Data Visualizaiton through Graphic Representation

Computer Science Department

Saint John's University

Collegeville, MN

adampbachmeier@gmail.com

Author:
Adam Bachmeier

Supervisor:
Dr. Mike Heroux

May 16, 2014

Abstract

Data Visualization is an interdisciplinary field that uses gathered information to provide meaningful data in graphic form. Jacques Bertin was the first to use these techniques in the field of semiotics. Differentiation of data marks, data presentation, and value representation provide a guideline for data visualization basics. By using these techniques, data sets can create appropriate data visualizations. Data can be represented in 2D, 3D, or 4D. With future advancement, data visualizations will dynamically provide real-time data that can provide dynamic usage.

Contents

List of Figures

1 Introduction

In an attempt to explore the field of Data Visualization, I have taken upon the task of studying data from the Annual Report which is produced by the College of Saint Benedict and Saint John's University. The goal is to create accurate useful maps that better illustrate the data presented in the tables currently presented in the charts. I will be focusing on the students who have come to theses colleges from out of state for my demonstration.

2 Content

2.1 Background

In association with data mining, data visualization has become a field of study that combines statistics, mathematics, computer science, and visual graphics. While still a fairly young area of study, visual informative graphics systems are applicable to disciplines that use data of all kinds. Beginning with applied semiotics, the study of symbols, Jacques Bertin pioneered the modern aspects of data visualization. He established three rules that act as the technical foundation of this area of study. Statistical graphics has become widely used to communicate information and has grown through major paradigms with the invention of personal computers as well as the expansion of data through the Internet.

Data visualization dates as far back as the collection of data itself. In fact, "a primitive coordinate system of intersecting horizontal and vertical lines that enabled a precise placement of data points was used by Nilotic surveyors as early as 1400 BC" [15]. However, this does not imply that modern statistics were immediately implemented. The origin of natural science stemmed from philosophy and was based on empirical evidence rather than theory or logic.

Popular methods such as, "line, bar, and pie charts, originate from the eighteenth century" [6]. Prior to this time period, large amounts of data were scarce due to the time and effort that it took to adequately record accurate information. Only a select group of statisticians and mathematicians would have been interested in using data to provide greater transparency in a particular field of study. Rationality was in opposition to the heuristic approach to science and did not "disdain the atheoretical plotting of data points with the goal of investigating suggestive patterns" [15].

Although, "there were applications of data-based graphics in the natural sciences, it was only after Playfair applied them within the social sciences that their popularity began to accelerate" [15]. While others preceded him in the role of creating different graphing techniques, Playfair's adaptation of graphic methods to subjects such as economics and finance, boosted the popularity of statistical graphics. The reason Playfair focused on graphing data was because he believed that "making an appeal to the eye when proportion and magnitude are concerned, is the best and readiest method of conveying a distinct idea" [14].

In the early 19th Century, French cartographer Jacques Bertin began to collect different maps and diagrams from every discipline and study. When the collection reached his own height, he sorted each piece into two groups: 'reasonable' and 'aberrant.' The sorting was accomplished by distinguishing the efficiency of legibility; 'seeing' provides instant perception, while 'reading' requires greater effort on behalf of the viewer. By focusing on the aspects of each group, Bertin was able to perceive a basic concept of visual variables. "He gradually summarized the properties of the graphic image by the XY plane and the Z attribute: six visual variables (size, value, texture, colour, orientation and shape), three meanings (quantitative, ordinal and differential) and three types of sign (point, line and zone)" [10]

In the mid 19th Century, Semiology, the study of symbols and their associated meanings, became a prevalent field of study. Both Ferdinand de Saussure and Charles Peirce, Swiss and American linguists respectively, formed contemporary studies of Semiology. For many semiotic theorists, language was viewed as the central semiotic system, and terms from linguistics have been borrowed and extended in its application to various disciplines outside linguistics [11]. Bertin saw his 'visual variables' as devices that expanded beyond the scope of modern semiotics and still relied on the power of human perception. "To help designers and statisticians in understanding how to create diagrams, networks, and maps with relevant information that could more easily be understood, Bertin developed a semiotic theory for graphic communication" [11].

He published his first work, titled, *Semiology of Graphics: Diagrams, Networks, Maps*, in 1967. It became "one of the subject's founding texts and represents one of the earliest and most comprehensive attempts to theorize how we perceive and interpret different representations of through shape, pattern, and color" [6]. Based on his studies at the Gestalt School of Psychology, Bertins works established that graphs are read in three stages:

1. **Identify External Identification:** Understanding the components that are involved in the

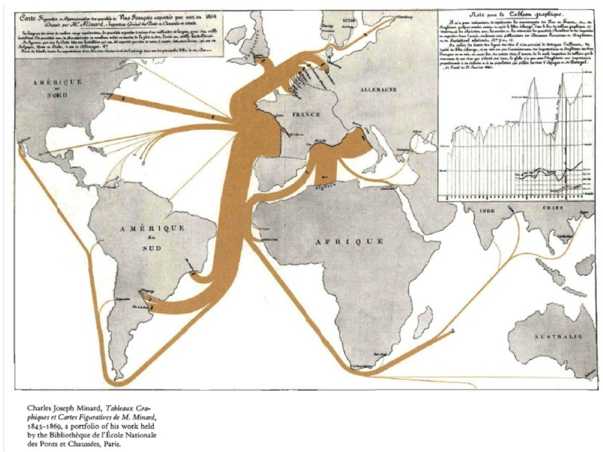

Figure 1: 1864 Shipping Traffic: An example of a chart that Bertin would have collected

graphic, as well as prior knowledge of the data set domain. Can the viewer see and read the data being presented, and can we visually distinguish different categorical and quantitative values legibly?

2. **Identify Internal Identification:** Understanding how the components are mapped to the visual variables they represent, and relating the graphical elements to cues in the legend. Can the viewer judge the relative order or ranking in terms of their magnitude?

3. **Perception of Marks:** The comprehension and meaning of each mark based on its location and visual symbol. This mark could be one of three implantations (point, line, area). Does the graph utilize the most accurate visual variables according to their interpretive precision?

"Although 3-D, animation, and interactive graphics did not yet exist, it [Semiology of Graphics] is well worth studying for the insights it gives about how the design of graphics should be approached and the kinds of questions that need to be considered" [11]. Bertins analysis of graph making includes different levels of perception by the user, new meanings of graphs, the types of components that can be included on graphs, as well as how they can be organized. He recognized the transformation from data to comprehension as taking place in two phases: Data Information to Graphic Image, and then to Visual Comprehension.

Data Analysis in relation to use with computers began in the 1960's with the invention of Graphical User Interfaces (GUIs). Prior to this point, computers had command line interfaces. "In the mid-1980's, advances in computer graphics hardware prompted research on visualization, the use of interactive, visual representations of data to amplify cognition" [9]. Early uses of visualization research focused strictly on sets of scientific data, but later branched to include more abstract information.

At about this time, the first English translation of *Semiology of Graphics* was put into print. "The timing of the English translation couldn't have been more perfect; it was published just around the time personal computers emerged as media, which partially accelerated the growth of information design as a discipline" [5]. Fueled by the prospect of cheap, powerful computing, tools and software for plotting data is now readily available for purchase. "Statistical graphics became widely used to communicate information, to decorate and enliven scientific presentations, and to store information" [15].

Most recently, the Internet has provided a means of greater data accessibility and in response, data visualization has become a cross-disciplinary field. Enormous amounts of data, and non-data, in every study imaginable are readily available for analysis. "Studies that were either too expensive, too tedious, or too difficult can now be done with the click of a mouse" [15]. This also means that there is more opportunity

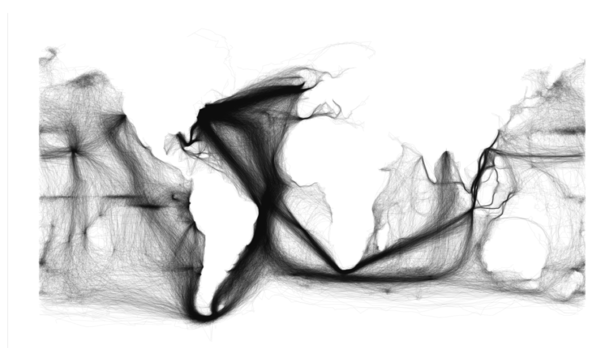

Figure 2: 18th and 19th Shipping Traffic: An example of how much more data can accurately be quantified

for data to be misconstrued or misinterpreted, but improvements in software that is used to create data visualization will hopeful mitigate misinformation.

Though its beginning were rather slow to gain traction, visual graphics have revolutionized the state of data interpretation. From Playfields bar and pie charts, to Bertins visual variable guidelines, graphs now are a constant part of our lives to help us absorb data sets at a single glance. The conversation with graphics will become multidimensional and interactive with the right guidance. With the incorporation of new ways to represent data, Bertins variables will gain new components and add to the plethora of ways in which we can accurately, and compellingly present data.

2.2 Technical Analysis

To demonstrate the application of semiology in relation to computer science, data from the 2012-2013 Annual Report was used to create a geographic mapping of the Fall 2012 graduate distribution by state. A number of important steps were taken to construct this information in a visually satisfying layout. Both data collection and selection processes were undertaken in order to determine which data are useful and usable in the graphic visualizations presented in 4, and 5. Color, size, saturation, and variable selection are all used in order to for the geographic visualization to perform its goal.

2.2.1 Data Collection

The Annual Report, provided by the Office of the Registrar for the College of Saint Benedict and Saint John's University, contains the number of students, separated by class, who have attended the two schools respectively, over the past academic school year. These data are displayed in rows and columns, which holds value for providing exactly how many students graduates from each school came from each region and state. Presenting data this way fails to provide a snapshot of the information in a way that is easily absorbed by the viewer. This is more accurately portrayed using a geographic mapping of the contiguous United States.

It is important to note that data being used is secondary data, meaning that the primary data of each individual has already been collected and redistributed in the Annual Report. This particular secondary data set is two levels of abstraction away from the most accurate information that would still provide useful results. Both city name and county name would provide a smaller geographic region that allow for more accurate findings. However, "since geographic space is continuous, one can always imagine a more precise survey, a more refined enumeration, which means that geographic information always results from human choice" [2]. Although city or county data would provide greater amounts of geographic detail, the amount of data that it would takes to accurately map increases exponentially with each layer of abstraction. Map accuracy is a combination of both the degree of precision of the information, and the degree of precision of the representation.

2.2.2 Data Selection

Each of the 50 states is represented in the data set. While all of this information is useful in the table format, there are a few anomalies that present an issue when presenting the data on a map. Both the states of Alaska and Hawaii, while important, represent geographic issues in both distance and size that complicate the accuracy associated with data and their maps. Washington D.C., due to its classification as a district, rather than a state, also presents an issue with the tool used for mapping.

Both Saint John's University and the College of Saint Benedict attract the majority of their students from inside Minnesota due to the proximity of the location to the student's hometown. In fact, 76% of Saint John's University graduates and 78% of College of Saint Benedict graduates were from Minnesota in 2012. [1]. While this information is interesting, it represents an outlier within the data set. The second highest numbers are both from Wisconsin and each represent 3% of the total graduate class respectively. By removing the Minnesota graduate data, a mapping of the remaining 48 states is able to display patterns that would have otherwise been difficult to perceive.

While the data presented in the Annual Report includes numbers from each class, the focus lies on the graduate class numbers. The graphic included in the Technical Analysis represents the graduate data from a 2012-2013 Annual Report. By focusing on the graduate numbers as opposed to composite students numbers from each state, the graph allows the comparison to data from both past and future reports. For example, composite graphs from 2012-2012 and 2012-2013 would contain 75% of the same information as only the incoming and graduating classes would provide different data.

2.2.3 Future Data Collection

A compilation of graphics from previous Annual Reports will be able to show patterns and trends over the span of the years gathered. A series of these geographic maps could be put together as a video, a gif, or into an interactive application that allows the user to switch between years in order to quickly see the differences between them. If the interactive application is useful to the viewer, it may be of value to include the maps in future Annual Reports provided by the College of Saint Benedict and Saint John's University.

2.2.4 Variable Selection

"Because color is three-dimensional, it is possible to display two or even three dimensions using pseudocoloring" [16]. With the data provided, it is important to differentiate between quantitative and qualitative data. "Color hue refers to different colors such as red and green. Symbols with different hues readily imply differences in quality. Color hue is a good choice for showing qualitative data" [7]. Red is the color chosen to represent data for Saint John's University graduates and blue is the color chosen to represent College of Saint Benedict graduates. These colors are pleasant to look at because they are primary colors. The RGB value for red and blue on the computer screen are (255, 0, 0) and (0, 0, 255) respectively. "Color value refers to different shads of one hue, such as dark or light red. Map symbols with different values readily imply differences in quantity. Value is a good choice for showing quantitative data" [7]. Due to the varying amount of students from different states, it becomes applicable to apply the darker values to the values that report higher amounts of students, and the lighter values to those with fewer amounts of students.

This variable assignment is referred to as a bivariate color sequence. The first variable is a hue, or a continuation of hues, and the second variable is the amount of saturation or lightness applied to the particular color hue. Figure 3 provides an excellent example of a data map that maps more than one variable successfully.

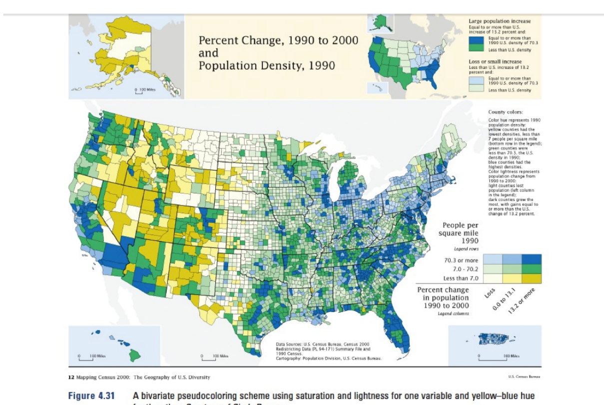

Figure 3: A map showing the percent change in population from 1990 to 2000 using a bivariate color scheme

"There are many considerations that go into making a color sequence that displays desired quantities without significant distortions, thus making it unlikely that any predefined set of colors will exactly suit a particular data set and visualization goal" [16]. In this case, the combination of the Saint John's University data, mapped to red, and the College of Saint Benedict data, mapped to blue, combine to make variations of purple.

2.2.5 Map Construction

"In any problem involving more than two components, a choice must be made between the construction of several maps, each one forming an image, and the superimposition of several components on the same map" [2]. A map containing information solely obtained from Saint John's University data is equally as interesting as a map designed strictly for the College Saint Benedict. Therefore, a composite map acts as a supplemental map in a set with the others, as opposed to a stand-alone map. The reason for this is that, "we do not seem to be able to read different color dimensions in a way that is highly separable" [16]. This issue stems from an inability to differentiate minute fluctuation levels in hue and saturation levels. However, "ignoring saturation can produce some mighty confusing graphics, where individual colors stand out strongly from other symbols for no apparent reason" [3]. This would be the case if the ratio of Saint John's graduates to Saint Benedict graduates was used to determine the color of each state. A state with 1 SJU graduate to 0 CSB graduates represent a 100% SJU graduate rate. This would result in the entire state being associated with a vivid red as opposed to the neutral white that a weighted result would give.

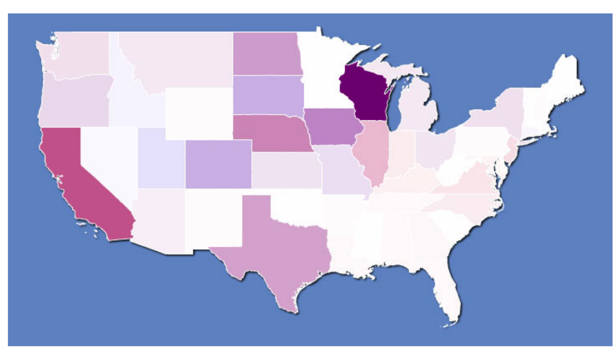

Figure 4: Contiguous map showing CSB/SJU 2012 graduates by Home State excluding Minnesota

Using the TargetMap tool on the Internet, the data could be compiled from a spreadsheet with the data from the Annual Report. A continuous color scheme was the best option in terms of representing the colors blocking by size. Although quartiles are easy associate real numbers with, it is more difficult to be able

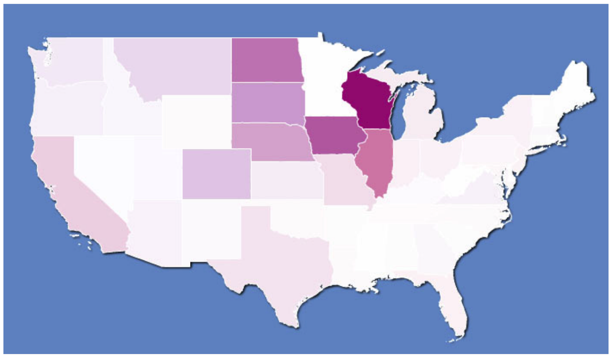

Figure 5: Contiguous map showing CSB/SJU 1980-2012 graduates by Home State excluding Minnesota

to tell the degree to which a color actually represents the data it shares.

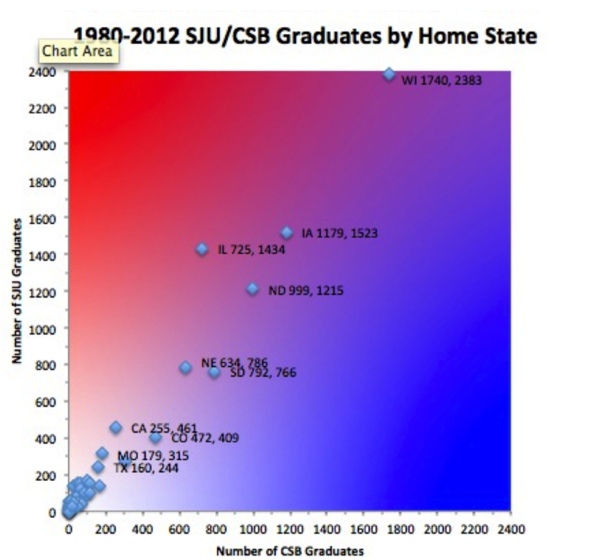

Figure 6: A scatterplot of the variables presented in 5

The following maps include data from 1980 to 2012 showing a 32 year compilaiton of data. In the future, we can use this compilation of data maps and sets to create interactive applications in order to better see patterns through large amounts of data. Although there are many ways to present data, the visualization of data through graphics presents the best way of presenting hundreds of points of visual cues, through colors, shapes, patterns, and sizes in order to create a singular representation of multiple data sets.

2.3 Future Trends

Data visualization, while useful, has much room for improvement in the future. There are four specific areas in which data visualization can grow in future trends.

1. The availability of more types of visualizations
2. The changing of visualizations from being static charts to dynamic interactive representations of information
3. The ability to visualize more data and more complex data
4. The role of industry standards

A combination of some or all of these categories will lead to advancements in the data visualization techniques.

2.3.1 Types of Data

Given how far technology has advanced within the last decade, it would not be unreasonable to assume that advancements in data visualization has made great strides as more tools are developed. Unfortunately, this does not seem to be the case. In the business setting, line graphs, bar graphs, scatter plots, and maps still seem to dominate the majority of visual data representation. While these different data display methods still accurately present data, it is in the interest of all that new types of visual representations are made available for different types of data. Currently, "there are six categories of data visualization techniques: geometric techniques (i.e., scatter plots), hierarchical techniques (trees), pixel-oriented techniques (circle segments), graph-based techniques (directed acyclic graphs), icon-based (shape coding, such as stick figures) techniques, and hybrid techniques (combinations of the previously mentioned techniques). [13] These categories do not inhibit new categories from being created or combining techniques to create new ones. However, it is important to note that the creation of new visualization techniques does not justify the means of their use. Only advancements in comprehension or interpretation are justifiers for replacing current visualization techniques.

Scaling within different data categories is also an area of semiology that has room for further advancement. "Multidimensional scaling combined with graphs promises to be an extremely interesting area of research. Scaling can be applied to graph-based data when using different graph representations and distance measures in order to visualize the impact each approach has. [12]In the case of the data presented in the Technical Analysis, Minnesota represented the obvious outlier in the data set that was collected. In an attempt to better display the differences between other states represented on the geographic mapping, Minnesota results were omitted from the data set before being applied to the visualization. It can be understood by the viewer that the data was omitted, but in a more advanced display, it might be possible to scale into certain portions of the graph in order to see more concentrated results. For example, home zip code data from graduates just around the Twin Cities area would provide beautiful data visualizations that would be hard to interpret if the size of the map remained at a size to include the entire nation.

The areas of data visualization that have seen the greatest increase in interest are the use of video presentation and animation. As early as 1986, Mackinlay, an American information visualization expert made the statement that, “Animation and 3-D presentation appear to be very powerful techniques for presenting symbolic information and should be incorporated into future tools. [8] Today, websites like YouTube provide a platform for individuals or businesses to upload content as videos that can then be shared. In the future, tools that are used to create videos and 3D animation will become cheaper and increasingly user-friendly, hopefully leading to a greater user base of individuals committed to creating useful data visualizations.

2.3.2 Dynamic Interactions

Another area of data visualization that promises improvement within the next three to five years is that of dynamic interaction between the user and the graphic. “Dynamic visualizations allow users to get their hands on the data, by rotating, zooming in or out, and panning visualizations. [15] This also relates to the type of scaling that was referred to in **??**. If the data can be presented in a way that lets the viewer interact with an interface, an executive user can digest the information at a speed that is may be different from a user that need to analyze individual data points. “A natural progression of visualizations for data mining purposes is from visualizations that are static views of data to presentations of information that the user can interact with a number of ways. [15] This is the direction that data visualizations are headed in the future and the technology for this implementation is already available.

“One of the greatest challenges to the future of visualization is the integration of data from many different sources. [4] As with the case of the data that

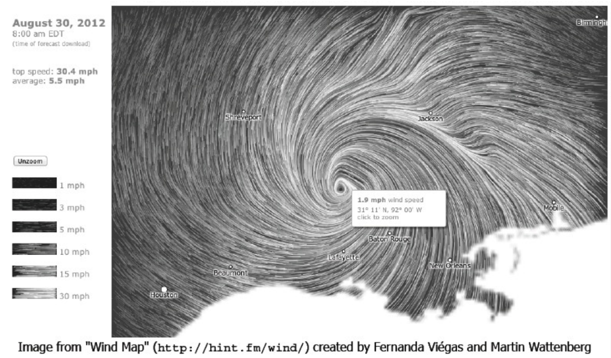

Figure 7: An example of an interactive map where the user’s mouse displays the wind speed of a particular point.

was presented in the Technical Analysis, it was best to use a map that was a combination of maps, one that used data from Saint Johns University graduates and the other that used data from College of Saint Benedict graduates. This map was then presented in tandem with a scatter plot that provided the exact data points and colors that were associated with each state. With the correct tools, these two different data visualizations could be joined using a technique known as brushing. Brushing is the synchronization of two of more visualizations of the same set of data. Mousing over a particular state on the map of the United States would automatically highlight that data point on the scatter plot. Two graphic displays with the same data gives the user a multiple but synchronized view of the same data set. This process can be extrapolated even further using a process known as drill-through, during which, “a visualization allows selection of specific subsets of data points by clicking on them or even drawing a bounding box around them. [15] This specification allows the user to subset these points allowing for dynamic user interaction.

2.3.3 Size and Complexity

In reality, data mining is extremely limited by the amount of data that can accurately be used in data representation techniques. It is of no use to collect huge data sets if only a small fraction of what is being collected is utilized. Hopefully, greater advances in data collection will push data visualization software to a point where it can quickly and accurately compile large amounts of data. Data size directly relates to the speed at which the software is able to preform tasks presented by the user. These issues are currently present with individuals who work in supercomputing or calculating astronomically complex numbers.

Complexity of data is an issue that relates back to the types of data that can be represented in data visualizations. There are already "several different methods of representing web documents using graphs. It is possible to create other, more elaborate representations that include even more information, such as information about html tags or document elements such as sentences, paragraphs, sections, tables, images, lists and so forth. [12] Data from the internet is useful to a people like programmers who make web pages as well as administrators who manage server farms. The compilation of this data in a visually meaningful way would be very useful. "Other types of complex structured data, such as software code, can also be visualized with this method. [12] In addition, regular text documents could visualizations such as word clouds to draw common themes using word frequency counting to quickly determine the contents of a text.

It is important to look at data size and complexity to see trends for the future, but it is equally important to see the future trends in software complexity. "The research on intelligent presentation applies artificial intelligence techniques to part of the user interface design problem-that of choosing an appropriate graphical presentation of relational data. [8] While choosing the correct data representations is a kind of art, some representations obviously work better than other. In the future, it is likely that artificial intelligence will be able to assist in the selection of appropriate graphics based on the data that is being displayed. It may also be the case that the interface of the software will be based on the competence of the individual at sitting in front of the keyboard. "Future work with these techniques can address other aspects of user interface management systems, perhaps choosing or adapting the dialogue specifications appropriate to the observed skill level of the user. [8]

2.3.4 Standards Between Tools

Lastly, there is room for advancement the interconnectedness between different applications that use data sets differently. This has been a problem with users since the implementation of file types. How can data that is stored differently be automatically compiled in a way that matches correlating data? Data points, models, visualizations and even transformations currently support dozens if not hundred of different file types that take large amounts of work to get to seamlessly work together. In the future, we can hope to remove the barriers and create applications that work together. Hopefully the Internet will be able to provide this platform as a working ground for those who attempt this lofty goal.

3 Conclusion

By studying the graphs that were constructed using the data from the Annual Roport, it is accurate to say that even after 32 years, there are still anomolies in the total number of students coming to each school from out of state. Both schools have shown increases in the number of students from states such as California and Texas. However, the greatest surprise was a large amount of Johnnies seem to come from Illinois compared to the smaller amount of Bennies that originate from that area. This may be due to football recruitment around the Chicago area. Further exporation into this could lead to a more detailed explanation. High amounts of students from the states immediately surrounding Minnesota was expected. Further studies could be done with other informtaiton provided in the annual report as well as a continuation on this project with the inclusion of more data as the Registar's Office provides.

References

[1] Office of the registrar annual report. Report, College of Saint Benedict/Saint John's University, 2013.

[2] Jacques Bertin. *Semiology of graphics.* University of Wisconsin Press, Madison, Wis., 1983.

[3] Cynthia A. Brewer. Color use guidelines for data representation, 1999.

[4] Paul A. Fishwick. Aesthetic computing. 2006.

[5] Suguru Ishizaki. Seminology of graphics: Diagrams, networks, maps. *Design Issues*, 28(4):108–109, 2012.

[6] Andy Kirk. Data visualization a successful design process : a structured design approach to equip you with the knowledge of how to successfully accomplish any data visualization challenge efficiently and effectively. 2012.

[7] John Wood Denis Krygier. *Making maps : a visual guide to map design for GIS.* Guilford Press, New York, 2005.

[8] Jock D. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2):110–141, 1986. 22950.

[9] Jock D. Mackinlay. Technical perspective: Finding and telling stories with data. *Commun. ACM*, 52(1):86–86, 2009. 1435438.

[10] Takashi Morita. Reflections on the works of jacques bertin: From sign theory to cartographic discourse. *Cartographic Journal*, 48(2):86–91, 2011. Morita, Takashi 1; Email Address: morita@hosei.ac.jp; Affiliation: 1: Hosei University, Tokyo, Japan; Source Info: May2011, Vol. 48 Issue 2, p86; Subject Term: CARTOGRAPHERS; Subject Term: CARTOGRAPHY; Subject Term: EARTH scientists; Subject Term: SEMIOTICS; NAICS/Industry Codes: 541360 Geophysical Surveying and Mapping Services; People: BERTIN, Jacques; Number of Pages: 6p; Illustrations: 1 Color Photograph, 2 Black and White Photographs, 2 Diagrams, 2 Maps; Document Type: Article.

[11] Glenn J. Johnson Wayne P. Myatt. Making sense of data iii a practical guide to designing interactive data visualizations. 2012.

[12] Adam Schenker. Graph-theoretic techniques for web content mining. 2005.

[13] Tom Soukup and Ian Davidson. *Visual Data Mining : Techniques and Tools for Data Visualization and Mining*. Wiley, New York, 2002. Accession Number: 75798. Publication Type: eBook. Language: English.

[14] Ian Spence. No humble pie: The origins and usage of a statistical chart. *Journal of Educational and Behavioral Statistics*, 30(4):353–368, 2005.

[15] Howard Wainer and Paul F. Velleman. Statistical graphics: Mapping the pathways of science. *Annual Review of Psychology*, 52(1):305, 2001. Wainer, Howard Velleman, Paul F.; Source Info: 2001, Vol. 52 Issue 1, p305; Subject Term: GRAPHIC methods; Subject Term: NOUNS (Grammar); Number of Pages: 31p; Illustrations: 8 Diagrams, 14 Graphs; Document Type: Article.

[16] Colin Ware. Information visualization perception for design. 2004.

4 Reflection

There has been a compilation of coursework during my education that has contributed to the work that I have completed up to this point. Several courses stand out as major pieces that have provided the knowledge needed to accomplish researching the State of the Field for Data Visualization. These courses include: Advanced Placement Statistics (Math 124), Design 2D/4D (Art 118), Design 3D/Drawing (Art 119), Computer Art I (Art 218), and Computer Graphics (CSCI 321).

The Art courses that I have taken have exposed me to the world of using creating and manipulating art digitally. Both Design courses that I have taken have explored the use of line, texture, space, negative space, lighting, and color through both two-dimensional and three-dimensional space. This exposure to the elements of design was useful in the creation of maps as a state of data visualization. The Computer Art class focused on the use of tools using the Photoshop application. I used Photoshop to create and manipulate all of the graphs that were used throughout the semester while working on this project. By learning how to create layers as well as layer masks, it was easy to change the transparency of two maps to create the correct colors for the maps. It was also useful to know how to convert RGB values from a 0-255 scale to a hexadecimal scale. The mapping program that I used received hexadecimal values, but I had to convert those numbers into hexadecimal so the colors were the same in Photoshop.

The background that I had in math, especially Advanced Placement Statistics allowed me to view the data that I obtained with a learned eye. I was wary of trying to make the data say what I wanted it to say. Statistics can be tricky and can be very easily manipulated to make it look like the results suggest something other than what they should. Data mining, or finding relevant information from data sets, was already completed when it was provided for you in a textbook. However, the project has allowed me to experience data mining first hand in an attempt to find new information, specifically about the number of students that attend the College of Saint Benedict and Saint Johns University from out of state.

With the conclusion of Computer Graphics this semester, I feel confident in my ability to expand further on this project. The creation of models in 3D space coupled with my knowledge from the 3D Design course I took holds potential with what the project holds for the future. I believe that with the proper resources, I could be able to create an interactive map or an application that would be able to scroll through

32 consecutive maps for the data provided by either college or the set of the two combined.

This project has been an accumulation of the research skills and writing skills that I have developed throughout my academic career. I have taken a topic that is of interest to me and explored the state of the field that it is a part of. The exploration of graphic representation has introduced me to the field of Data Visualization and its relationship to Computer Science. By calling upon coursework that I had already completed, I was able to bring the whole project together.