a challenge for conventional visualization methods. Some of these types of data are so complex it is difficult for researchers to know where to begin in seeking insight into trends and anomalies hidden within. We need new and innovative ways to explore such massive amounts of complex data.

In this tutorial we provide a brief history of glyphs in scientific visualization and conditions in which their use is appropriate and beneficial. We make the case that these conventional, simple glyphs should be extended and complexified into what we call ‘hyperglyphs’, highly complex visual structures designed to encapsulate much more information within a single glyph and which, when thousands are arrayed in 3D space and with user training and adaptation, can significantly enhance perception and information assimilation leading to new knowledge and insight. We show how our work has led to the creation of successful startup commercial ventures in the field of visual analytics. We provide a wide range of examples from diverse fields including education, physiology, meteorology, public health, and social media. Participants will have the opportunity to get hands-on experience with interactive glyph design using free open source tools.

## **Background: The Goals of Data Visualization**

"*No, I can't explain the dance to you; if I could say it, I wouldn't have to dance it!*"

— Isadora Duncan

### **The Big Data Glut**

In 2009 WIRED Magazine declared "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." [1] Perhaps that's going a little too far, but we have definitely entered the era of "Big Data," and there's no turning back. The costs of collecting and storing data continue to drop, and smaller organizations are now facing the same kind of data glut that once only challenged large corporations and governments.

### **Getting to the Right Questions**

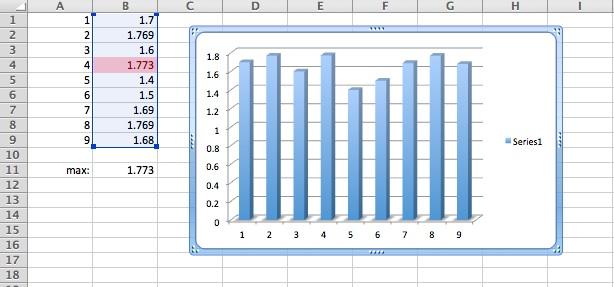
Many of the difficulties occur at the beginning, when the organizations first start to examine their archived data. At this point it is an open question what to look for. How do you get answers when you don't know the questions? This is precisely the sort of job that visualization can best address.

When you know what you are looking for, simple algorithms such as statistical tests can get you an answer quickly without human intervention. For example, given this list of numbers:

1.7  
1.769  
1.6  
1.773  
1.4  
1.5  
1.69  
1.769  
1.68

How do you find the maximum? In today's world there are probably thousands or even millions of ways to solve this problem using available tools. If you were going to find the answer manually, you would go through the list asking yourself repeatedly, "Is this number bigger than any I have seen before?" and noting the new number if the answer was yes. Working through this list you find 1.773 is the maximum. But if the list gets overly long, or you have many lists to process, or you want to reduce human error, a more automated approach is needed.

An approach would be to load the list into a spreadsheet program like Excel, OpenOffice or Google Docs (Figure 1).

******

***Figure 1: Finding the maximum value in a list with Excel***

Then a simple formula will identify the maximum value in the column:

|  |
| --- |
| cell B11:  =MAX(B1:B9) |

For every such statistical question you can ask there are many ways to get answers from the numbers in the "big data" archives. But you have to know what questions to ask! Hyperglyph visualization tools can help you *before* you know the questions.

### **Common Uses of Data Visualization**

In our work with users we have identified these main uses for data visualization:

* **data exploration and discovery**
* Often a "first look" at data with a visualization tool can reveal unsuspected patterns at the outset.
* **as a planning tool for other analysis**
* Once data patterns are identified in a preliminary visualization this often leads to more focused questions and suggests new lines of inquiry.
* **in an iterated loop of analysis and display**
* Once the relevant questions are identified and statistical tools are used to process and summarize data, visualization can be useful in an iterated loop of analysis and display, helping to hone in on the best answers in both categories.
* **to display results of analysis**
* Once conclusions are reached a visual display can be useful to help explain results to clients, investors, funding sources, partners and other stakeholders.

The big data glut has introduced a challenge which is difficult to address using traditional tools [ Data exploration with no specific question can be costly, time-consuming, with no guarantee of return on investment so tools which can expedite the process of finding the proverbial ‘needle in the haystack’ are of extreme value. The ability to visualize massive amounts of unstructured data is becoming an important capability but it is still not very common in traditional visualization applications.

### **What is a Glyph?**

First, we need to better understand the concept of a glyph before we consider hyperglyphs. Thankfully, some excellent work has already been done to characterize the concept of a glyph in data visualization by Lie et al and Borgo et al [2,3]. Borgo et al [3] list features of a glyph which include:

(a.1) a glyph is a small independent visual object that depicts attributes of a data record;

(a.2) glyphs are discretely placed in a display space; and

(a.3) glyphs are a type of visual sign but differ from other types of signs such as icons, indices and symbols.

Or more broadly:

(b.1) a glyph is a small visual object that can be used independently and constructively to depict attributes of a data record or the composition of a set of data records;

(b.2) each glyph can be placed independently from others, while in some cases, glyphs can be spatially connected to convey the topological relationships between data records or geometric continuity of the underlying data space; and

(b.3) glyphs are a type of visual sign that can make use of visual features of other types of signs such as icons, indices and symbols.

### **A Brief History of Glyphs**

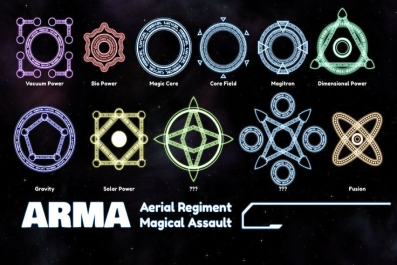
The English word ‘glyph’ has been used since the 18th century in reference to a carved or inscribed symbol, similar to a pictogram. More recently it has been used in reference to basic font elements. Other terms such as ‘grapheme’ or ‘morpheme’ are related.

The basic concept of a glyph as a graphic symbol began several thousand years ago with petroglyphs followed by more advanced graphic primitives including Egyptian, Mayan, and Enochian hieroglyphics (Figure 2).

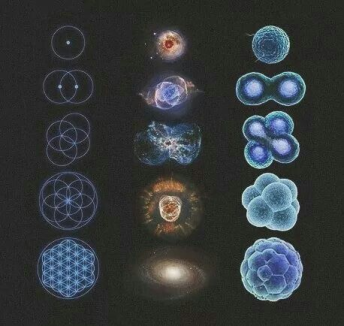
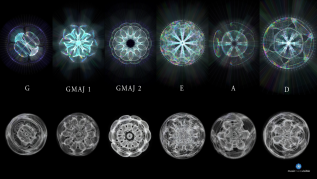
***Figure 2: Petroglyphs and heiroglyphs***

Today, glyphs have become somewhat ubiquitous. They can be found in street signs, video games, and of course social media in the form of emojis (Figure 3).

***Figure 3: video game glyphs (left) and emojis (right)***

All of these in one way or another fall within the category of ‘glyphs’ because they contain information which conforms to a convention either explicitly or metaphorically so people can understand and use them. Nature and its underlying pure mathematics can manifest organic structures which take on glyphic forms and inspire meta-glyph structures in art, science, and even religion (Figure 4).

***Figure 4: Symbol-like images from nature (left) and pure mathematics (right)***

Figure 5 below shows the frequency of occurrence of the term “glyph” since its initial inception.



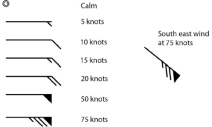
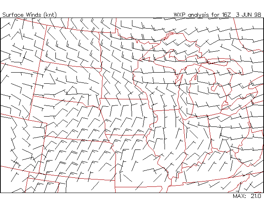
***Figure 5: Frequency of occurrence of ‘glyph’ over time***

### **The Arrow Glyph**

In scientific visualization, an arrow glyph is often used to show the magnitude and direction of a vector flow. Glyph features such as size, color, and orientation can be put to effective use in revealing complex structures in scientific datasets. However, these are typically a single object representing a single data value in a 3-dimensional space.

### **Weather Glyphs**

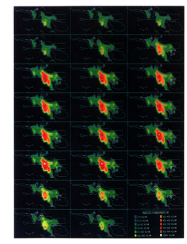
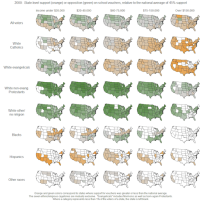
The glyph perhaps most familiar with the general public is the “wind barb” on a weather map designed to represent wind speed and direction at various locations for a specific time (Figure 6).

***Figure 6: Weather glyphs***

**The Principle of Small Multiples**

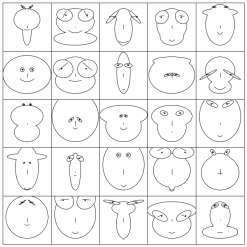
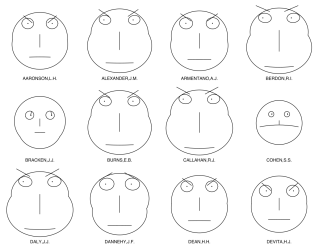
Edward Tufte [4] elaborated on the value of visualizing arrays of small visualizations, each of which were similar but with subtle variations in change, making it easy for our built-in pattern recognition skills to reveal patterns, often unexpected, in the data. He describes this as the Principle of Small Multiples. Some examples are shown below (Figure 7).



***Figure 7: Principle of small multiples***

**Chernoff Faces**

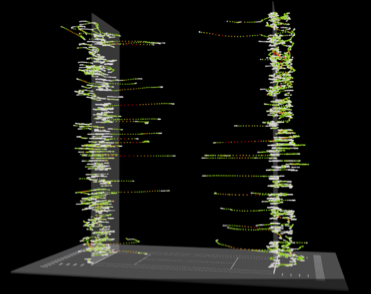
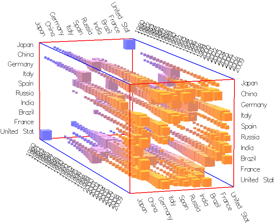
Chernoff Faces leverage the human mind’s ability to quickly differentiate one facial expression from another. Typical facial features including eye location, mouth expression, and feature distribution were used to represent subtle variations in data. Examples are shown in Figure 8 below.



***Figure 8: Chernoff Faces***

**Space-Time Cubes**

Hyperglyphs are useful for displaying spatially located data that changes over time, especially geo-located (positioned on the Earth). Humans have been using maps of regions on Earth for at least about 4300 years, since at least the dynasty of Sargon of Akkad, but for most of that time only dots, or colored dots, or pins, or possibly flags have been used as markers and which are unable to render changes over time. Only recently, applications such as GeoTime, Datascape, and ArcGIS include ways of visualizing complex 3-dimensional spatiotemporal data. Examples are shown in Figure 9 below.



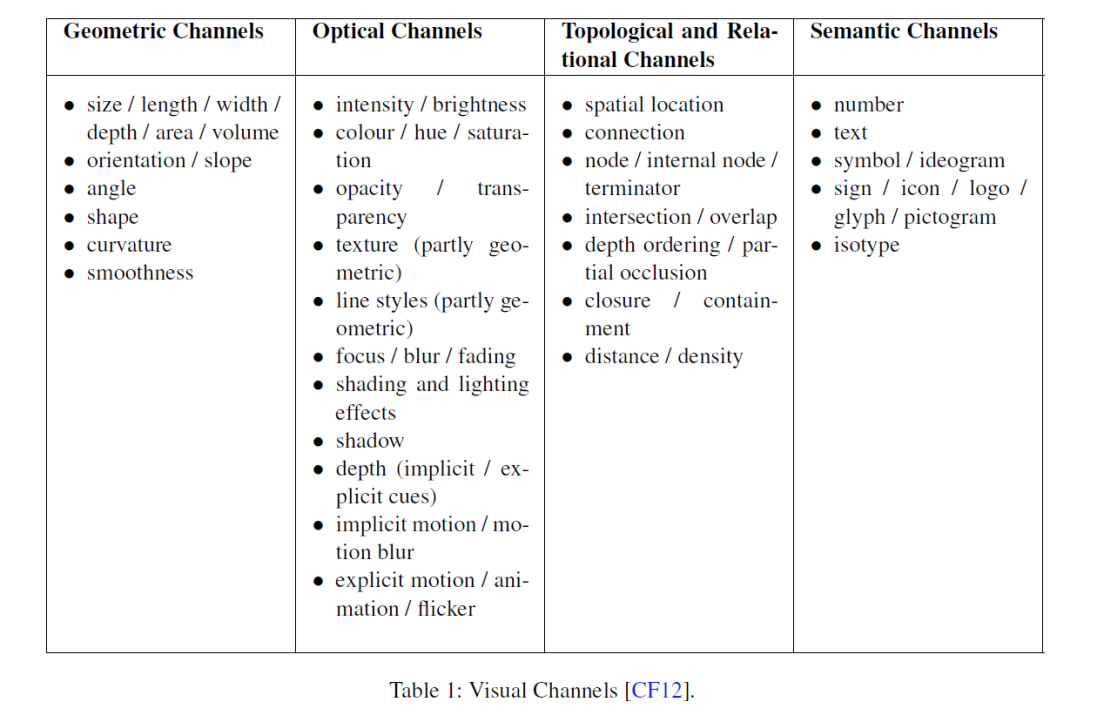
***Figure 9: Space-time Cubes***

### **What is a Hyperglyph?**

In this tutorial we will extend the concept of a glyph to include features and capabilities such as nested object ‘parent-child’ relationships, dynamic animation, and interactivity. This extended concept of a glyph leads us to the term “hyperglyph” to emphasize the added complexity of representing potentially many variables or dimensions in a single glyph.

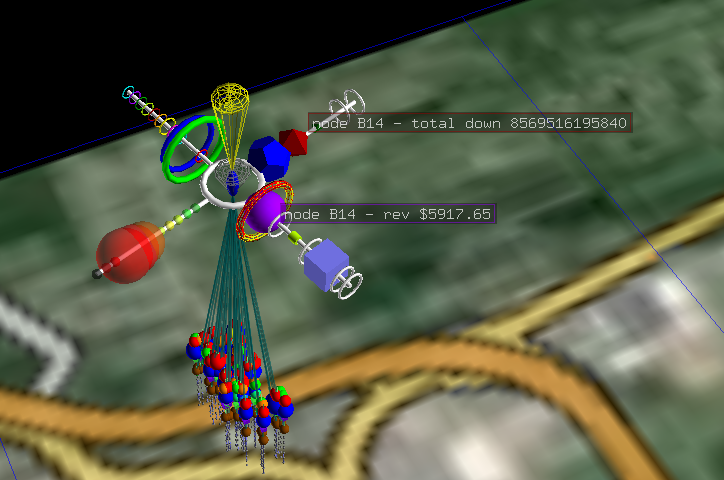
One way to describe a hyperglyph might be “a multimodal information packets consisting of a wide range of perceptual properties, currently dominated by visual information.“

Borgo et al [3] list parameters they call ‘visual channels’ summarizing the many ways in which a glyph can be varied. Their list is shown in Figure 10 below.



***Figure 10: Visual Channels***

One of the innovations of hyperglyphs is to use aggregates of simpler geometric primitives in a coordinated design to display complex geo-located data. This allows higher-dimension data, and data with higher complexity, to be displayed in a geo-located context. For example, the glyph shown below in Figure 11 is from a visualization of internet customers in a small town, showing the nodes through which data passes, and allowing a number of values indicating customer type, data rates, and revenue to be displayed at the same geographical point.



***Figure 11: Glyph design for internet node traffic data***

### **Begin With the End in Mind**

One important lesson we have learned is that even though you *may not know what you are looking for*, you should still make sure to *know what your choices are*. When you begin a project of data exploration, know what decisions in the future will be influenced by the results. If data analysis isn't determining actions, it may be a waste of effort. Be sure to identify the choices you have going in.

### **Dominant Factors to Consider in Glyph Design**

* Number of children
* Inter-glyph and intra-glyph spatial arrangement
* Spacing and scale
* Ratios proportions and distribution of glyph elements
* Color palettes and color ranges
* Understanding the hierarchical structure
* Why are the glyph elements arranged the way they are?
* What is the rationale for the arrangement
  + Visualizing a process
  + Visualizing a thing
  + Visualizing the changes to the thing
  + Visualize the processes that change the thing
* Deterministic structure can be rule specific/ topic specific
* Creating visual lists
* Temporal layout -- time lines vs time rings vs time spiral
* The use of texture maps for disambiguation and categorical distinction
* what data is the glyph going to contain
* what is the range of each data element (max min mean mode etc) that is going to be mapped to each glyph element
* who is the glyph for… different levels of understanding of the data should influence the glyph design and levels of complexity
* how many glyphs will be grokked at the same time?

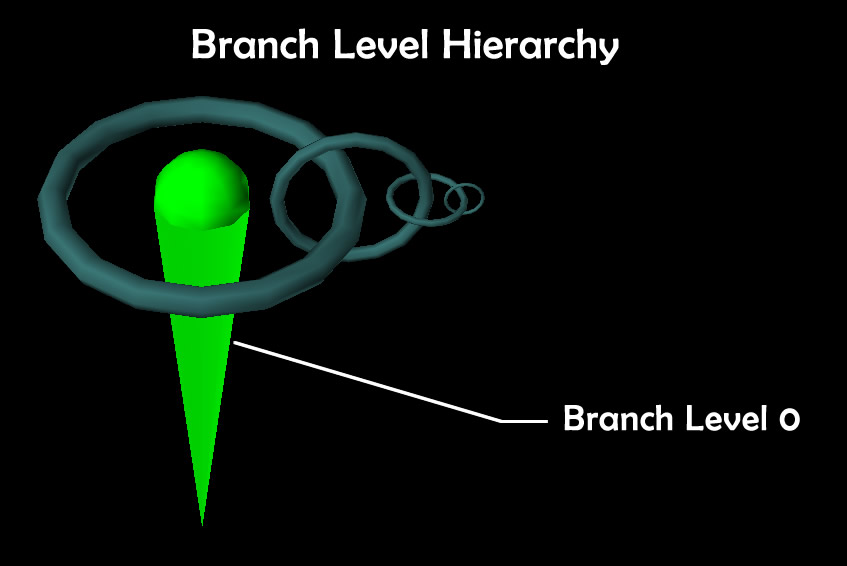
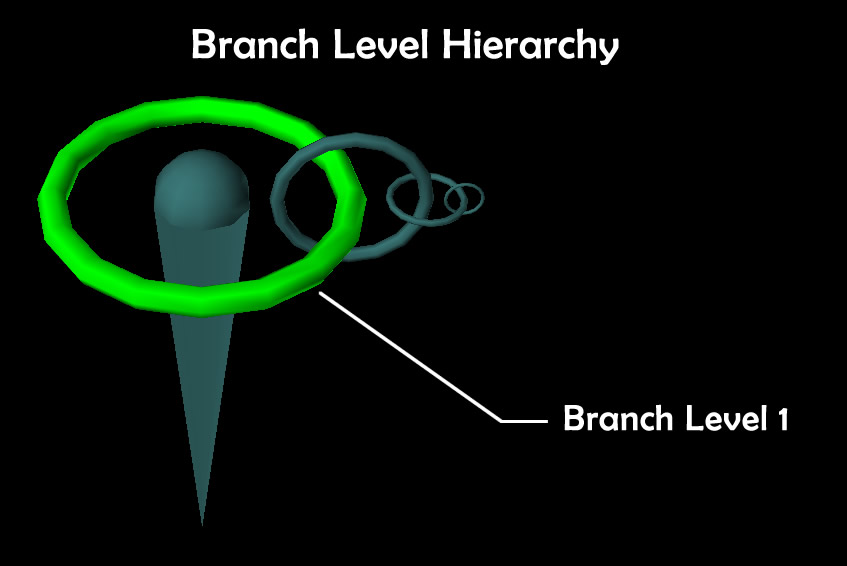
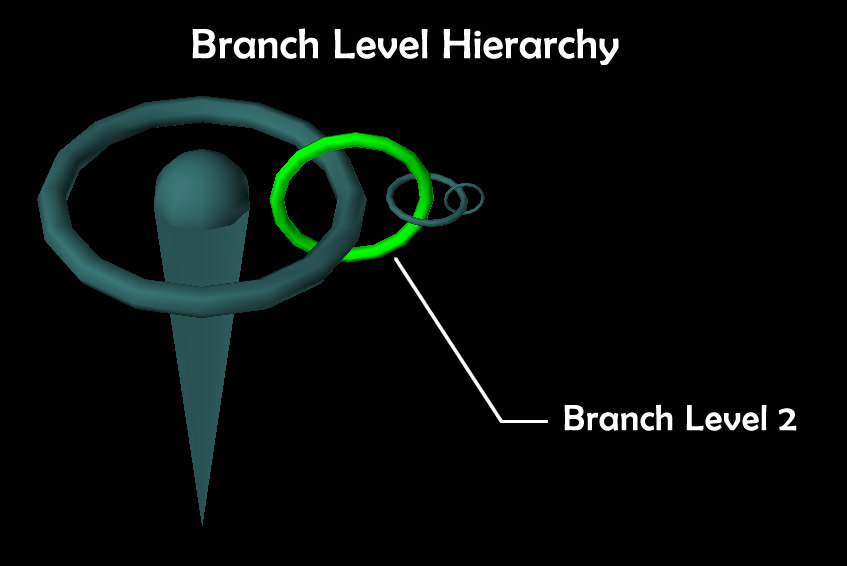
### **Visualizing the Forest AND the Trees**

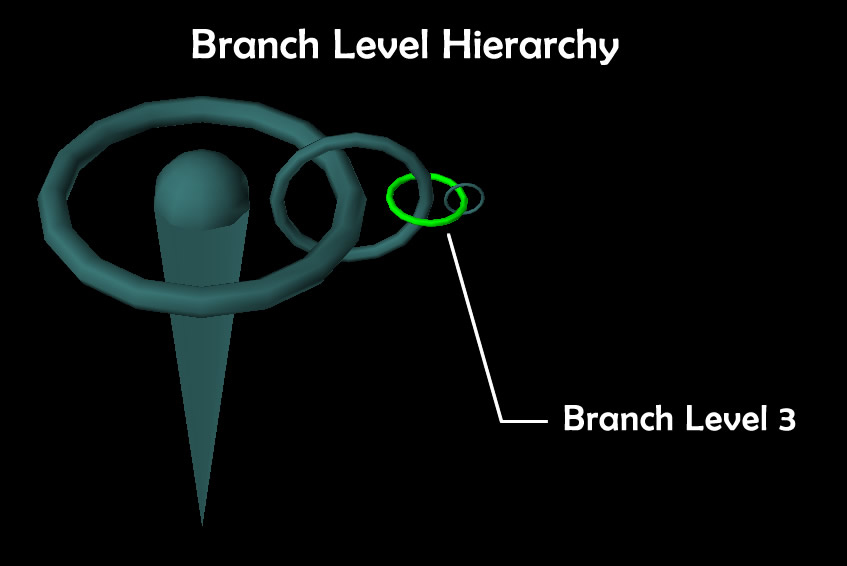
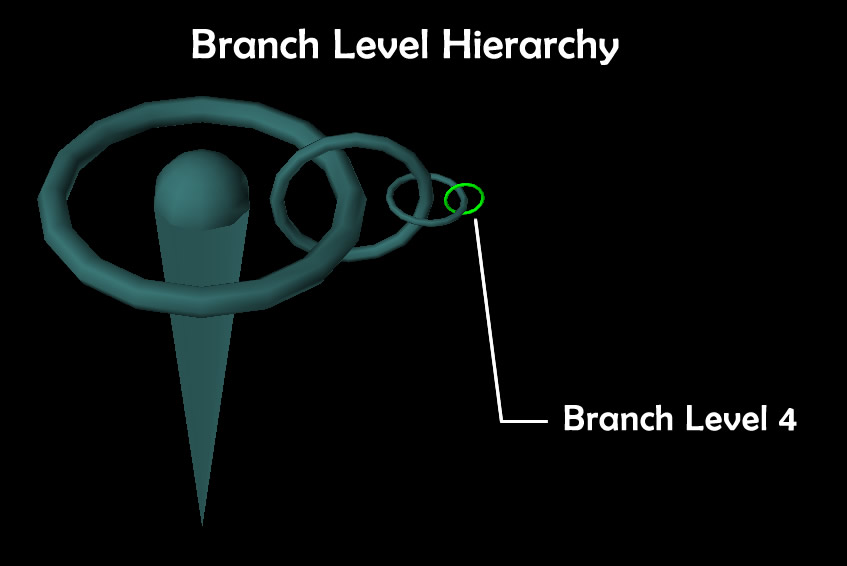
One of the most important skills for any visualization method is to learn the conventions for interpretation. When using hyperglyphs, a situation is being presented in which one is exploring both the forest and the trees. The “forest” consists of considerations of spatial topological distribution of data in spaces which are not necessarily Cartesian. The “tree” consists of a complex self-contained structure of Information, in this case graphical in nature. By distributing massive numbers of these complex hyperglyphs or ‘trees’ we blur the boundaries between forest and tree. In a sense, the tree becomes the forest!

**Hyperglyph Branch Levels**

The best tool to demonstrate what we mean by the term hyperglyph is the ANTz visualization application. ANTz, as well as its SynGlyphX commercial spinoff, includes built-in features which are designed to use a variety of simple geometries to build a complex cell-like 3D geometric structure or hyperglyph. The ANTz hyperglyph uses a parent-child relationship between geometries and supports up to 28 generations or ‘branch levels’ by default. We have not attempted a real-world data application which goes beyond branch level 5, but artistic explorations can be quite spell-binding at deeper levels.

ANTz supports up to 28 Branch Levels. Objects on one Branch Level are Parents to the objects on the Branch Level 'below' it. Examples of objects on Branch Levels 0 through 4 are shown in Figure 12 below.

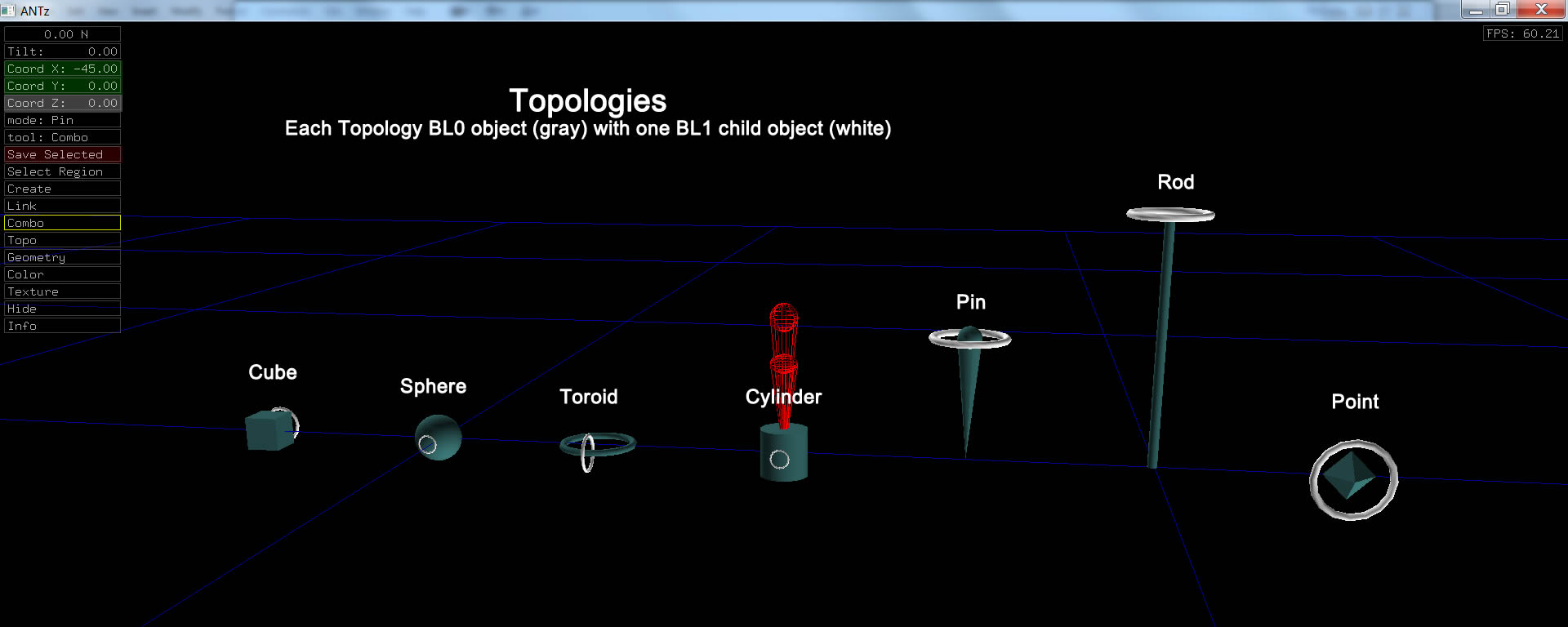
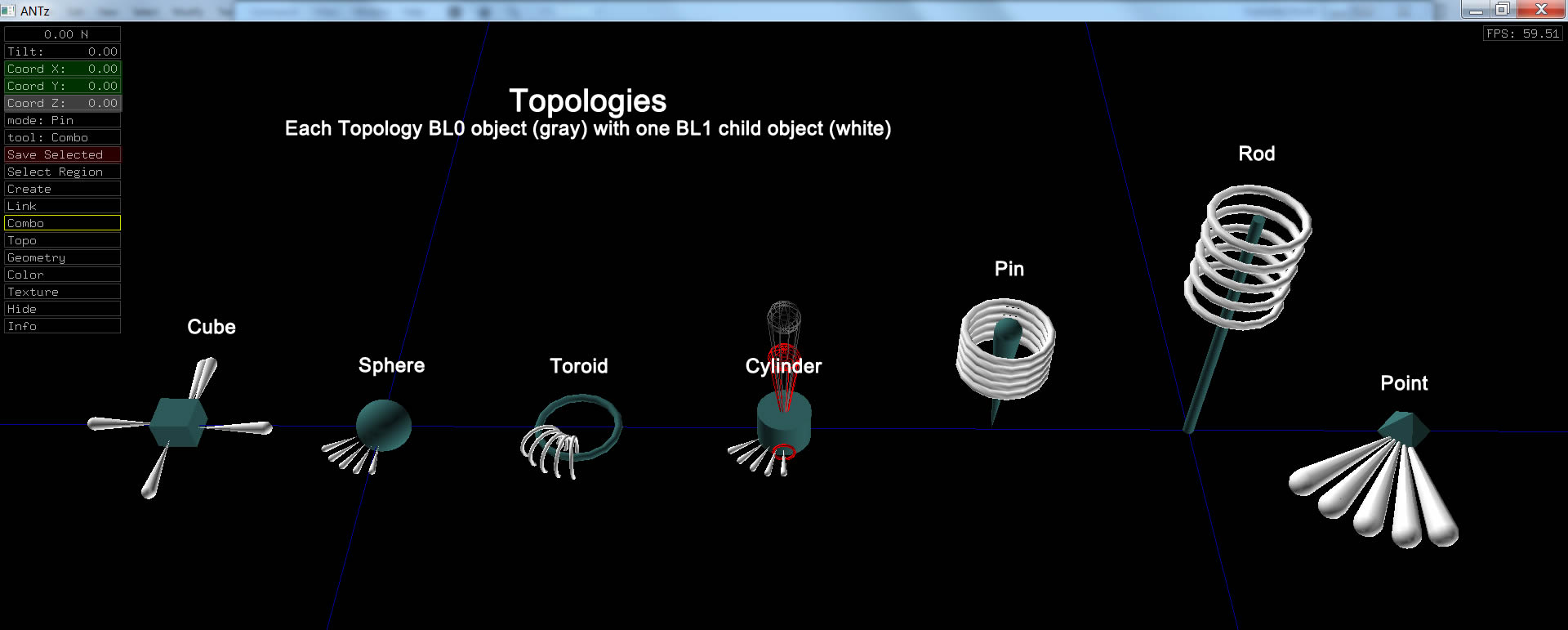
  

***Figure 12: Examples of objects on branch level 0 through 4***

**Hyperglyph Topologies**

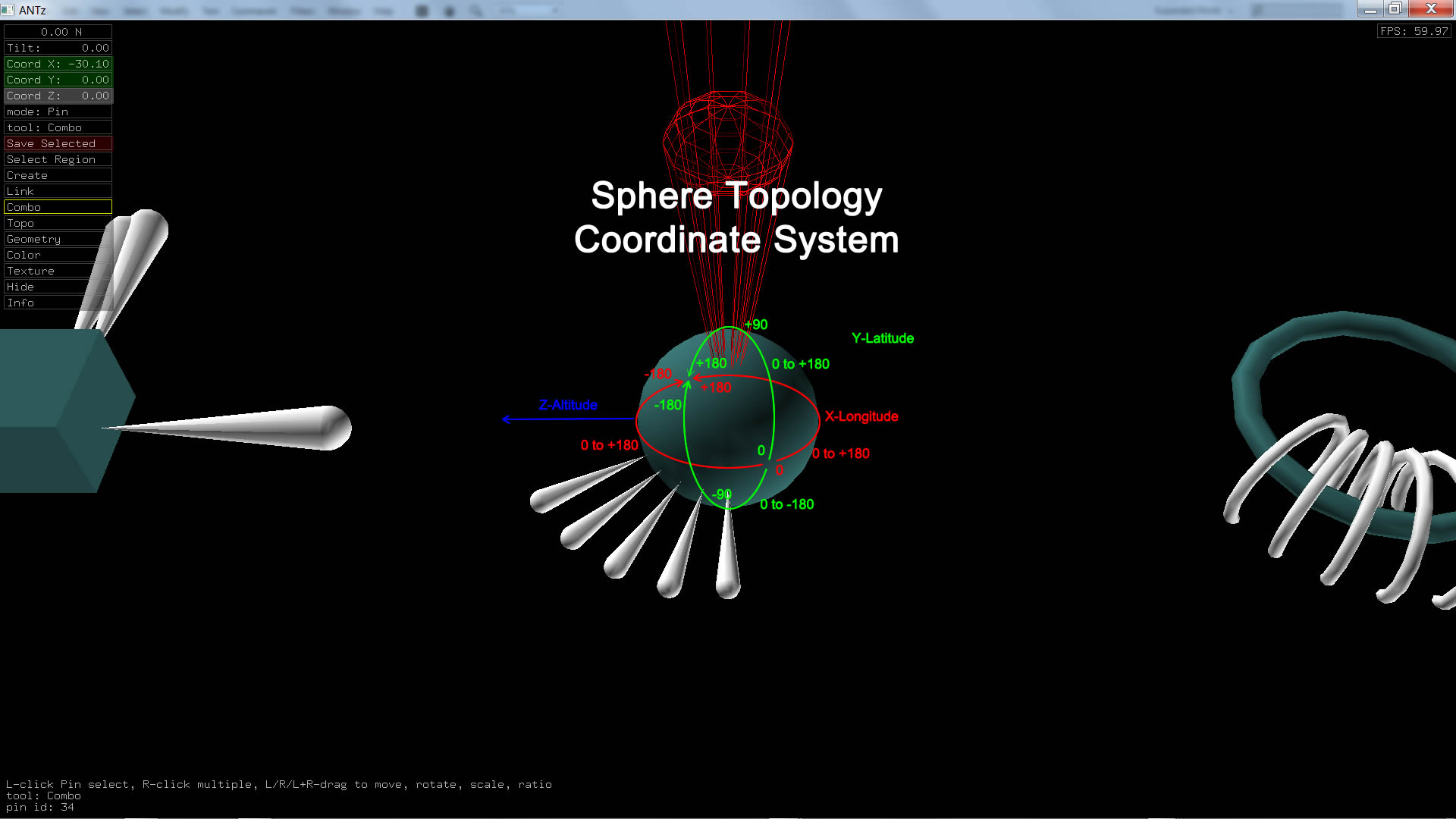
ANTz hyperglyphs use a set of 8 'topologies' to provide a diverse but consistent set of coordinate systems which constrain how we can distribute objects around their parent. ANTz topologies use the Google Earth (formerly Keyhole) Keyhole Markup Language (KML) spatial coordinate system for all 7 topologies. This means by default objects are distributed in ANTz and around their parents using a longitude-latitude-elevation convention for x-y-z spatial distribution by default.

[](http://www.edworlds.com/antz/toroids/tutorials/antz_topologies_one_child.zip) [](http://www.edworlds.com/antz/toroids/tutorials/antz_topologies_five_children_mac.zip)

***Figure 13: ANTz Topologies***

**Example of a Topological Coordinate System: the Sphere**

The coordinate system for the sphere is the simplest example of how the KML format is used. Child objects translate\_x, translate\_y, and translate\_z parameters correspond to longitude, latitude, and elevation respectively.



***Figure 14: Sphere Topology***

## **Examples**

The following examples are created using ANTz data visualization application. ANTz [9] is a 3D data viewer for interactive visualization of complex datasets. Spatiotemporal cues based on shape, color, size, texture and time enable the user to identify patterns for further exploration. The following examples will help illustrate the utility of the techniques of hyperglyph-based visualization. The workshop will provide detailed step-by-step instructions on how to work with these examples.

Technology pioneer Jaron Lanier, who coined the term *virtual reality*, also came up with the term *microexperiments* [6] to describe the way we interact with our environment. If you pick up an object and rotate it back and forth as you look at it, this activity involves muscle motion and vision working in a coordinated way. We are very skilled at receiving information about our environment this way. Analogously, when we interact with a 3D model on a computer screen by moving a mouse or other pointing device, we are using these same skills to receive information about simulated objects, and it works very well. Whenever possible we recommend experiencing the demos directly, but for now this static presentation will have to serve.

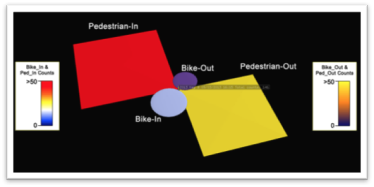
Mark Pesce, creator of VRML, referred to the ideas behind ANTz as ground-breaking for synergetic modulation of immersive data visualization in a perceptual cybernetic loop [7,8].

## **Arlington County Bicycle and Pedestrian Traffic**

#### **Problem Statement**

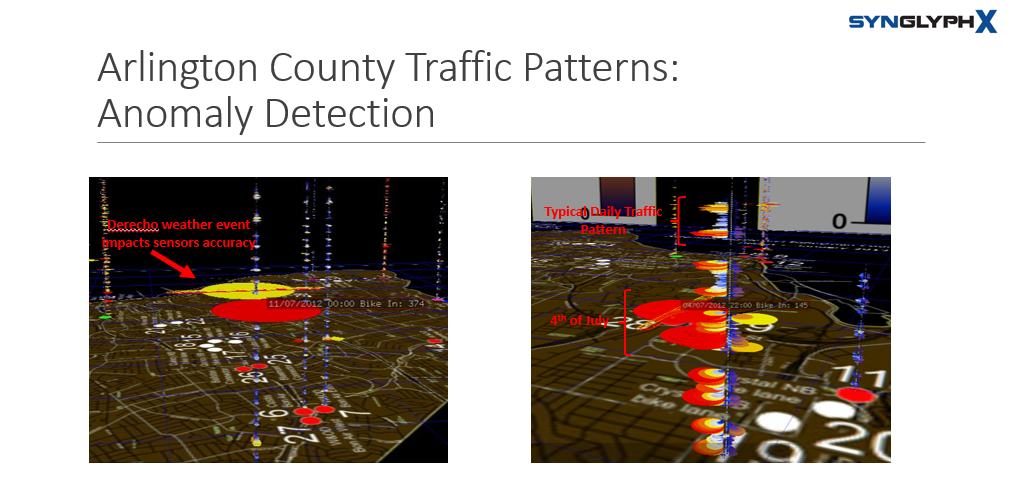
Arlington County, Virginia, just west across the Potomac River from the United States Capital, has an excellent network of bicycle trails. Electronic sensors capture information about how many bicycles and pedestrians pass by key points, as well time measured in 15 minute increments. Over time this data has accumulated in sensor log files that are usually viewed in spreadsheet form, if at all. An administrator of this data complained that he was "drowning" in it. We took the challenge to visualize some of the data with our glyph-based tools to see what patterns might emerge.

#### **Hyperglyph Design**



For this problem the only values in the data were location, date/time, and volume of riders/pedestrians. Location would map to a horizontal position, overlaid on a map of Arlington County, and date/time would determine vertical position. The type (bike or pedestrian) and the count for each 15-minute interval were the remaining data values, and they were mapped onto two glyphs: spheroid disks for bikes and cubes for pedestrians. Both the size and color for each glyph were based on the count (bigger and red = more traffic; smaller and blue = less traffic).

#### **Visualization**

The resulting visualization is shown below. Each vertical tower of hyperglyphs represents a "timeline" of events for the particular spatial location of sensors. Interactive examination of the resulting 3D model revealed several features: daily patterns, weekly patterns, weather-related patterns and holiday-related patterns.****

#### **Observations**

It became clear that it was possible sometimes to distinguish commuter traffic, which has a morning and an evening peak, with recreational traffic, which has a single midday peak. As expected big holidays such as the fourth of July brought out more recreational trail users. A large anomaly was also discovered — a huge number of bicycles on a single date in the middle of the night — that was traced back to a sensor failure due to heavy rains and wind. The hyperglyph visualization approach suggests further questions, such as which days tend to conform more or less to the morning-evening bicyclic behavior and why.

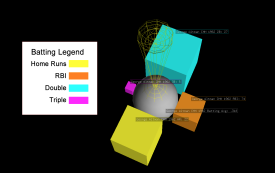
### **Lahman Baseball Database**

#### **Problem Statement**

Professional sports is a huge source of big data. The goal of this study was to identify features of interest which either provide new insights or confirm existing insights.

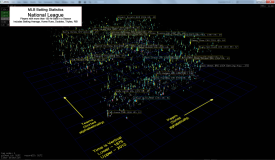
#### **Hyperglyph Design**

Our design idea was to attempt to visualize the entire American and National League data for batting with a minimum of 100 at-bats per season for every player. We simplified the effort to five parameters, batting average, home runs, RBI’s, doubles, and triples. This results in a hyperglyph with 5 glyphs. The hyperglyph is shown below.



#### **Visualization**

The resulting visualization consisted of over 150,000 glyphs and 30,000 hyperglyphs. Career batting data for more than 4000 players was visualized.



#### **Observations**

A distinct gap during the period of World War II was immediately apparent. The transition from home runs being infrequent to frequent was also observed. The use of performance-enhancing drugs may explain a variation in a player’s performance over the course of their career. If we compare Barry Bonds with Ty Cobb, for example.

Numerous additional examples will be explored in the tutorial as time permits. You can view some of these at the authors site [10].

**Summary**

Hyperglyphs offer a new and powerful approach to exploratory visual analytics by providing ways of visualizing complex, multivariate, unstructured data which varies in both space and time. Hyperglyph design consists of integrating multiple channels of information into a meaningful whole. Visualization of massive amounts of these individual hyperglyphs can lead to insight and questions for further exploration which may not have been considered otherwise.

**References**

1. “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”, https://www.wired.com/2008/06/pb-theory/
2. A. Lie, J. Kehrer, H. Hauser, “*Critical Design and Realization Aspects of Glyph-based 3D Data Visualization*”, p. 24-37, Proceedings of the Spring Conference on Computer Graphics (SCCG 2009), April, 2009
3. R. Borgo, J. Kehrer, D.H.S. Chung, E. MacGuire, R. S. Laramee, H. Hauser, M. Ward, M. Chen, “*Glyph-Based Visualization, Foundations, Design Guidelines, Techniques and Applications*”, Eurographics 2013 Proceedings, M. Sbert, L. Szirmay-Kalos, Eds., 2013
4. Tufte, Edward R. *The Visual Display of Quantitative Information*. Cheshire, Conn: Graphics Press, 2001. Print. [ISBN-10:0961392142]
5. *Visualization: The Second Computer Revolution* (1989) Richard Mark Friedhoff [ISBN-10:0716722313]
6. “Ten Years To The Singularity If We Really Really Try”, 2014, Ben Goertzel, <https://docslide.com.br/documents/ten-years-to-the-singularity.html>
7. *An Afternoon With Mark Pesce”,* [*http://hyperreal.org/~mpesce/interview.html*](http://hyperreal.org/~mpesce/interview.html)
8. *“*Final Amputation: Pathogenic Ontology in Cyberspace“, <http://hyperreal.org/~mpesce/fa.html>
9. ANTz Visualization Application, <http://www.openantz.com>
10. ANTz Visualization Examples, <http://www.edworlds.com/antz/toroids/>