The Applicability of Geovisualization Techniques in
Exploratory Data Analysis Methodology

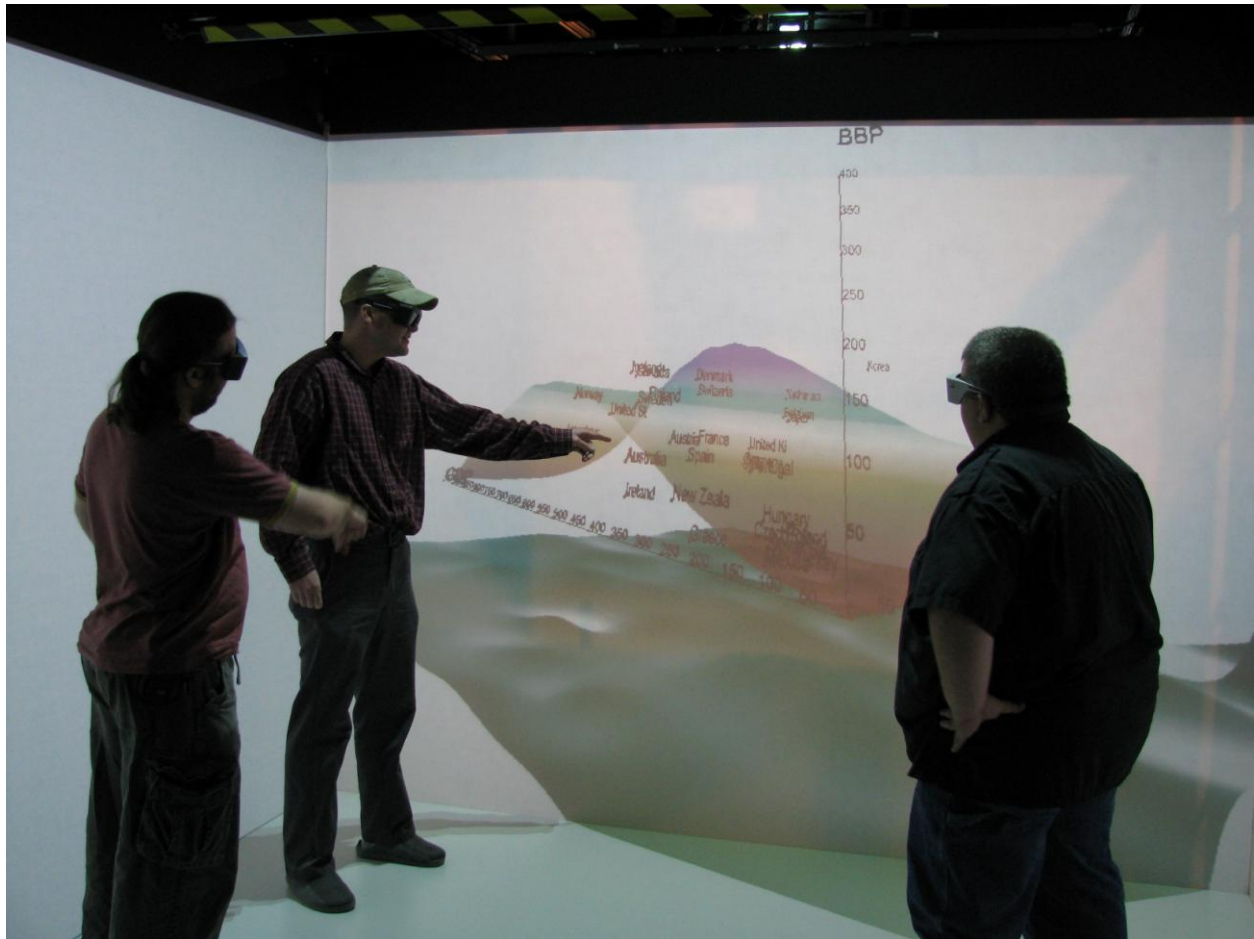by: Isaac Emrick, Mike Harman, Josh Johnson, Frank LaFone.

Contents

The Applicability of Geovisualization techniques in Exploratory Data Analysis methodology

by: Isaac Emrick, Mike Harman, Josh Johnson, Frank LaFone.

Abstract

Three dimensional geovisualization and exploratory data analysis form a natural pair in a virtual environment. The strengths of 3D geovisualization augment the analysis process while adding utility and purpose to the geovisualization software. We integrated 2.5D geovisualization software, immersive technology and EDA principles, which allowed for the development of a prototype 3D EDA system. This system was capable of exploring the strengths and weaknesses of the theory and identifying the major technological hurdles than we need to overcome in order to gain reasonable functionality and control. With debugging and additional product development to improve user control, maneuverability, recognition, and brushing capabilities, a working immersive environment for EDA is possible. It is unclear if this prototype and any future manifestations will yield a marked improvement over the 2.5D workstation environment already available, but it is clear that a highly functional working immersive environment with greater user controls will outperform the working prototype.

Introduction

In the absence of a proper tool, human kind has but two choices, we must make do with the tools at hand or build a better tool. A prototype 3D immersive environment for EDA is just such a tool. This prototype is a patchwork process of software and hardware drawn together like a collage of not quite right tools. We utilized the best available tools to create an environment as close to an ideal virtual EDA environment as we could. Our prototype was successful at many levels, and the development identified many processes and functions that are fundamental and in some cases obligatory to the creation of a highly functioning immersive EDA environment. This report will go about examining the current literature, explaining our objectives, chronicling our attempt to construct a prototype highly functioning immersive EDA environment, and what we witnessed, learned, and scorned along the way.

Literature Review

Above all, Exploratory Data Analysis (EDA) is the process of exploring data in its raw form. It is about attempting to develop linkages between data points without imposing assumptions on those linkages. Cox and Jones write, "In *exploratory data analysis*, attempts are made to identify the major features of a data set of interest and to generate ideas for further investigation…" (Cox and Jones, pg. 135) EDA brings explanatory power to research efforts without obscuring any of the raw information contained in the data itself.

EDA is comprised of two key components – data description and data modeling (Chatfield, pg. 6). Data description is the initial exploration of patterns and linkages within the data. Most of this work is built upon techniques first developed by Tukey in the late 1970's. The key to this process is a non-destructive or transformative impact upon the data itself.

Equally important to the description of the central data is the capturing and understanding the outlier data. Largely, this process is achieved through the application of different data viewing methods, or visualization. More specifically, Anselin suggests "representation of data by means of multiple and simultaneously available 'views', such as a table, a list of labels, a bar chart, pie chart, historogram, stem and leaf plot, box plot, or scatter plot." The descriptive process leads to the formulation of data models that help researchers to develop testable hypotheses. These models are designed to iteratively 'tease out' relationships between the data points, particularly between outlier data points.

EDA brings several strengths to data analysis. Tukey suggests that, assuming a researcher enters the process with no prior knowledge of the data, EDA allows researchers to explore data linkages with no inherent bias. Eherenberg takes minor issue with this, suggesting that prior knowledge cannot be avoided and therefore should not be ignored. However, this prior knowledge should not be allowed to prejudice the findings (Cox and Jones, pg. 135). These biases run the danger of creating a form of tunnel vision blinding the researcher from potential lines of inquiry. Sibley notes that, "Several writers have noted that hypothesis confirmation only discourages researchers from considering new ideas" (Sibley, pg.4). EDA is an incredibly resistant approach, as outliers are taken into the account in the explanation and descriptive process.

Exploratory Spatial Data Analysis (ESDA) is a variation of EDA. ESDA has grown from a recognition that the power of GIS allows for the exploration of spatial data in ways hitherto unobtainable. "It is widely recognized that many of the geographical analysis techniques of the 1960's failed to take advantage of the visualization and data manipulation capabilities embodied in modern GIS" (Anselin, pg. 253). Hence GIS technology allows for the intersection of EDA

5

techniques of data discovery but in a highly interactive graphical format. Researchers can interact directly and instantaneously with spatial data through common GIS functions (Anselin, pg. 253). ESDA is an open field for research and development, both in analytical tools and visualization tools. "Many ESDA techniques have been developed quite recently and this remains an area of very active research" (Anselin, pg. 254).

Interactive Human-Computer Interfaces (HCI) are key to the development of EDA and ESDA models, especially in regard to their use in GIS. As Anselin noted, GIS is a well developed pallet of tools for exploratory data analysis. Nöllenburg further suggests that there are three components to a well developed HCI: effectiveness, efficiency, and satisfaction. He notes that there are two main classes of interaction: focusing and linking/brushing techniques. Focusing essentially links the researcher with the data itself, allowing the researcher to move around and within the data. It is an "interactive modification that selects what to see in a single display and how it is seen" (Nöllenburg, pg. 276). Linking and brushing techniques focus upon query techniques. Linking is "selecting data in multiple views", whereas brushing is "selecting display objects by pointing on them or encircling them on the screen" (Nöllenburg, pg. 279). Clearly the two are closely related. The HCI is a method by which the researcher can effectively explore the data as it is visualized in the EDA/ESDA technique.

Visualization is part and parcel of EDA and ESDA techniques. Authors commonly move directly into visualization techniques as a first step in EDA (Charlton, Chatfield, Sibley, and others). Visualization takes two forms – textual descriptive and graphic representation. Common statistical techniques including histograms, stem and leaf plots, and box plots allow the researcher to quickly identify outliers and see patterns within the data. Further techniques such as scatter plots represent the graphicacy of the data, as suggested by Tukey, Sibley, and

Eherenberg separately. Nöllenburg identifies four key goals of visualization: exploration, analysis, synthesis, and presentation. Viewing "a data set in a number of alternative ways to prompt both hypothesis and their critical reflection" is especially important (Nöllenburg, pg. 261).

The realm of visualization is much broader than techniques described by EDA or ESDA. Certainly the visual is a key element to the human condition. Batty argues it is the most important of senses (Batty, pg. 317). There are almost limitless ways that a researcher can visualize data but the accessible methods that most closely parallel real-world experience have more analytical power. "Batty et al argue that it is this 'connection' between the user and the model that is the distinctive measure of the virtual environment" (Raper, pg. 178). For this reason, a virtual environment is the most compelling approach to EDA/ESDA because it represents a more natural representation of reality. "Virtual environments function directly as geo-representations through their reconstruction of environmental content, geometry and topology" (Raper, pg. 178). Nicholas Hedley suggests that Virtual Reality technology must represent multi-attribute data, support multiple views into the same dataset, such as ego- and exocentric viewpoints, use real tools and natural interaction metaphors, and support for a shared workspace (Hedley et al., pg. 121). Virtual environments are most effective in a true three dimensional model, such as a Cave Automatic Virtual Environment (CAVE). Topological as well as spatial relationships are preserved more accurately in true three dimensional models compared to two-and-a-half dimensional representations.

Although EDA/ESDA is tightly coupled with generic visualization techniques, the integration of EDA/ESDA and three dimensional visualization and interaction has largely been unexplored. As noted, 3D visualization is a more natural representation of the natural universe.

EDA/ESDA seeks to maintain the integrity of the data that the researcher seeks to explore. Therefore it seems natural that a visualization technique that most closely represents data in its most natural form would closely adhere to the goals of EDA/ESDA. As the literature of both fields suggests, each field brings with it a set of capabilities and constraints. The combination of EDA/ESDA with virtual environments may present challenges, but it also presents a potentially powerful new approach for data exploration and analysis.

Objectives

Our objective was to utilize three dimensional geovisualization for the purposes of conducting exploratory data analysis. Specifically, we examined the use of CAVE technology utilizing ESRI's ArcScene product to explore their efficiency and effectiveness in EDA. We designed and constructed a prototype system to allow users to interact with multivariate data in an immersive environment. Further, we identified the strengths and weaknesses of CAVE visualization as a tool for EDA.

Methodology

Nearly any analysis of a dataset examining the impacts of two independent variables on one dependent variable could be used to meet the study's objectives. However, intuitively, it makes sense to pick a set of data that relates to the real world. Although EDA dictates that one should initially analyze the data with as little bias as possible, some understanding of how the variables are related might facilitate the user's understanding in the CAVE environment, thus allowing for a richer exploration of the CAVE's effectiveness for EDA.

We examined the relationship of economic output (Gross Domestic Product), population density (inhabitants/km^2), broadband Internet adoption rates (subscribers/100 inhabitants) for 30 modern, industrialized countries located worldwide, but mostly in Europe. We obtained this data set for the year 2004 from the Organization For Economic Co-operation and Development (OECD) website. The dataset was specifically chosen to be both manageable and intuitive among a wide audience of users, potentially making relationships reasonably evident and evaluation of the technology easier.

We encountered and resolved two issues prior to data entry. First, we preconditioned the data to constrain the variables to similar scales. GDP had a maximum value that was highly disparate compared to the maxima of population density and broadband adoption, forcing us to divide each country's GDP by 1000. Secondly, we noted which parameter we assigned to each axis, facilitating interpretation of the data during visualization process.

The dataset was imported into ArcMap as X,Y data from Excel spreadsheet format. GDP and population density were the X, Y coordinates, respectively. The resultant event data layer was exported as a feature shapefile. The shapefile was then added to a new ArcScene project. Within ArcScene we converted the shapefile into 3 dimensional data, with the 3rd dimension (Z) represented by broadband adoption.

To enhance the visualization of relationships within the dataset, we used the WinSTAT toolbar in Excel for analysis, perform a simple multiple regression, and generate a multiple regression equation. We entered our X and Y variables into the multiple regression equation, generating a series of Z variables. The Z variables and their X and Y counterparts were used to create a TIN, which was a completely flat surface. The TIN was then converted to a shapefile, from which a 3D object was created using the base heights from the TIN.

The data was analyzed farther through a multiple polynomial regression. We converted

the raw data into a series of multiple order values X, X^2, X^3, X^4, Y, Y^2, Y^3, Y^4, and interaction

terms XY, $(XY)^2$, etc. These terms were used as independent variables and a series of multiple

polynomial regressions along with one mlitiple linear regresion on the X, Y and Z data. The

analysis was preformed in the WinSTAT add on toolbar for Microsoft Excel. The multiple

polynomial regression procedure was repeated with the least significant term removed one at a

time. After the series of regressions the analysis with the lowest p value <0.0001 was used to

generate a multiple polynomial regression plane. The analysis yielded a function that defines the

plane (Z = ($3.28045*10^{-9}*X^4$) + ($0.0014126*Y^2$) + ($-1.55858*10^{-6}*Y^3$) + ($1.88329*10^{-8}$); where

X = population density, Y = GDP, and Z = broadband adoption rate). The linear multiple

regression yeilded the folowing equation (Z = (.210334*X) + (0.218093*Y) + (1.384699) with a

p value of 0.0037. The multiple polynomial regression equation was then applied to a grid of

evenly spaced values across the data range in Excel. The X, Y, and generated Z values were

processed using the same procedure as the original data. The generated Z values and the 2D grid

points were splined into a polynomial multiple regression plane. The linear multiple regression

equation was used to plot the point boundaries of a plane covering the range of the data. This

X,Y and Z data was used to create a raster surface, which was converted into a polygon. This

polygon was projected in 3D using the generated Z values from the linear multiple regression

equation.

The ArcScene project, now consisting of 3D data points and a multiple regression plane

and surface, was projected into the CAVE. We added graphical X, Y, and Z axes with an origin

value of zero to provide a frame of reference for data visualization. Further, we adjusted data

point symbology to 3D Simple Marker Symbols (spheres), which caused the points that were

nearer to users within the CAVE to be larger than those that were farther away. Data points were

not discernable by any characteristic other than their relative spatial positions and location along the axes. We labeled the data points with their respective country name. However, the current version of ArcScene 9.2 does not facilitate labeling of individual data points. Therefore, we downloaded a 3D labeling toolbar from ESRI's ArcObjects Online website onto our computer following their directions (http://edndoc.esri.com/arcobjects/8.3/?URL=/ArcObjectsOnline/ Samples/3D%2520Analyst/3D%2520Visualization/3D%2520Text/Label3DToolbar/Label3DToo lbar.htm). The 3D labeling toolbar allowed us to not only label individual data points and features, but also customize the position of those labels relative to their respective features. The label names can be obtained from a any field in the feature's attribute table, e.g., country name. However, there is no provision within ArcScene to edit a feature attribute table. Therefore, we edited the feature attribute table with the Editor toolbar in ArcMap, then added the point feature into our ArcScene project. Many options in the 3D labeling toolbar allow changes to font style, size, and color. We used a 50 point, Times New Roman font in red to create legible labels that contrasted well with the background and data points. The billboarding option in the toolbar allows the labels to be readable regardless of viewing perspective at the workstation. However, once in the CAVE environment, the labels are static and were difficult to read depending on viewing perspective. Moreover, clusters of data points and their associated labels proved difficult to read, unless the data cluster was viewed at a close range. However, close range data viewing resulted in some loss of the perspective of relative positions of all data points. Therefore, zooming in and out of the data point cloud was necessary at times to view the labels and retain a sense of the relative positions of the data points. We also labeled the X, Y, and Z-axis with their respective representations, e.g., "GDP", for information and orientation purposes within the CAVE.

Results & Discussion

Moving EDA/ESDA into the CAVE environment presented a host of technical challenges that needed to be addressed. The fundamental problem arises from the application of non-spatial data in a spatial environment. Spatial data comes with a set of inherent characteristics that partially dictate behavior in spatial software systems, most notably a coordinate reference system. Non-spatial data does not. As a result ESRI products assign a default coordinate system to data that does not come with a coordinate system. This causes the project to display in a counter-intuitive way in the CAVE. However, this deficiency is simultaneously an advantage in that the system is robust in its acceptance of values within the dataset as a coordinate system.

Another technical problem we encountered was created when our data were outside of the bounds, i.e., near and far planes, that are created by OpenGL for display. Outside of these planes, the standard assumes the user is incapable of viewing the data and therefore does not render the scene. The problem is that the lack of a common coordinate system means that ArcScene informs OpenGL to draw the near view plane too far and the far view plane too near. In other words, the project clips out scene features too soon when a user zooms in or out. The correction for this is two-fold. First, the default view angle in ArcScene has to be expanded to exceed 130 degrees. Secondly, ArcScene must be zoomed in fairly close on the project in order to reset the near and far planes to properly view the scene in the CAVE.

By default, ArcScene uses 2D symbology. Viewing this in the CAVE environment makes it difficult for users relative positioning of points. All points need to be converted to true 3D objects. Furthermore, sizes of symbols have to be adjusted. However, symbology that works in an endogenous view might not work as well in an exogenous view. Symbol sizes need to be adjusted to be the best fit for both far and near viewing, which functionally means neither is done

exceptionally well. Colors have a major impact as well. Symbology that can be visualized in an orthographic view on a 2D computer screen does not always work in the stereoscopic views of the CAVE environment. Therefore, final symbology choices should be done in the CAVE prior to viewing.

To assist with the analysis of the data we added the linear plane and a polynomial regression surface to the scene, enhancing our visualization of data clusters and outliers. The method used to generate surfaces dictates their visibility in the CAVE. Most generated surfaces must have some thickness or they will not be viewable from all angles in the CAVE. However, surfaces that have thickness values that are too large may obscure some of the data points. Setting the transparency of the surface to 20–50% allowed all data points to be viewed in relation to and through the surface.

Beyond the current technical limitations of display that we were able to solve, there were a number of technical issues that are currently insurmountable. The ability to effectively brush data points would greatly enhance exploratory data analysis yet there is no method for doing so in the CAVE. Effective interactivity is unavailable in the CAVE environment beyond simple navigation. You can approximate brushing through the use of multiple operators, with one person at the workstation and one person in the cave. However, we found this to be largely ineffective. The addition of labels, while not a true interactive technique, does significantly increase data interpretation by allowing for the identification of individual points. Unfortunately, there is no way to simultaneously compare multiple attributes between points.

To provide a frame of reference, axes must be incorporated into the scene. However, there is no way to automatically derive these axes from the data with our protocol. To overcome the absence of this functionality, we manually created axes from a second set *?* of data points.

The axes we used were not graduated, which would give the user a sense of scale. It is possible to create these graduation marks, but it is a painstaking and time-consuming process.

Although we solved many technical issues that we encountered throughout the process of creating our prototype, some major conceptual hurdles remain. When we first viewed the data cloud, we started from an exogenous point of view. The natural thing for a researcher to do in such an environment is to orient themselves around the cloud of data points. In the CAVE system, the center of the universe is the operator. To move around the cloud in the exogenous view requires that the data cloud to be the center of the universe. The endogenous view in the CAVE places both the user and data at the center, but the user loses the ability to see the complete data set. There are a few benefits to using the endogenous view. This view allows the user to understand the micro-level relationships of data. The immersion in the data creates a unique perspective not easily recreated using other technologies. The endogenous view also allows the user to examine the data from a several unique perspectives: from the origin, from one data point's location, or from all points above the regression plane. These perspectives may highlight relationships, anomalies or outliers which may be otherwise obscured. However, in the endogenous view, the problems with orientation and control were exasperated.

Motion and animation can positively or negatively impact exogenous and endogenous views in the CAVE. The auto-rotate function in ArcScene can be used to rotate the data cloud, which enhances the visualization of all the data points when viewed from afar. Auto-rotation overcomes the center-of-the-universe problem in the exogenous view, but requires the complex interaction between the workstation and the CAVE operators. It is impossible for the CAVE operator to directly control the cloud rotation. This is yet another example of the lack of interactive control that the user experiences within the CAVE environment. From an endogenous

point of view, rotation can help illuminate more obscure linkages. The major problem with auto-rotation is that when viewing the data cloud from within, anything less than a 6-sided CAVE results in the loss of the view of data points as they rotate onto non-existent walls. Moreover, using the wand for motion-control was impractical with auto-rotation enabled because as the scene moved, the reference points within the CAVE also moved. It became impossible to predictably move relative to a moving point.

It is evident that the CAVE has much less interactive functionality than either a workstation or traditional EDA methods. Ignoring the differences in interactivity, it is helpful to compare CAVE (3D), workstation (2.5D) and traditional EDA (2D) visualizations. There are immediate issues with representing three variables with a two-dimensional analysis. The 2.5D view in ArcScecne adds a number benefits over traditional 2D analysis. The user gains multi-dimensionality because the representation of all variables is made possible. 3D visualization allows the user to gain multiple perspectives by moving around and through the data, creating a conducive environment in which to visualize data features such as clusters and outliers. Moreover, the CAVE has a much larger viewing area that allows for collaborative exploration of the data. The CAVE also provides immersion that allows a unique perspective which may or may not be useful depending on the data set and user(s). The single most important difference between the 3D environment of the CAVE and the 2.5D representation at the workstation is that the CAVE is user-centric but the workstation is data-centric. EDA is data-centric; any technology that puts the user above the data will have limited utility.

In EDA analysis, the CAVE has serious technical and conceptual limitations. Some of the technical limitations can be overcome, but they present impediments to the application of this approach. Other current technical limitations render CAVE-based EDA impractical. The

damaging problems for CAVE-based EDA are the conceptual problems for which there are no

solutions. The workstation environment presents itself as a more efficient and effective method

for EDA than traditional 2D methods (e.g., stem and leaf plot) or the CAVE.

Conclusions

The prototype that we developed for the CAVE provided little tangible improvement over

EDA at the workstation level. Does this mean that CAVE technology is inappropriate for EDA?

Our prototype does not definitively answer that question because of inherent limitations, most

notably data and process issues. From a data perspective, the low sample size of our dataset

allowed for intuitive visual interpretation. Therefore, the CAVE lends little additional power. A

larger dataset with more complexity may increase the analytical capacity of the CAVE, over

other visualization methods. Furthermore, the visualization of different datasets with complex

structures, e.g., a double helix, may be enhanced in the CAVE (Appendix A). The development

of the prototype revealed technical limitations that are relatively easy to overcome while

simultaneously rendering the CAVE currently impractical. The utility of the CAVE would be

greatly improved by solving these technical problems. Unless the limitations of interactivity

within the CAVE are addressed, effective EDA is impossible. This is a technically resolvable

problem that may allow the CAVE to become an effective and powerful EDA tool.

Appendix A: Additional Constructs
Additional constructs developed and used in the process of prototype development.

The artificial construction of data is mentally challenging. The process is in many ways like EDA in reverse. The user must visualize what the are trying to develop and think in terms of points and basic grid coordinate systems. This line of thought must then be purseude untill the data construct is complete and functional. Consider the following example, try to construct a portion of DNA in which the relative distances between base pairs of nucleic acids and the number of pairs per revolution in the double helix are contextually correct. The data you must develop should include proper relative spacing between nucleic acid pairs. This will determine the size of the helix. The number of pairs necessary for the helix to complete one revolution will determine the geometric function necessary to identify the locations of each pair. The length of a strand of DNA per each helix revolution, divide by the pairs per revolution will provide the height increase for each new point. After all of these figures are developed they must be converted to a 3D grid system of X,Yand Z values which form a right-handed spiral.

This idea can be converted into 3D virtual reality.   The average width of a DNA base pair is 22-26 angstroms. The average distance between base pairs is 3.4 angstroms. Based on the model by Watson and Crick as told by the unreferenced website Wikipedia, it is estimated that approximately 10.4 base pairs the double helix would complete a revolution. This basic math indicates a 240 unit width, a 34 unit height, and a pair every 34 degrees from the beginning.

The reality of all of this is DNA exists in several states and often the double helix itself is coiled. So now, all we have to do is visualize the orientation of a double helix of these dimensions rotating to the right on a coiled central axis. This new layer of mathematics adds complexity and girth. Each new wrinkle of detail we try to model complicates the data

generation workload. The helical helix proved to be very difficult to create within the time allowed for this class. However, a straight coil of DNA with proper relative spacing was creatable, and proved to be a valuable learning experience in 10th grade geometry, and geovisualization.
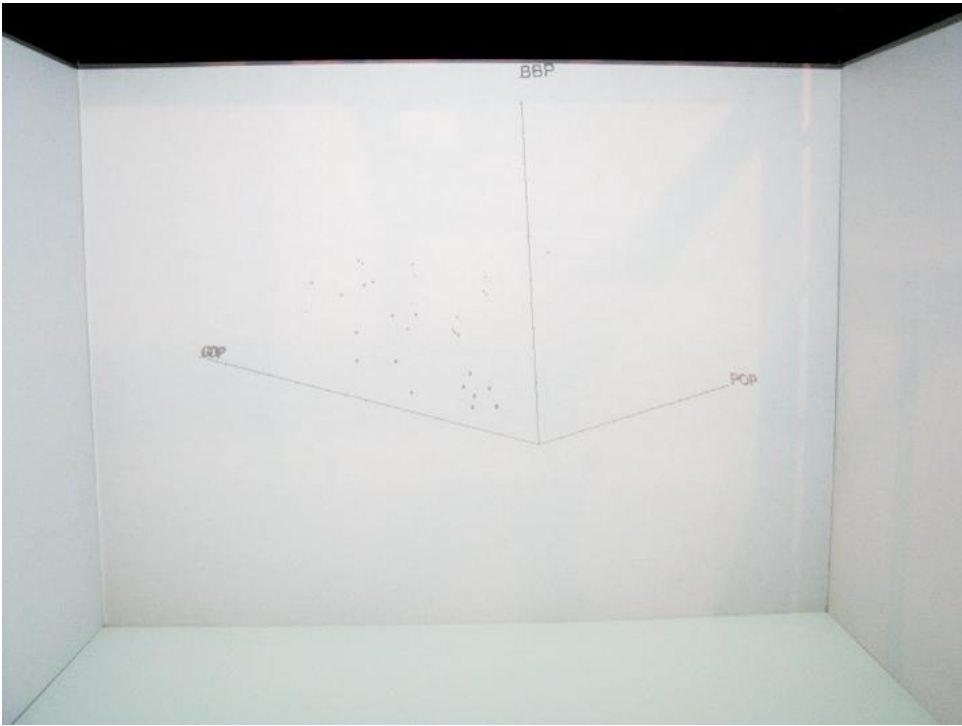
Appendix B: Figures

Figure 1: Basic model showing only axes and points.

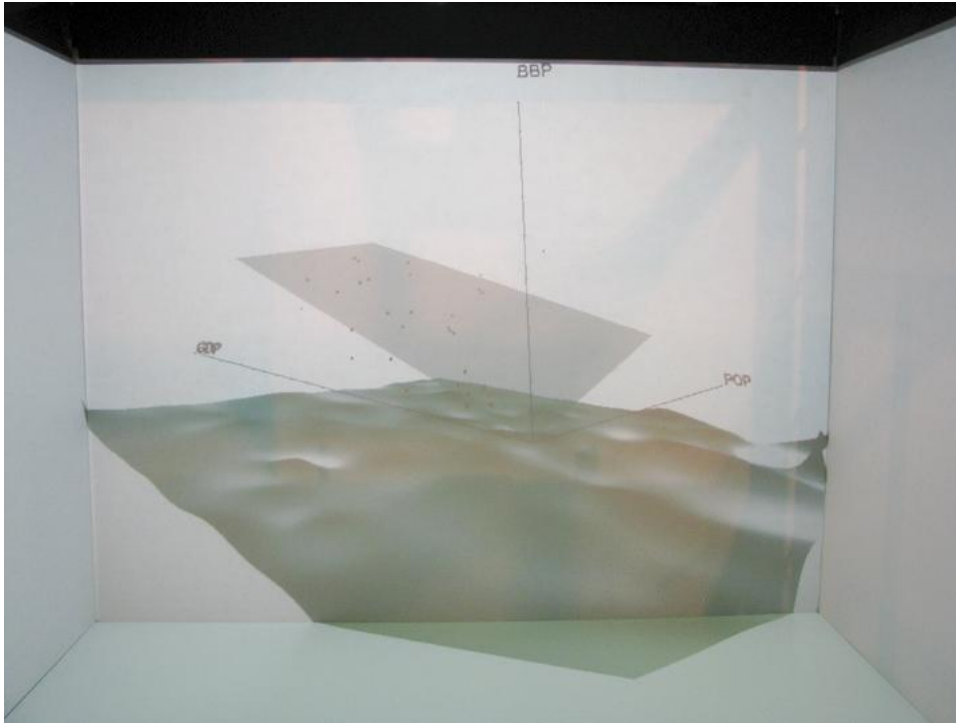Figure 2: Model enhances with the addition of a floor.
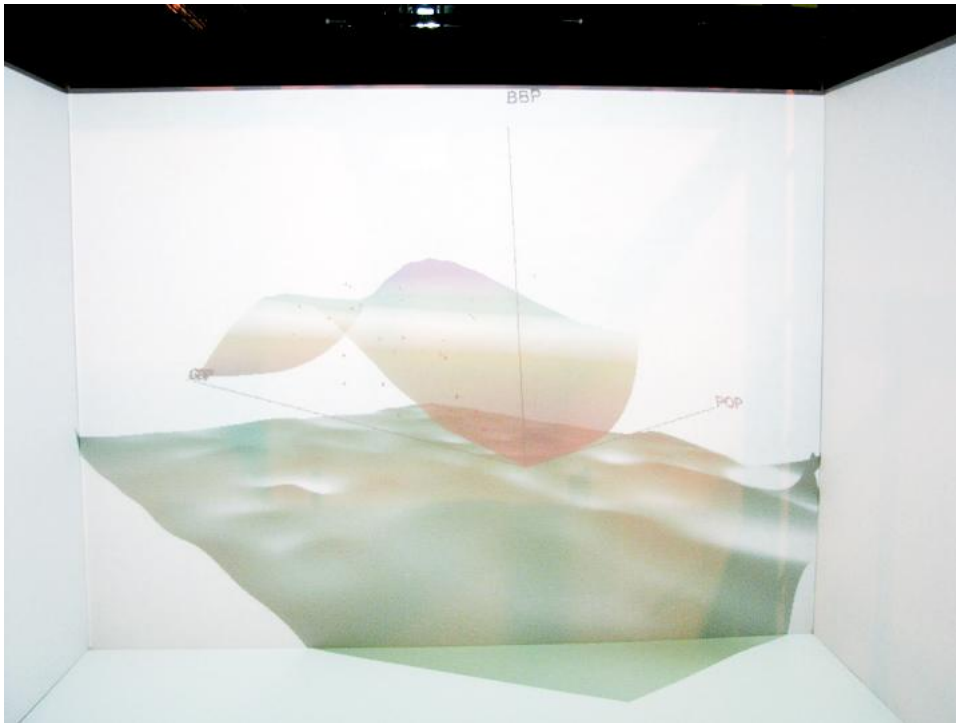
Figure 3: Model with multiple linear regression plane.

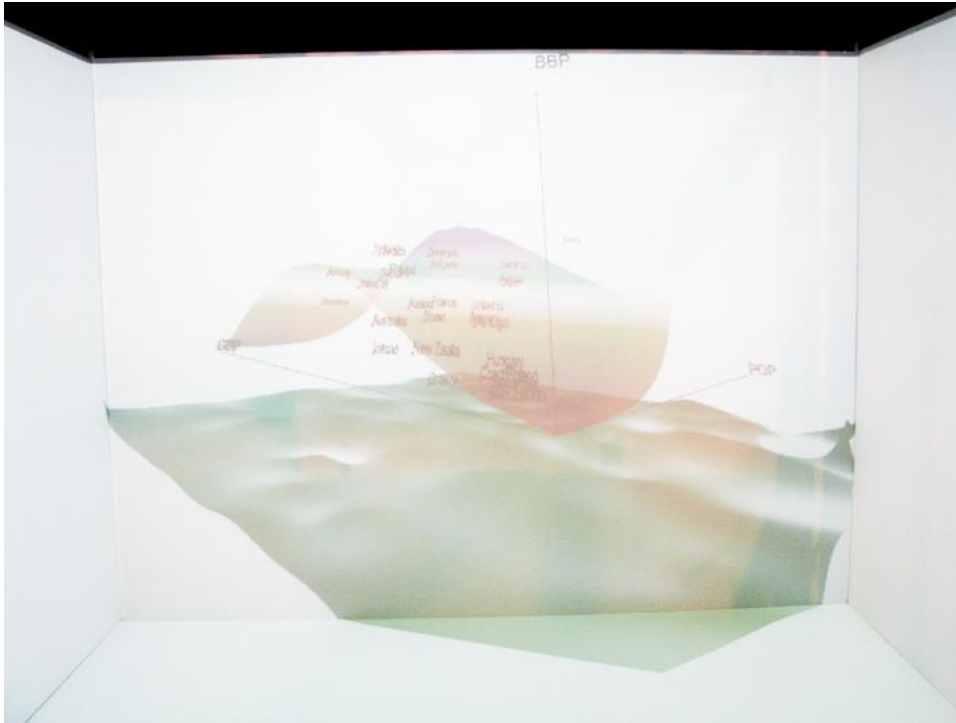Figure 4: Multiple linear regression plane replaced by multiple polynomial regression surface.

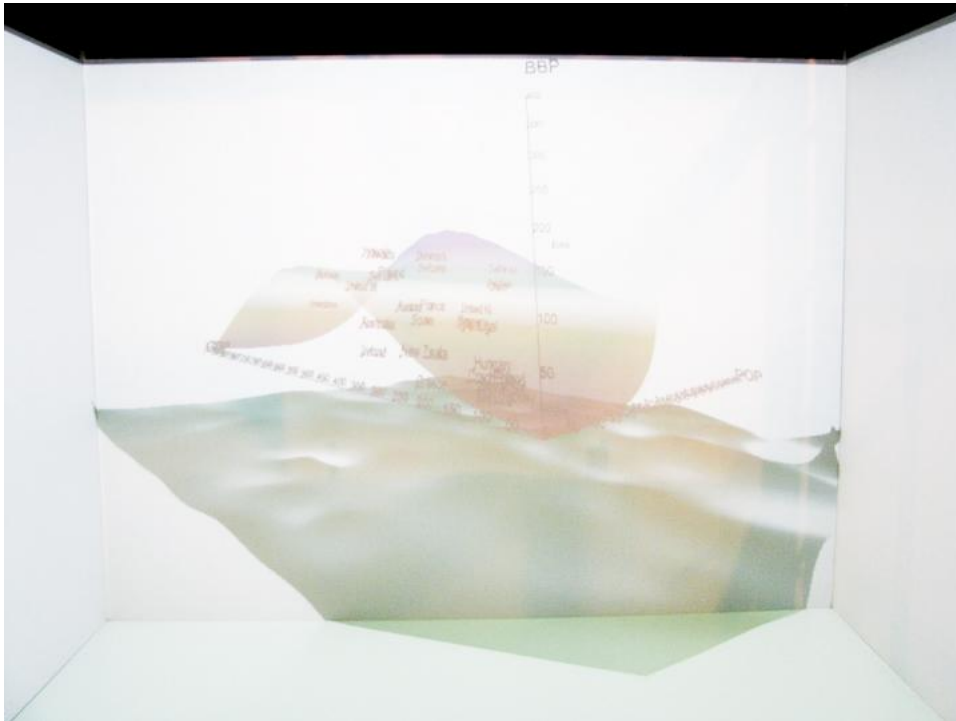Figure 5: EDA model is enhanced with data point labels.

Figure 6: Complete model with data cloud, axes, regression surface, floor, and labels to enhance the analytical power of the visualization.

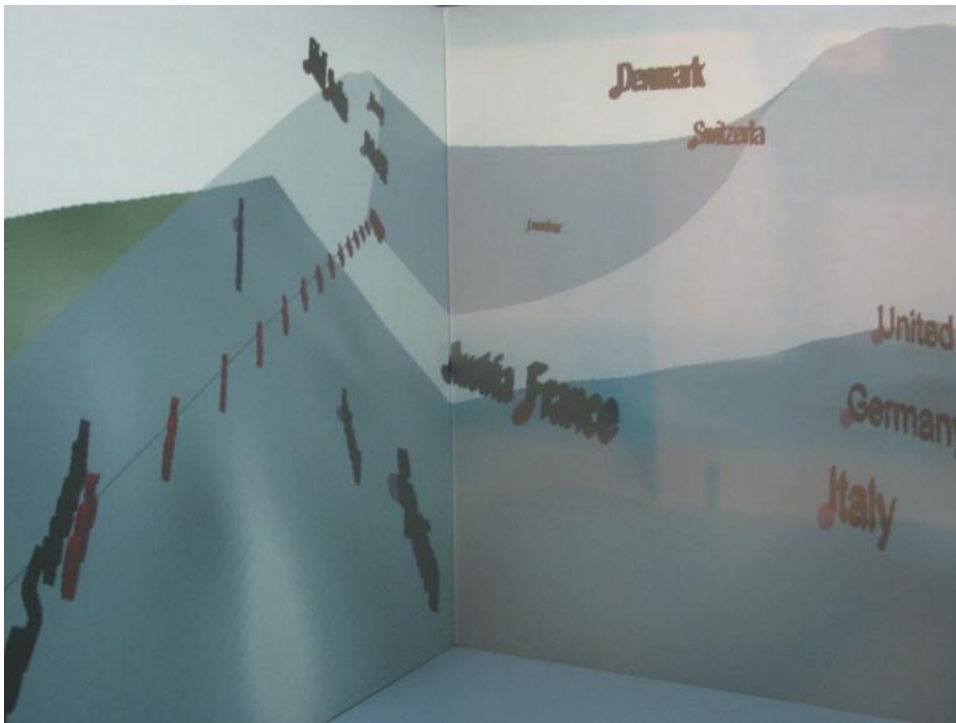Figure 7: Exogenous view of model.

Figure 8: Endogenous view, note the disorientation and difficulty in assessing the relationships between data.
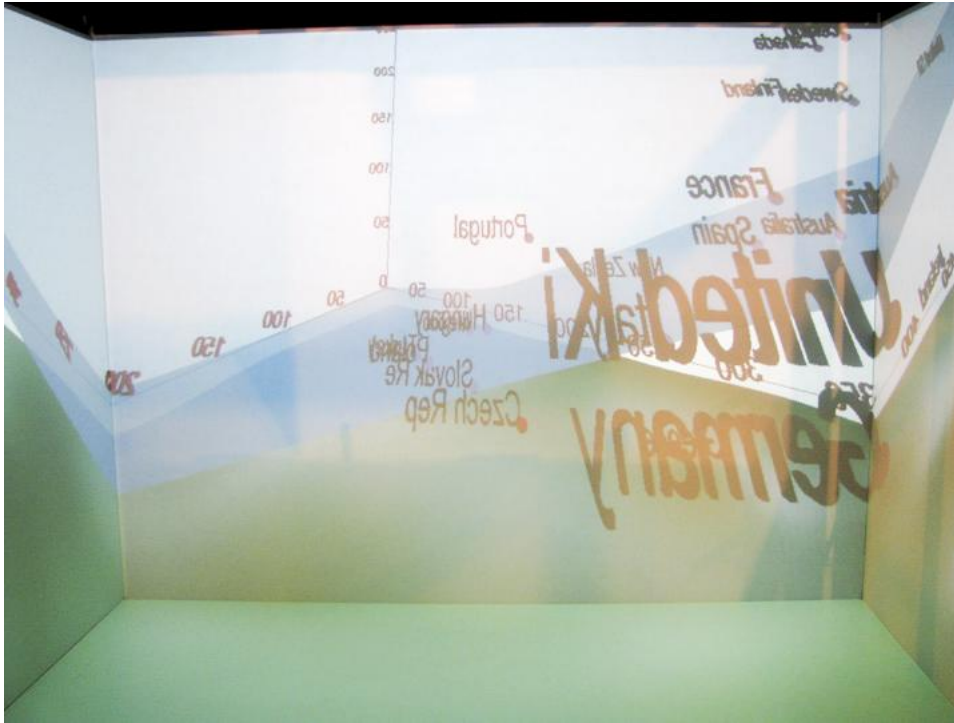
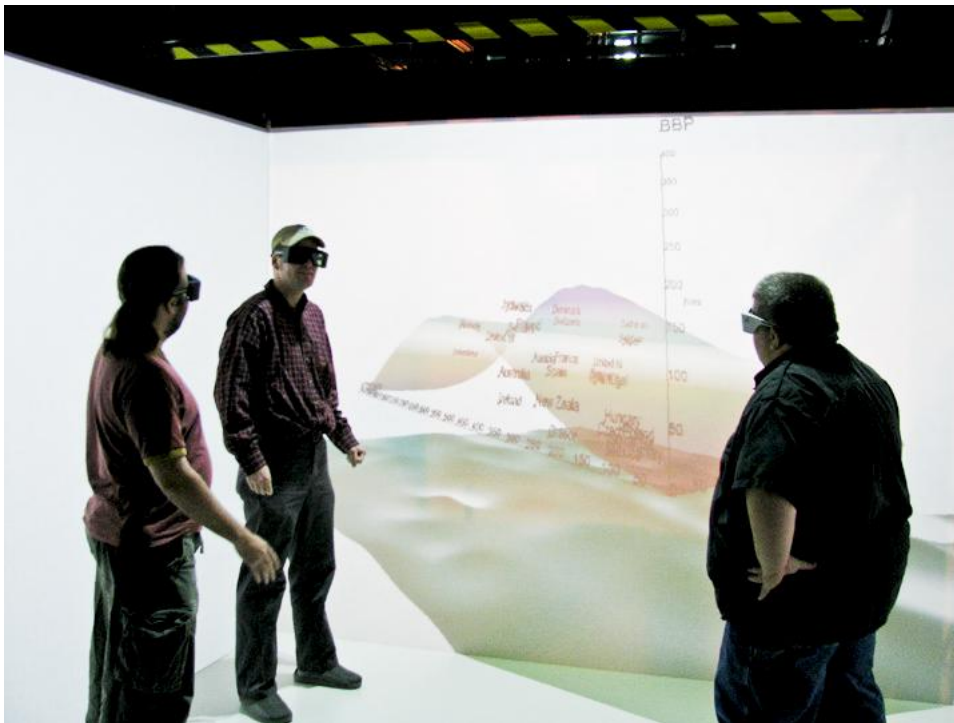Figure 9: Endogenous view towards the origin showing that the labels do not rotate with the user.

Figure 10: Collaborative immersion.

23

Appendix C: Data

Country	Broadband penetration (per 100 inhabitants, 2004) (Z)	Population density (inhab/km2) (X)	GDP per capita (USD PPP) (Y)
Denmark	18.8	126.1	36,208
Netherlands	19.0	400.5	38,389
Norway	14.9	14.4	54,298
Switzerland	17.3	183.0	41,336
Iceland	18.3	3.0	37,986
Sweden	14.5	20.2	37,171
Korea	24.9	484.9	24,750
Finland	15.0	15.6	34,331
Luxembourg	9.8	181.8	81,781
Canada	17.8	3.3	36,610
United Kingdom	10.5	247.3	33,925
Belgium	15.6	345.8	34,603
France	10.6	115.1	32,709
Germany	8.4	230.8	33,470
United States	13.0	32.0	45,489
Australia	7.7	2.8	35,150
Japan	15.0	338.1	33,111
Austria	10.2	98.7	37,437
New Zealand	4.7	15.6	25,927
Spain	8.4	87.3	31,638
Ireland	3.4	62.0	41,490
Italy	8.1	195.4	29,754
Czech Republic	1.6	130.6	21,933
Hungary	3.6	108.3	18,702
Portugal	8.2	114.5	21,918
Greece	0.4	84.3	28,864
Slovak Republic	1.1	110.0	17,875
Poland	2.1	121.9	15,875
Turkey	0.7	94.6	12,619
Mexico	0.8	52.9	13,553

Bibliography

Anselin, L. "Interactive techniques and exploratory spatial data analysis." in Paul A. Longley, Michael F. Goodchild, David J. Maguire, and David W. Rhind, eds. *Geographical Information Systems: Principles, Techniques, Management and Applications, Vol. 1*. 2nd Ed. New York: John Wiley & Sons, 1999: 253-266.

Batty, Michael. "Virtual Reality in Geographic Information Systems." in S. Fotheringham and J. Wilson, eds. *Handbook of Geographic Information Science*, Blackwell, Oxford, UK: 317-334.

Charlton, Martin E. "Quantitative Methods and Geographic Information Systems." in John P. Wilson, and A Stewart Fotheringham, eds. *The Handbook of Geographic Information Science*. Blackwell Publishing, 2008: 379-394.

Chatfield, Chris. "Exploratory data analysis." *European Journal of Operational Research*, 1986, 23(1): 5-13.

Cox, N. J., and K. Jones. "Exploratory Data Analysis." in Neil Wrigley, Robert John Bennett, eds. *Quantitative Geography: A British View*. Routledge, 1981: 135-143.

Fischer, M. M. "Spatial analysis: Retrospect and Prospect." in Paul A. Longley, Michael F. Goodchild, David J. Maguire, and David W. Rhind, eds. *Geographical Information Systems: Principles, Techniques, Management and Applications, Vol. 1*. 2nd Ed. New York: John Wiley & Sons, 1999: 283-292.

Getis, A. "Spatial Statistics." in Paul A. Longley, Michael F. Goodchild, David J. Maguire, and David W. Rhind, eds. *Geographical Information Systems: Principles, Techniques, Management and Applications, Vol. 1*. 2nd Ed. New York: John Wiley & Sons, 1999: 239-251.

Hedley, Nicholas R., Mark Billinghurst, Lori Postner, Richard May, and Hirokazu Kato. "Explorations in the Use of Augmented Reality for Geographic Visualization." Presence, *Vol. 11, No. 2, April 2002,* 119 –133.

Nöllenburg, Martin. "Geographic Visualization." in A. Kerren et al, eds. *Human-Centered Visualization Environments 2006*. LNCS 4417: 257-294.

Openshaw, S. "Developing appropriate spatial analysis methods for GIS." in D. Maguire, M. F. Goodchild, D. Rhind, eds. *GIS Principles and Applications Volume 1*. London: Longman, 1991: 389-402.

Raper, Jonathan. "Multidimensional geo-representations for exploration," *Multidimensional Geographic Information Science*. CRC, 2000: 173-193.

Sibley, David. *Spatial Applications of Exploratory Data Analysis*. Concepts and Techniques in Modern Geography No. 49. Norwich: Geo Books, 1988.