INVESTIGATING A KNOWLEDGE DOMAIN: INTERACTIVE VISUALIZATION OF THE GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY BODY OF KNOWLEDGE

A Thesis
Presented to the
Faculty of
San Diego State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Geography
with a Concentration in
Geographic Information Science

by
Marilyn Ruth Stowell
Summer 2014
SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Marilyn Ruth Stowell:

EXAMINING A KNOWLEDGE DOMAIN: INTERACTIVE
VISUALIZATION OF THE GEOGRAPHIC INFORMATION SCIENCE
AND TECHNOLOGY BODY OF KNOWLEDGE

André Skupin, Chair
Department of Geography

Ming-Hsiang Tsou
Department of Geography

Eric Frost
Department of Geology

07/11/2014
Approval Date
DEDICATION

I dedicate this thesis to the memory of my dad, George Stowell and to my mom, Leslie Stowell. This achievement is the product of years of your love and support. Thank you.
ABSTRACT OF THE THESIS

Investigating a Knowledge Domain: Interactive Visualization of the Geographic Information Science and Technology Body of Knowledge 1

by

Marilyn Ruth Stowell
Master of Science in Geography with a Concentration in Geographic Information Science
San Diego State University, 2014

This research compared the effectiveness and performance of interactive visualizations of the GIS&T Body of Knowledge 1. The visualizations were created using Processing, and display the structure and content of the Body of Knowledge using various spatial layout methods: the Indented List, Tree Graph, treemap and Similarity Graph. The first three methods utilize the existing hierarchical structure of the BoK text, while the fourth method (Similarity Graph) serves as a jumping off point for exploring content-based visualizations of the BoK. The following questions have guided the framework of this research: (1) Which of the spatial layouts is most effective for completing tasks related to the GIS&T BoK overall? How do they compare to each other in terms of performance? (2) Is one spatial layout significantly more or less effective than others for completing a particular cognitive task? (3) Is the user able to utilize the BoK as a basemap or reference system and make inferences based on BoK scorecard overlays? (4) Which design aspects of the interface assist in carrying out the survey objectives? Which design aspects of the application detract from fulfilling the objectives?

To answer these questions, human subjects were recruited to participate in a survey, during which they were assigned a random spatial layout and were asked questions about the BoK based on their interaction with the visualization tool. 75 users were tested, 25 for each spatial layout. Statistical analysis revealed that there were no statistically significant differences between means for overall accuracy when comparing the three visualizations. In looking at individual questions, Tree Graph and Indented List yielded statistically significant higher scores for questions regarding the structure of the Body of Knowledge, as compared to the treemap. There was a significant strong positive correlation between the time taken to complete the survey and the final survey score. This correlation was particularly strong with treemap, possibly confirming the steeper learning curve with the more complex layout. Users were asked for feedback on the perceived “ease” of using the interface, and though few users said the interface was easy to use, there was a positive correlation between perceived “ease” and overall score. Qualitative feedback revealed that the external controls on the interface were not inviting to use, and the interface overall was not intuitive. Additional human subjects were recruited from the professional GIS community to participate in testing remotely. These results weren’t significant due to small sample size, but helped to verify the feedback and results from the controlled testing.
Although few of the results obtained in this study were statistically significant, it has revealed opportunities to improve upon this interface in future iterations of the BoKVis, in addition to opening the door to content-based visualization of the BoK.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 INTRODUCTION</td>
</tr>
<tr>
<td>1.1 BoKVis Development/Spatial Layouts</td>
</tr>
<tr>
<td>1.2 Research Objectives</td>
</tr>
<tr>
<td>2 LITERATURE REVIEW</td>
</tr>
<tr>
<td>2.1 Knowledge Domains</td>
</tr>
<tr>
<td>2.2 Spatialization</td>
</tr>
<tr>
<td>2.3 Human Computer Interaction</td>
</tr>
<tr>
<td>2.4 Hierarchical Layouts</td>
</tr>
<tr>
<td>3 METHODOLOGY</td>
</tr>
<tr>
<td>3.1 BokVis Development</td>
</tr>
<tr>
<td>3.1.1 Similarity Graph</td>
</tr>
<tr>
<td>3.1.2 Interface Enhancements</td>
</tr>
<tr>
<td>3.2 Human Subjects Testing</td>
</tr>
<tr>
<td>3.2.1 User Survey Creation</td>
</tr>
<tr>
<td>3.2.2 Focus Group Testing</td>
</tr>
<tr>
<td>3.2.3 User Testing</td>
</tr>
<tr>
<td>3.3 Analysis of Results</td>
</tr>
<tr>
<td>4 DISCUSSION AND CONCLUSIONS</td>
</tr>
<tr>
<td>4.1 Final Results</td>
</tr>
<tr>
<td>4.2 Limitations</td>
</tr>
<tr>
<td>4.3 Conclusions</td>
</tr>
<tr>
<td>REFERENCES</td>
</tr>
<tr>
<td>Section</td>
</tr>
<tr>
<td>-------------------------------</td>
</tr>
<tr>
<td>A  FLASH SURVEY</td>
</tr>
<tr>
<td>B  SURVEY DOCUMENTS</td>
</tr>
<tr>
<td>C  QUALITATIVE FEEDBACK</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Technology Used- Software and Programming</td>
<td>20</td>
</tr>
<tr>
<td>Table 2</td>
<td>Survey Accuracy and Completion Time by Method</td>
<td>36</td>
</tr>
<tr>
<td>Table 3</td>
<td>Independent Sample T-Tests to test if the differences between the means are statistically significant between methods</td>
<td>36</td>
</tr>
<tr>
<td>Table 4</td>
<td>Correlation between Total Survey Time and Average % Correct on Multiple Choice portion</td>
<td>36</td>
</tr>
<tr>
<td>Table 5</td>
<td>Multiple Choice Questions and ID</td>
<td>38</td>
</tr>
<tr>
<td>Table 6</td>
<td>Survey Accuracy- Percent of subjects with the correct answer, by question</td>
<td>38</td>
</tr>
<tr>
<td>Table 7</td>
<td>Which is the parent element of the other: Database Design or Conceptual Model?</td>
<td>39</td>
</tr>
<tr>
<td>Table 8</td>
<td>In which knowledge Area can one find the most information on accuracy?</td>
<td>39</td>
</tr>
<tr>
<td>Table 9</td>
<td>How many Units does the Knowledge Area &quot;Organizational and Institutional Aspects&quot; contain?</td>
<td>39</td>
</tr>
<tr>
<td>Table 10</td>
<td>In which of the following Knowledge Areas can you find out about the role of economics?</td>
<td>39</td>
</tr>
<tr>
<td>Table 11</td>
<td>Paired Sample Correlations between Multiple Choice Question Time Stamps</td>
<td>40</td>
</tr>
<tr>
<td>Table 12</td>
<td>&quot;In which Knowledge Area(s) is the candidate likely to be most qualified for this position?&quot;</td>
<td>43</td>
</tr>
<tr>
<td>Table 13</td>
<td>&quot;In which Knowledge Area(s) is the candidate likely to be least qualified for this position?&quot;</td>
<td>43</td>
</tr>
<tr>
<td>Table 14</td>
<td>Independent Sample T-Test: Scorecard Question</td>
<td>44</td>
</tr>
<tr>
<td>Table 15</td>
<td>Human Subject Status</td>
<td>49</td>
</tr>
<tr>
<td>Table 16</td>
<td>Human Subject's Prior Knowledge of the BoK</td>
<td>49</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1: Potential applications for mapping knowledge artifacts onto the GIS&T BoK. (Ahearn et al., 2013) ..............................................................2

Figure 2: Spatial Layouts utilized for visualizing the GIS&T Body of Knowledge ..........3

Figure 3: Treemap Layout- Zoomed into Unit level within the Analytical Methods Knowledge Area (Analytical Methods → Geometric Measures) ..........................4

Figure 4: Treemap Layout- Zoomed into the Learning Objectives level (Analytical Methods → Geometric Measures → Shape) ......................................5

Figure 5: TreeGraph Layout zoomed into Learning Objectives level (Analytical Methods → Geometric Measures → Shape) .......................................6

Figure 6: Indented List Layout zoomed into the Learning Objectives level (Analytical Methods → Geometric Measures → Shape) ....................................7

Figure 7: Similarity Graph Layout (Analytical Methods → Geometric Measures) ........8

Figure 8: Methodology Flow Chart ................................................................19

Figure 9: Similarity Graph Methods Flow Chart .............................................21

Figure 10: Full Extent of projected 2D coordinates after MDS, visualized in ArcMap 22

Figure 11: Zoomed in projected coordinates .....................................................23

Figure 12: Example of XML Structure for Help Menu Elements .....................25

Figure 13: User Survey Stages Flow Chart .....................................................27

Figure 14: Boxplot distribution of Total Score by Method (outliers labeled with survey completion time) .................................................................33

Figure 15: Boxplot distribution of total survey completion by method (outliers labeled with total survey score) ......................................................34

Figure 16: Average Minutes per Question- Multiple Choice ............................40

Figure 17: Visual results of copying/pasting the job ad (query 1) and resume (query 2) onto the BoK ............................................................................42

Figure 18: Visual results of two scorecards (course 1 and course 2) mapped onto the BoK ..................................................................................44

Figure 19: Average Minutes per Question- Entire Survey ...............................46

Figure 20: Subject perceptions- “How easy was this program to use”? ............48

Figure 21: Participant Agreement/Qualification ............................................61

Figure 22: Spatial Layout associated with survey ...........................................62
Figure 23: Experience/Expertise in Geography .................................................. 63
Figure 24: Experience/Expertise in GIS ............................................................. 64
Figure 25: Computer Interaction .................................................................. 65
Figure 26: Age Group ..................................................................................... 66
Figure 27: Educational/Professional Status .................................................. 67
Figure 28: Gender ............................................................................................ 68
Figure 29: Color-blind/Vision-Impaired ....................................................... 69
Figure 30: Prior GIS&T Knowledge ............................................................... 70
Figure 31: GIS&T BoK Introductory Information .......................................... 71
Figure 32: Survey Procedure ........................................................................ 72
Figure 33: Additional Survey Instructions ..................................................... 73
Figure 34: Multiple Choice 1 ......................................................................... 74
Figure 35: Multiple Choice 2 .......................................................................... 75
Figure 36: Multiple Choice 3 .......................................................................... 76
Figure 37: Multiple Choice 4 .......................................................................... 77
Figure 38: Multiple Choice 5 .......................................................................... 78
Figure 39: Introduction to real-world scenario questions ............................ 79
Figure 40: Real World Scenario 1 - Job Ad/GIS Professional ...................... 80
Figure 41: Multiple Selection Interface 1 ...................................................... 81
Figure 42: Multiple Selection /Open Ended Answer Interface for real-world scenario ...... 82
Figure 43: User Feedback Interface ............................................................... 83
Figure 44: Survey Submission ...................................................................... 84
ACKNOWLEDGEMENTS

I want to thank my family for their constant support throughout my entire college experience. Mom and Natalie, thanks for asking about my thesis progress during every phone call, and thanks to my Dad for knowing me well enough to suggest that I do a Masters in GIS before I had ever even taken a GIS course. I couldn’t have done it without you all pushing me and encouraging me. Thank you.

I would like to thank my friends and colleagues that I have had the pleasure of working with throughout my SDSU experience. Extra special thank you to Fangming who saved my life multiple times when I had hit a wall with programming! Thanks to those who stood by me while I griped about my thesis instead of just doing it…Jenny, Maddie, Daniel, Claire, Gabe, Tony and again to Mom and Natalie…

So many faculty members have been instrumental in my successes at SDSU, but I would like to thank Harry for assisting with all of my technical difficulties! Also special thanks to Dr. Weeks for being such an incredible teacher and positive motivator.

A very important person I would like to thank is my advisor, Dr. Skupin. I had the pleasure of taking and TAing every one of Dr. Skupin’s courses offered (sometimes more than once…) at SDSU throughout my time here, and am continually amazed at how much I still am able to learn from him. My sister told me once, “If you find a professor that you learn well from, take every one of their courses that they offer because it will never be a waste.” And that’s exactly what I did, and I haven’t regretted a moment. So, thank you, Dr. Skupin, for encouraging me to pursue my Masters when I was undecided, and forcing me to learn programming (despite my resistance!). You have been a great mentor to me over the past 6 years and I truly appreciate your commitment to your students’ successes. Thank you for turning me into one of those success stories!

Finally, thank you to my thesis committee (Ming, Dr. Frost, and Dr. Skupin) for their guidance and leadership. It has been such a pleasure working with you all over the past few years on such a diverse range of projects outside of thesis work. I look forward to collaborating in future endeavors.
CHAPTER 1

INTRODUCTION

Published in 2006, the Geographic Information Science and Technology Body of Knowledge (GIS&T BoK) has become one of the most influential documents in the area of GIS&T education, both in the United States and internationally. It was created by the University Consortium for Geographic Information Science (UCGIS) as a key component of their Model Curriculum Project, aiming to better prepare students for GIS related professions. The BoK has been primarily distributed as a printed document, created with contributions from professionals and scholars from over 80 institutions. It consists of a four-level hierarchy of 10 knowledge areas, 73 units, 329 topics, and over 1600 learning objectives, designed to support the development and assessment of GIS&T curricula (Prager & Plewe, 2009).

The current hierarchical structure is merely one way of presenting the configuration of elements within the GIS&T BoK. Visualization allows for exploration of meaningful spatial layouts, potentially revealing gaps, connections and currently hidden relationships between elements. Thus, the research presented here aims to create and compare four different spatial layouts of the GIS&T BoK. Main objectives are to improve accessibility and understanding of the structure and content of the GIS&T BoK through visualizations that lay out the top-down hierarchy as well as exploit the content of BoK elements in a bottom-up manner. Our efforts at developing this BoKVis application is situated within a larger project that is developing foundations for the next iteration of the BoK (termed BoK2), that in addition to this visualization component (termed BoKVis) will consist of ontological, participatory, and immersive components (Ahearn et al., 2013; DeMers et al., 2013).

In knowledge space, one is able to engage in activities and produce knowledge artifacts, yet to this point there hasn’t been a means in which to situate these activities within a knowledge domain. However, by applying geographic concepts to abstract space, the knowledge domain becomes a reference system or base map onto which other knowledge artifacts can be overlaid, whether these be syllabi, CVs or whole curricula, as shown in
Figure 1 below (Ahearn et al., 2013). An outside application allows for input of weights or scores assigned to specific BoK elements, as shown in Figure 1; the resulting input is termed a BoK Scorecard, and could be mapped directly onto a BoK basemap.

![Figure 1: Potential applications for mapping knowledge artifacts onto the GIS&T BoK. (Ahearn et al., 2013)](image_url)

In this thesis, the knowledge domain of GIS&T will serve as the reference system; the users will be able to explore this domain via the interactive visualization interface. Research in human-computer interaction acknowledges that flexible and dynamic tools are especially useful for exploring unfamiliar data (A. M. MacEachren & Kraak, 1997). Since the BoK is likely unfamiliar to many individuals, this research investigates how users are able to utilize this interactive tool when asked to complete particular cognitive tasks. Additionally, we will develop and compare different spatial layouts of the BoK elements, three of which are driven by the hierarchical structure of the BoK, and one based on the textual content.
1.1 BoKVis Development/Spatial Layouts

The existing hierarchy-based visualizations of the BoK are of the type treemap, tree graph and indented list (Figure 2). These three spatial layouts were created by transforming the printed document into an XML file, which preserves the hierarchical structure, then implementing different spatial layout algorithms of that hierarchy. The fourth spatial layout, the similarity graph, breaks up the existing hierarchical structure and the topology is constructed based on the content of the BoK text.

Figure 2: Spatial Layouts utilized for visualizing the GIS&T Body of Knowledge

Johnson and Shneiderman (1993) introduced the treemap as one method for visualizing hierarchical data. It is supposed to allow the user to better understand the structure and content of large data repositories. The user is able to click on a portion of the display which will then reveal its constituent components, or leaf nodes, as shown in Figure 3 and Figure 4. The main goal of that approach is to make large hierarchical data sets more easily understood.
by the user. To this end, they focus on efficient space utilization, interactivity, comprehension, and esthetics. The treemap effectively uses display space and reduced graphical complexity (Wood & Dykes, 2008), which makes it useful for large data sets such as the BoK. Although computational procedures for displaying hierarchical data continue to develop, few studies test the usability of the methods (Stasko, Catrambone, Guzdial, & McDonald, 2000). Stasko et al. (2000) compared the treemap to a radial display to test the treemap’s performance and critiqued the treemap for its complexity and failure to convey the entire hierarchical structure. This research compares the treemap to other spatial layouts and further tests this method based on the user’s ability to complete particular tasks.

Figure 3: Treemap Layout- Zoomed into Unit level within the Analytical Methods Knowledge Area (Analytical Methods → Geometric Measures)
Figure 4: Treemap Layout- Zoomed into the Learning Objectives level (Analytical Methods → Geometric Measures → Shape)

The tree graph also utilizes the BoK hierarchy to determine its spatial layout, in the form of connecting lines from a parent element to its children, as a node-link relationship (Figure 5). This is useful for examining the structure of the BoK in that the user can easily reveal the lower levels of the BoK while still being able to view the upper levels, which in this case contain the parent elements.
Figure 5: TreeGraph Layout zoomed into Learning Objectives level (Analytical Methods → Geometric Measures → Shape)

The indented list utilizes a hanging indent to distinguish the various levels from one another, when the children elements are revealed (Figure 6). This layout is useful for the BoK because it is structured logically and is intuitive to interact with based on the commonality of this type of layout for accessing hierarchically structured data, such as file management systems.
Unlike previous methods, the similarity graph derives each element’s geometry, or position, by their similarity to other elements. This similarity measure is determined by the textual content of each element in relation to all other elements. In this spatial layout the hierarchical structure of the BoK is preserved, as each Parent element is connected to their children via links (Figure 7).
Figure 7: Similarity Graph Layout (Analytical Methods → Geometric Measures)

The visualizations were designed to allow the user to explore the structure and content of the Body of Knowledge, with a range of interactive features:

- Expansion of nodes upon click to reveal descendent elements
- Toggle Interactivity and Animation on/off
- Radio button activation of any/all level of BoK
- Text query with dynamic symbolization of results
- Element Scaling
- Dropdown-menu “scorecard” overlays
- Radio Buttons to change spatial layout
- Hover activated help menu
- Sorting of Elements based on scores
1.2 Research Objectives

The primary objective of this research is to improve accessibility and understanding of BoK through interactive visualization, and gain insight into the user’s experience utilizing the various different layouts when asked to complete tasks related to the BoK. To do this, four interactive visualizations of the BoK have been developed, utilizing different spatial layouts. The treemap, tree graph and indented list methods lay out the hierarchical tree structure inherent within the BoK as it currently exists. The fourth method breaks up this existing structure and instead utilizes a spatial layout based on the similarity of each individual element. A survey designed for user testing will allow for direct comparison and analysis of their performance and effectiveness. Since the similarity graph is still under development, the usability study does not include this method for comparison to the other three completed methods. The following questions guide the approach of this study:

1. Which of the spatial layouts is most effective for completing tasks related to the GIS&T BoK overall? How do they compare to each other in terms of performance?
   This question is intended to compare the spatial layouts to each other, by testing the user’s ability to complete general cognitive tasks. Results will be determined by the number of correct answers on the survey, and will be refined to assess performance through analysis of the time it took a user to come to the correct answer.

2. Is one spatial layout significantly more or less effective than others for completing a particular cognitive task?
   This question is intended to compare the spatial layouts to each other and assess them individually in terms of performance based on specific tasks. Results will be determined through analysis of the time it took a user to come to the correct answer, on a question-by-question basis.

3. Is the user able to utilize the BoK as a basemap or reference system and make inferences based on BoK scorecard overlays?
   This question is intended to determine the overall impact of the visualizations for uses beyond exploration of the structure or content of the BoK. The survey will test for the user’s ability to make inferences about mappings onto the BoK, and will be analyzed in terms of correctness and time to complete particular tasks.
Some of these questions are open ended and will encourage discourse regarding future design aspects based on qualitative analysis.

4. Which design aspects of the interface assist in carrying out the survey objectives? Which design aspects of the application detract from fulfilling the objectives?

These questions are intended to test the usability of the application. Through qualitative analysis and feedback from the user about their reactions and experience with the interface, the strengths and weaknesses of the application will be examined.

For this research the hypotheses are as follows.

- Null Hypothesis (Ho): There are no statistically significant differences among the visualizations.
  - A. There are no statistically significant differences between the three visualizations in terms of survey accuracy.
  - B. No statistically significant differences between the three visualizations in terms of recorded time to correctly answer questions (time stamp). This hypothesis will be tested by comparing accuracy and response times indicated in the database of survey results.

- Alternate hypothesis 1 (HA1): The treemap will prove less efficient (longer response times) for completion of cognitive tasks due to the more complex spatial layout.
CHAPTER 2

LITERATURE REVIEW

This research is significant for the field of GIS&T specifically, yet is incorporated into the wider context of utilizing different spatial layouts, or spatializations, as a means for interpreting knowledge spaces. Mapping various elements onto spatializations has been successful in the past (A. Skupin & Fabrikant, 2003; A. Skupin & Hagelman, 2003; A Skupin, 2009), yet the literature only shows preliminary research in mapping knowledge artifacts onto a Body of Knowledge spatialization.

This section reviews the literature associated with conceptualizing knowledge domains and examines efforts in visualizing knowledge spaces. Additionally, the literature on spatializations confirms the usefulness of visualization for conceptualizing multi-dimensional data. The literature pertaining to human-computer interaction and cartographic design informs the BoK visualization design aspects and computational considerations.

2.1 Knowledge Domains

Knowledge domain refers to an abstract information space containing knowledge actors, activities and artifacts all sharing a common theme or epistemology (Ahearn et al., 2013). It is a broad term that spans beyond one’s “technical field” or discipline and embodies many diverse ways to derive knowledge, such as communities, genomes and networks (Boyack, 2004). This abstract information space conceptually would take up an area, useful for expressing relationships among artifacts. These knowledge spaces can intersect each other, as we see with many interdisciplinary studies. The idea that a knowledge domain occupies an abstract space and is constructed of knowledge elements and activities incorporates the notion of ontology. Chandrasekaran et al. (1999) define ontologies as “content theories about the sorts of objects, properties of objects, and relations between objects that are possible in a specified domain of knowledge” (Chandrasekaran, Josephson, & Benjamins, 1999, p. 20). They further explain that an accurate ontology and associated
vocabulary is necessary in order to have an effective knowledge representation, which can be conceptualized and shared. Gärdenfors defines a conceptual space as a “collection of one or more quality dimensions” in which objects are situated in terms of their properties (Gärdenfors, 1995, p. 5). For this thesis, the domain ontology will be that of the domain of GIS&T (as presented in the GIS&T BoK document), and the elements contained in this body of knowledge will be positioned within this conceptual space; efficiently and effectively representing the elements of the GIS&T BoK is the foundation of this visualization effort.

Current studies in Information Science revolve around finding the best method in which to organize large databases of literature and citation networks, in order to better navigate and display these knowledge spaces (Shiffman & Borner, 2004). Due to the number of online, textual data repositories encompassing these domains, there is a need for an effective way in which to understand and the complex composition of information space (A. Skupin & de Jongh, 2005). Thus far, information scientists have focused primarily on utilizing available technologies to create a big-picture of the relationships and components contained within all scientific knowledge (Borner, Chen, & Boyack, 2002). This research visualizes the relationships between elements in the GIS&T knowledge domain, thus working towards mapping knowledge at a smaller scale.

The content of knowledge domains typically is structured using one of two approaches, the network and the multivariate approach, which would be too difficult for a human to understand if just presented in tabular form. Because of this, it is critical that we reduce the number of dimensions and allow for the user to view the information from a variety of perspectives (Shiffman & Borner, 2004).

In visualizing these knowledge domains and mapping science, it is especially important to note the interconnectedness between elements and represent them spatially in order to discover patterns. These patterns in turn will let users understand the relationships inherent within the knowledge domain (Borner et al., 2002). Because knowledge domains are continuously developing and merging, the visualizations must be capable of efficiently monitoring the changes. As such, specific computational techniques must continue to develop to address the complexities inherent in mapping a field of knowledge (Chen, 2004). This research further develops these computational techniques, through programmatically flexible coding, which allows for a smooth
transition between BoK1 and BoK2, and future interdisciplinary visualization efforts.

There have already been efforts towards visualizing the GIS&T BoK. Prager and Plewe (2009) utilized visualization in their assessment of the GIScience Curriculum, which used a radial layout to display the BoK. With this visualization method, a GIS course could be mapped onto the display, in turn showing how the course is situated within the domain of GIS&T (Prager & Plewe, 2009). Painho and Curvelo (2008) created an interactive tool called the GIS&T BoK e-Tool Prototype which was designed to allow for management and exploration of the BoK. This utilized a network spatial layout, connecting the leaf nodes to their parent with a line. Both of these visualization efforts aim to find an effective means in which to interact with the BoK for various applications. This research aims to further interact and understand the content of the BoK, and to implement content-based spatial layout algorithms to reveal relationships that may not be apparent within the hierarchal layouts. Additionally, the user testing here examines the degree to which these various layouts are effective for examining a knowledge domain.

2.2 Spatialization

Spatialization is the combination of the use of spatial metaphors to make sense of an abstract concept, and the subsequent transformations of data necessary to aid in comprehension (A. Skupin & Fabrikant, 2008). This is becoming increasingly significant with the growing number of digital datasets, and limited means by which to make sense of large amounts of data. The ideal spatialized visualization should represent data in the simplest form for the user to comprehend, while including the information necessary to adequately see and understand patterns in the data (A. Skupin & Buttenfield, 1997).

A metaphor is “the mapping of one domain of experience into another, more coherent, powerful, or familiar one” (Couclelis, 1998). This type of abstraction will ideally enhance the knowledge the user gains from viewing a visualization of a dataset. For example, the user will inherently understand spatial relationships, such as spatial autocorrelation, as described in Tobler’s First Law (Tobler, 1970). Intuitive geographic concepts, such as distance, would help to explain how similar or dissimilar elements in visualization are to one another if they are meaningfully placed to resemble their actual relationship to each other. “Distance, or closeness, is a geometric notion and so we enter
the realm of geometry... When words are represented by vectors, the closeness of vectors can represent the similarity of their meanings (Widdows, 2004, p. xi).” Additionally, the terminology commonly used in geography such as map, space, place, scale and region (A. Skupin & de Jongh, 2005) could all be used to organize large datasets into more meaningful layouts.

Using spatial metaphors, along with traditional geographic techniques for modeling and displaying geographic phenomena, one can apply meaningful spatial locations to non-geographic data by noting relationships between elements (A. Skupin & Fabrikant, 2008). However, the fundamental principles behind the computational techniques and cognitive implications of spatializations are still being developed (A. Skupin & Fabrikant, 2008).

So far, much of the research in spatialization has been studying various techniques for visualizing high-dimensional data, such as bibliographic data, or web networks. However, techniques have progressed to begin to examine geographic data, such as census data. Skupin and Hagelman (2003) took this a step further and applied spatialization methods to address change in census data over time by visualizing trajectories through multi-dimensional attribute space (A. Skupin & Hagelman, 2003). Various studies utilizing the self-organizing map have also been extremely useful for demonstrating relationships in complex datasets (Kohonen, 1990; A. Skupin & Fabrikant, 2003), and have been more recently applied to knowledge space to expand on this functionality, and allow for mappings of various artifacts onto a SOM.

There has also been successful research involving spatialization of non-geographic, textual information. Skupin and Buttenfield (1997) applied spatialization techniques to information spaces by taking two issues of the New York Times and breaking down the constituent (about 100) articles into lists of key terms. They then used mathematical algorithms and similarity matrices to create coordinates for the key terms location in the geometric space. Once they had derived the coordinates they could then see and analyze the relationships between the articles. The authors assert that this concept works, is replicable, and could potentially be used to explore large collections of information or data repositories. It also questions how other spatial metaphors such as slope, could be used effectively (A. Skupin & Buttenfield, 1997). Skupin and Fabrikant (2003) explain the usefulness of using a cartographic approach for all spatialization tasks, geographic and
otherwise. They maintain that geographic expertise is a key component in making decisions regarding dimensionality reduction techniques and spatial layout. Many of the conceptual and computational principles mentioned here are fundamental in visualizing the GIS&T Body of Knowledge.

2.3 Human Computer Interaction

MacEachren et al (1992, p. 101) identify that data visualization is primarily an “act of cognition” which draws on the human ability to “identify patterns and to create and impose order.” An additional study confirms the power of graphic representations and suggests the need to better understand the human process of acquiring and understanding spatial information (Buttenfield & Mackaness, 1991). Spatialization is particularly effective for information retention because it is designed for the user’s interaction with the interface to be as intuitive as possible. Because of this, an important element in designing knowledge domain visualizations is designing the way in which the user will interact with the interface. Most spatializations follow up complex computational procedures with interactive visualizations for best results (A. Skupin & Fabrikant, 2003). The notion of interactivity is somewhat debatable, seeing as how reading a static map could be considered “interacting” with it (Crampton, 2002; A Skupin, 2004), however for this thesis, interactivity will refer to the human responding to the visualization through a computerized interface, or “direct manipulation” (Slocum, 2001).

Research in interactivity has primarily focused on combining different methods of interaction to maximize the amount information a user can comprehend. Crampton (2002) identified the four types of interactivity as: the Data, the Representation, the Temporal Dimension and Contextualizing Interaction, and each type was ranked typologically to determine effectiveness (Crampton, 2002). Russell et al (2003) explain that interactivity also serves to lessen the cognitive demands of the user, and increases the amount of knowledge that can be obtained and retained from visualization (Russell, Sharpe, Brodaric, Boisvert, & Logan, 2003). Additionally, Friedhoff and Benzon (1989) argue that quality visualization tools can help to eliminate much of the extraneous information processing tasks that users typically have to apply (Friedhoff & Benzon, 1989). This moves the method of acquiring information into preconscious visual processing, which is a product of lower order
information processing that comes involuntarily to the user. Because of this, visualization tools can provide most of the inferences and perceptions for users intuitively (Larkin & Simon, 1987). This thesis will assess the effectiveness of the four spatial layouts, based on the user’s ability to complete particular cognitive tasks, and a time-stamp will help to refine how long it took the user to process and understand this information. Much of the research on interactivity has been related to animations (Slocum, 2001), and there remains a gap in research utilizing interactivity to study abstract information and knowledge domains in particular.

One can’t assume that there is a “most effective” display method without first looking at the individuals and their different abilities to comprehend particular displays. The user will essentially dictate how effective a display is, due to their particular cognitive capabilities. Benyon (1993) explains that all humans exist within a specific social context, which is a result of social and political factors. Examples of individual differences would be sex, education, personality, cognitive skills, motivation, goals, and mood (Benyon, 1993). Because of this, the survey will ask the user for particular information designed to better understand the user’s ability to acquire information. For example, the survey will ask for the participant’s age, gender and level of GIS experience. This will not completely alleviate the issue of individual differences, which often complicates usability testing, but will help to hone in on some of the nuances behind users’ survey results.

The complexity of visualizations should not be underestimated, and human subjects often need longer amounts of time to understand the patterns. The layout should, in fact, come second to the content (Tufte, 1997). Garrett (2003) created a five-tiered standard in which to format dynamic layouts, primarily for web-based visualizations. This standard incorporates strategy, structure, skeleton, and surface, assisting the designer in specific design techniques for visualizing certain levels of abstraction, with the goal of creating the most intuitive and effective map (Garrett, 2003). Maceachren and Kraak (1997) argue that flexible and dynamic visualization tools are preferred for exploring data that may be unfamiliar to the user. They maintain that the display should be oriented towards the user’s ultimate goals and the user should be able to explore the visualization and then be able to answer certain questions on their own. They also identify room for improvement in future research in incorporating multiple perspectives to facilitate understanding of geographic or
abstract data (MacEachren & Kraak, 1997). This research applies these design considerations in order to maximize the users’ ability to acquire knowledge, while preserving data integrity.

2.4 Hierarchical Layouts

Network analysis is an effective framework in which to process and visualize data that can be represented with a node-link type structure. Networks, or “graphs”, have an inherent structure with quantifiable measures and relationships between elements, which can be effectively applied to many different types of networks, but has proven particularly useful for analyzing textual content (Borner, Sanyal & Vespignani, 2007). The visualization techniques behind network visualization remain challenging, particularly finding an optimal graphic layout (André Skupin, 2014). The complex issues of the layout are exemplified when coupled with the usability factor (Shneiderman & Aris, 2006). Many tools have been created to handle this issue through dynamic layouts that can be changed on the fly to support the users’ end goals. Some of these tools include Gephi, Network Workbench, Sci2 and Pajek. Although these tools are computationally flexible and advanced (Börner, 2011), they lack control over the visualization’s appearance.

Hierarchical visualization also encounters many of these same issues in finding the balance between optimal layout and user involvement, thus becoming the foundation for many studies. There has also been research done directly comparing spatial layouts in terms of performance and usability. Barlow and Neville (2001) compared four 2-D visualizations of hierarchies: the organization chart, icicle plot, treemap and tree ring in terms of performance and found the icicle plot and tree ring to be superior to the treemap, and comparable to the organization chart (Barlow & Padraic, 2001). Additional research surrounding hierarchical data visualization expands beyond the single hierarchy and begins to explore the interconnectedness between multiple related hierarchical structures (Robertson, Cameron, Czerwinski, & Robbins, 2002).

Hierarchical visualization techniques are evolving from static tools and visuals towards dynamic, mutable interfaces. Collaborative hierarchical visualizations have emerged as a robust way in which to have access to shared ideas, skills and knowledge, focusing on flexibility and interactivity with multiple users. (Isenberg & Carpenter, 2007) With the
growing number and size of data repositories, there is an increasing desire to be able to monitor layout and attribute changes when data evolve over time (Tu & Shen, 2007). Comparative research aims to better develop hierarchical data visualization techniques, while focusing on the user’s ability to interact with the tools (Holten & Van Wijk, 2008).

This research addresses many of these challenges. The user testing performed will help to ensure that the user is able to utilize BoKVis to complete tasks related to the BoK hierarchy, while maintaining tight control of display elements and working towards collaborative development of future iterations.
CHAPTER 3

METHODOLOGY

The key purpose of this study is to examine the performance and effectiveness of interactive spatial layouts for obtaining information pertaining to a Body of Knowledge. The research was conducted in 5 stages: BoKVis Development, User Survey Creation, Focus Group Testing, User Testing, and Results Analysis. The first stage is the process of developing the visualization version to be used for testing. The second stage is the creation of the user survey. The Focus group testing, stage 3 refines the BoKVis application and survey, prior to administering the test. Then, the user testing (stage 4), tests users’ ability to obtain information about the Body of Knowledge based on their interaction with the visualizations, using an interactive survey. The final stage is the statistical analysis of the survey results (both qualitative and quantitative) and the production of the final conclusions of this research. Table 1 below lists the software and tools used throughout all 5 stages.

![Methodology Flow Chart](image)

Figure 8: Methodology Flow Chart
### Table 1: Technology Used - Software and Programming

<table>
<thead>
<tr>
<th>Developers</th>
<th>Software/Programming</th>
<th>Usage for this Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adobe</td>
<td>Flash CS5/Actionscript 3.0</td>
<td>Used to create the User Survey Interface - deployed through .HTML.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linked to MySQL database via network to store survey data output.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Create interactive buttons to move forward/back in the survey</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Record and store survey answers</td>
</tr>
<tr>
<td>ESRI</td>
<td>ArcGIS 9.3.1 - ArcMap</td>
<td>Preliminary Visualization of Similarity Graph Coordinates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Storage of survey results (both answers and timing results)</td>
</tr>
<tr>
<td>Microsoft</td>
<td>Excel 2013</td>
<td>Simple statistical analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Output of basic results through charts/graphs</td>
</tr>
<tr>
<td>IBM</td>
<td>PASW Statistics 18.0</td>
<td>Statistical/Quantitative analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Output of basic results through charts/graphs</td>
</tr>
<tr>
<td>Ben Fry &amp; Casey Reas</td>
<td>Processing (<a href="http://www.Processing.org">www.Processing.org</a>)</td>
<td>Programming Environment for Developing BoKVis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Additional Libraries Used : Ani (Animation Library), ControlP5 (GUI Elements), RiTa/Wordnet/OpenNLP (Natural Language Processing), Treemap (Treemap Layout)</td>
</tr>
</tbody>
</table>

#### 3.1 BOKVIS DEVELOPMENT

The creation of BoKVis has been an ongoing process, with contributions from SDSU graduate students since the start of its development. The main programmatic undertakings of this research were the creation of the similarity graph spatial layout, the design and implementation of user-interface enhancements, as well as ongoing debugging.

#### 3.1.1 SIMILARITY GRAPH

All of the programming for BoKVis was done using Processing, a free, open source programming language and environment that allows the user to easily see visual results of their coding. It is useful for all projects with visual outputs, and is capable of creating images, animations and interactive displays. There are many benefits of using this environment for the BoK visualization including being easy to learn, Java-based language, exportable applications or web applets, multi-server compatibility and accessibility to libraries (Fry 2008). This project also utilized a library called RiTa, which is a Natural Language Processing (NLP) library that can be leveraged in the Processing environment. This library contains built in text analysis functionality for quantifying text, or Lexicons.
In order to create the fourth visualization of the BoK1 based on its text, the similarity matrix of the content was computed within Processing, using each individual element as a separate “document” based on a unique ID. Each document’s value was determined with a Euclidean similarity measurement which calculates the distance between two document vectors (Borner, Chen and Boyack 2007). To do this, the code iterated through the entire tree and compare each lexicon to all other lexicons to compute a dissimilarity value (ranging from 0 to 1, with 0 being a perfect match and 1 being completely dissimilar). These methods created a 2071-dimensional matrix, with one dimension for every term vector. After calculation, this matrix was output to a text file to be accessible by the application while not negatively impact performance speed.

Because of the high dimensionality of this space, it needed to be reduced to two-dimensions to be visualized in the BoKVis in an optimum spatial layout that preserves the topology of the elements in relation to one another. The similarity matrix was brought into SPSS to calculate these new coordinates using Multi-dimensional scaling. The output of this process was exported to a text file containing the unique identifier of each element as well as X and Y coordinate.

In order to get an idea for how these coordinates will appear in a two-dimensional layout, an Excel spreadsheet was created from the text file in the previous step to be imported and projected in ArcMap. The overall shape of the points is circular, with the root node (the GIS&T BoK) towards the right side of the circle (Figure 10). It would be expected that the children of the root would be laid out surrounding the parent with just slight adjustments accounting for dissimilarity. Shown in (Figure 11) below, the root node is the largest green node, and the smaller, lighter green nodes represent its immediate children (or Knowledge Areas). The yellow nodes represent the next depth (Knowledge Units), and this zoomed
image shows that for the most part, the children are located near to the parent.

Figure 10: Full Extent of projected 2D coordinates after MDS, visualized in ArcMap
Figure 11: Zoomed in projected coordinates

Next, the coordinates were brought back into Processing as an attribute of each element. In order to get the coordinates to fit within the measurements of the screen, the map function reprojected the coordinates within the new bounds.

The existing code from the Tree Graph and Indented List was used as a starting point to create the Similarity Graph. This helped to ensure that the new visualization was following the same structural and programmatic formatting for easier adjustments with future development. Additionally, this allowed for the use of the tree topology as the input for the edges in the structure, connecting child elements to their parent. In the main method, a SimGraph class was added to adopt the same functionality as the other methods, which was already programmed in the code. This included interactive expansion of nodes, query, scaling, depth control (showing all elements at a certain level at once) and scorecard overlays.
The Similarity Graph spatial layout serves as a prototype for future work. In its current state it does not adequately allow the user to see the structure in its entirety. Using a method such as this should effectively group elements that are more similar closer to one another, thus distinguishing different clusters or groups, and allow for quantification of similarities and differences as well as cluster analysis of the BoK1. In its current state, it is difficult to do this at any depth of the visualization with any accuracy. The zoom and pan function are useful so that the user can have more control over the visualization and can see more of the structure and the relationships between elements. This could be remedied with the use of a different dimensionality reduction technique such as a SOM, which could optimize the layout for this purpose. This will require further research to determine the best spatial layout for display purposes. Additionally, with a dataset this size, an on-the-fly layout would be possible, rather than reading in coordinates from an outside source.

Future development will explore additional means of laying out the BoKVis based on its content, using various dimensionality reduction techniques such as SOM and Latent Dirichlet Allocation, which will facilitate spatial analysis of the inherent relationships and structure of the Body of Knowledge 1.

3.1.2 Interface Enhancements

BoKVis is an ongoing project and there is constant work being done to fix bugs and optimize performance quality. For user testing, the application needed to be in a stable state where users of all skill levels could understand and use the application without encountering too many issues. Development of a hover-activated help menu was the first development towards this end. A class was created to incorporate “IFGraphics” or interface graphics, which were compiled in an XML document (Figure 12). The new class would then read the XML document to determine which tool-tip would be displayed for each external interface control. The next step in modifying the applications for use in user testing was to separate the visualization methods from one another. The goal of the test is to compare the spatial layouts, so we wanted the user to only be able to access one layout at a time. To do this, the other visualization method buttons were grayed out, and the Tool Tip was adjusted to say that those methods were “deactivated”. The code had to be slightly adjusted to make sure that all interactive functionality was still enabled for all visualizations. At this stage it was decided
that the Similarity Graph wouldn’t be used for testing purposes as it wasn’t developed
enough to stand alone without the other methods for solving problems related to the BoK;
this was decided after the first round of focus group testing. After the remaining three
methods were modified for testing, they were exported to applications (for both Mac and
Windows) from Processing and compressed into zip files. They were then put into a
Dropbox for access for remote users along with instructional documents.

```xml
</HelpElement>
<HelpElement>
  <METexts>Adjust the size of the boxes surrounding the BoK elements. </METexts>
  <MName>imgBoxHeight.png</MName>
  <!-- Element METools, minOccurs=unbounded -->
  <METools>
    <METool>
      <ToolText>Small Box Size</ToolText>
      <PosX>780</PosX>
      <PosY>15</PosY>
      <Width>30</Width>
      <Height>10</Height>
    </METool>
    <METool>
      <ToolText>Medium Box Size</ToolText>
      <PosX>780</PosX>
      <PosY>26</PosY>
      <Width>30</Width>
      <Height>10</Height>
    </METool>
    <METool>
      <ToolText>Large Box Size</ToolText>
      <PosX>750</PosX>
      <PosY>37</PosY>
      <Width>30</Width>
      <Height>10</Height>
    </METool>
  </METools>
</HelpElement>
</HelpElements>

Figure 12: Example of XML Structure for Help Menu Elements

3.2 HUMAN SUBJECTS TESTING

The user survey was designed to see if test subjects were able to answer questions
regarding the BoK using only one spatial layout of the BoK to solve the problems. It aimed
to incorporate both multiple choice and open ended questions, providing “real world
scenario” questions and an opportunity to leave feedback. After the survey was developed,
small focus group studies were conducted to get initial feedback regarding the user interface,
allowing for correction of any bugs and potentially confusing elements prior to implementing
the user survey.

### 3.2.1 User Survey Creation

The user survey was designed to test the user’s ability to perform particular cognitive
tasks and accomplish particular conceptual goals such as locating, identifying and
categorizing (Koua, McEachren, & Kraak, 2006). Previous research in visualization has
shown that general cognitive tasks (such as query/navigation, search, match, analyze and
infer) can be translated into more specific tasks which will test that particular notion (Koua et
al., 2006). For example, to test the user’s ability to analyze the structure of the BoK, the
survey would ask the user “How many units does the knowledge area ‘Organizational and
Institutional Aspects’ contain?”, thus prompting the user to utilize the interactive
functionality to obtain that information (Figure 34). Synnestvedt and Chen (2005) argue that
simple tasks such as locating and identifying are often included in usability studies, whereas
more conceptually difficult tasks (for example: compare, associate, rank, cluster, correlate,
categorize) are often left out. This survey will include a sample of these tasks in multiple
choice format, and will also test for the participants’ ability to complete such tasks based on
mappings onto the BoK in an open-ended format.

To create the electronic survey in Flash, ActionScript 3.0 was utilized to implement
the questionnaire, move forward in the survey with buttons, and record survey answers
through radio buttons, check boxes and input text boxes. Each answer was assigned a value
as a global variable to later be called and sent through the network with a PHP file and stored
into a MySQL database on the SDSU Geography Department Server (Figure 13).
In the survey, users were first asked a series of multiple choice questions about themselves (experience with computers, prior knowledge of BoK, etc). Next, they were instructed to explore the BoKVis application and asked multiple choice questions pertaining to the BoK interface. The survey concludes with open ended questions about using the BoK in real-world scenarios, and an opportunity to leave feedback. Answers to all questions were recorded in the MySQL database, and additionally, timestamps were taken for each answer. Each user was provided with a survey ID number, provided on the instruction form, to indicate which of the three spatial layout methods was used to answer the questions on the survey.

3.2.2 Focus Group Testing

Focus group testing was performed to help improve upon the survey design. For this initial test, a paper survey form was created with sample questions relating to the GIS&T BoK to start honing in on what types of questions would yield interesting and/or thought provoking results. The main goals of this round of testing were to improve and reflect upon the survey questionnaire and to identify bugs in the BoKVis Application.

The participants were to read and follow the instructions on the paper survey and answer them using the body of knowledge visualization. For this round of testing, participants were able to use all four spatial layouts to find answers. They were advised to complete the survey, while taking notice of any glitches in the application. Additionally, they were asked to
identify questions that don’t make sense are too easy/too difficult and also provide suggestions on how the survey could be improved in further development.

Participants in this preliminary focus group were Dr. Skupin’s students, who meet weekly to discuss research projects. The group of three students were already familiar with the BoKVis project and thus were able to offer greater insight on survey design.

The main takeaway from this session was that the questions needed to be redesigned to require the user to use the program to solve relevant problems, or real-life scenarios where the BoKVis tool may actually be used. Multiple choice questions are useful for testing the basic functionality of the application (i.e.: navigation, query), but lack the depth to assess some of the more complex facets of the Body of Knowledge which require open-ended problem solving, and may or may not have one definitive answer.

Other feedback identified some bugs in the application. Additional feedback about the user survey was that there should be fewer multiple choice questions, because they tend to feel redundant and the wording is sometimes confusing. The multiple choice questions should be narrowed down to a select few which serve a distinct purpose (test to see if user understands the hierarchical structure of the BoK, or test if they understand how to navigate), before digging deeper to see if the user is able to use the interface to help find a solution for a real-world scenario.

This feedback influenced the redesign of the survey and application. SimGraph was excluded for testing until it has been further developed. Bugs in the interface were identified and fixed before the next round of testing and the survey was remodeled to include open ended questions. Two of the Multiple Choice questions were decidedly put first to serve as a “pre-test” to see if the users had basic knowledge of the structure of the BoK and the capability to navigate the interface.

This second session of focus group testing aimed to test the computer-based survey for functionality and identify areas of improvement in both the survey and BokVis application. The survey was created using Adobe Flash Professional and can be run through an internet browser or as a standalone application.

Participants were students enrolled in Geography 780, the Knowledge Visualization Seminar. They had a basic knowledge of the Body of Knowledge project, through brief introduction during the seminar, but had not worked with the BoKVis interface prior to this
focus group test. The students were instructed to open the BoKVis application. All three visualizations were activated so students could choose which they wanted to use to find answers. They then were given the file with the standalone Flash application of the survey to use for the testing.

The main takeaway from this focus group was that the survey needed more information about the study (what is the Bok? What is BoKVis used for?) as well as more instruction on how to use the application.

Feedback:

Application:
- “How do I collapse a branch?”
- Need explanation for the calculations: cumul? avg?
- When there is an element mapped on the BoK, what does the grey color mean?
- “First open ended question: what do the colors mean? Maybe the legend could explain that red means, the applicant is better that requested and blue means that he is not good enough.”

Survey:
- What is the study about?
- “The first 2 questions (GIS vs Geography) are very similar. I needed to start the test 3 times to realize the difference, so a user might answer the geography question differently if he/she knew there was a GIS question following it”

All of this feedback was incorporated into the application and survey before proceeding to the next stages. Additional instruction was added into the survey to provide more guidance on how to use the interface: specifically, instructions on how to open/close nodes, the meaning of the divergent color scheme, and a summary of the calculations used for cumul/avg. All of these instructional points could be later incorporated into a web-interface for the application, but since the human subjects testing is using the standalone application, it requires further explanation via the survey interface. An instructional document about the GIS&T Body of Knowledge,
BoKVis and the user study was created as a handout for participants to introduce the project prior to testing. Also, navigation tools (such as back buttons) were added to the survey, as per the request of the participants in this focus group. The back buttons are available for the user during the multiple choice questions asking for personal information in case the user would like to change an answer to a question, but the back buttons are not available during the timed multiple choice questions pertaining to the BoK.

### 3.2.3 USER TESTING

All of the controlled User Testing was done in the Visualization lab in Storm Hall (339) over the course of 3 weeks. This environment provided a controlled, intimate setting, where the tests could be set up for the subjects upon their arrival. The computers used for testing all had dual monitors, which was useful for simultaneously viewing the application and the survey interface. All of the computers had Windows 7, 64 bit OS and 24 GB of RAM. Seven participants could take the survey at once, under observation of the test administrator. The application was preloaded onto the desktop, as well as a shortcut to the user survey. Paper versions of the instructional document and consent form were provided at their workstation and they were instructed to go through the documents prior to beginning the survey (Appendix B). Participants were recruited from many sources:

- **Recruiting for controlled testing:**
  - SDSU Geography Courses:
    - Geography 101- Earth’s Physical Environment Lab (3 Sections)
    - Geography 321- United States
    - Geography 354- Geography of Cities
    - Geography 380- Map Investigation
    - Geography 385- Spatial Data Analysis
    - Geography 581- Cartographic Design
    - Geography 583- Internet Mapping and Distributed GIS
    - Geography 584- GIS Applications
    - Geography 596- GIS for Business Decision Making

- **Recruiting for remote testing:**
  - San Diego Regional GIS Council Meeting
This extensive recruiting provided a large sample size from various levels of experience and expertise. Students were given an oral explanation of project, BokVis and potential benefits of participating. The instructional document has general information as well in case users are testing remotely. They were instructed to open the application and user survey once they have had a chance to read through the documents then follow instructions on the survey from that point forward. The test proctor walked the room taking observations of people’s interactions, questions they may have and areas they get stuck and take note on how they go about solving problems.

Throughout the testing process, a few observations came to the forefront as subjects struggled with the same main issues with the application. Foremost, most users didn’t seem to use or understand the external controls, such as the level navigation, text query or dropdown menus. For example, they wouldn’t use the query feature, but would rather guess or “explore” to find answers to questions. The dropdown menu proved to be another difficult aspect of the interface, as it doesn’t look like a traditional dropdown menu that one would see in a Windows or Mac environment. It was also observed that subjects were not quick to use the navigation between the levels. With all of these issues combined, it can be assumed that the toolbar containing these controls is not user friendly and does not invite interaction, unlike the main BoK visualization. Verbal feedback from participants revealed that an optional video tutorial may be useful for helping to introduce the GUI. It was also suggested that a video demo would be helpful prior to starting the survey.

Subjects were observed utilizing the “full screen” mode for both the survey and application interfaces, indicating that resizability is a desirable feature in the application going forward. Most users explored the “box size” feature in Tree Graph and Indented List
in order to better see text. They also caught and brought to my attention a few bugs in the program regarding text overlap when using those methods.

Although addressed during focus group testing, the main question of “how to go back?!?” remained an issue. Additional instruction on using the shift-click to close nodes had been added to the survey instructions, however it wasn’t intuitive enough to where the users could figure out what to do if they had overlooked that comment in the directions.

Some recurring questions about the survey wording required additional explanation for some users. Specifically, what is meant by the wording of “overqualified” or underqualified” in the open ended questions. Also, the idea behind the scorecard wasn’t immediately understood by most users. Since this was the controlled portion of testing, the test proctor was able to clarify most of these questions for the subjects, however this ambiguity could potentially pose a problem for the remote users.

### 3.3 Analysis of Results

In total there were 75 surveys used for analysis: 25 for Indented list, 25 for Tree Graph, and 25 for Tree Map. The survey questionnaire contained 9 Multiple Choice questions about the participant, 5 Multiple Choice questions about the BoK, two “real-world scenarios” with multiple-selection (checkbox) responses and an open-ended text response, and lastly, an opportunity to leave feedback. The survey results data were processed to verify the validity of database results before proceeding with analysis.

One of the multiple choice questions was thrown out during preliminary data processing because only 9 subjects out of 75 got it correct. This question, “How many learning objectives deal with economics?” was likely difficult because of the limited functionality in all three spatial layouts for viewing all learning objectives, which was the primary manner in which to answer the question (Table 5). The program sometimes freezes when asked to display all learning objects, particularly in treemap mode (indicated in verbal feedback from two students during testing).

The distribution of the data were analyzed to identify potential outliers that may affect the results. All three methods had the same distribution range and survey accuracy scores were skewed in favor of higher scores, with the majority of scores falling in the 75%-%
100% range for all methods. Six surveys with low scores were identified as being outliers (Figure 14). The distribution of the total time taken to complete the survey was also analyzed. The Indented List method had the most variability in completion time. Four outliers were identified which had longer completion times, and all of those outliers received 100% survey accuracy suggesting a positive correlation between time taken to complete the survey and percent of questions correct (Figure 15). None of the outliers were removed from the dataset before statistical analysis because of the small sample size.

Figure 14: Boxplot distribution of Total Score by Method (outliers labeled with survey completion time)
After processing the data, two separate scores were calculated, one for percent of correct responses on the Multiple Choice section and one as a proxy for “correctness” on the multiple-selection questions. Additionally, the total time taken to complete each question was recorded, and a total survey time was calculated with the summation of these time stamps. In order to give every user an equal opportunity to explore the interface prior to beginning the multiple choice questions, one slide in the survey instructed the user to take as much time as they would like to explore the interface before proceeding onto the timed questions (Figure 33).

Once the data was processed, both qualitative and quantitative analyses were performed in order to begin developing answers to the posed research questions.
1. *Which of the spatial layouts is most effective for completing tasks related to the GIS&T BoK overall? How do they compare to each other in terms of performance?*

The Tree Graph had the highest average survey accuracy score (82%) for the multiple choice questions. The average score for subjects using the Tree Graph was slightly higher than the average scores of those using the other two methods, by about 3% (Table 2). 2-tailed Independent Sample T-Tests were computed for each pair of visualization methods to compare the means and test for statistical significance. As shown in Table 3 below, none of the tests yielded statistically significant results, indicating that there is not enough information to reject the null hypothesis that there is a difference between the means that couldn’t be accounted for by random chance. Thus, we cannot infer with any certainly that using a particular method would result in a better or worse results on the multiple choice portion of the survey.

**Next, Pearson correlations were computed to find the direction and strength of the relationship between survey accuracy and response time (**

Table 4). There is, in fact, a strong positive correlation between the total time the user took to take the survey and their final score results. This is supported by a Pearson’s Correlation of .368, significant at the 98% confidence level. This could be attributed to the user’s time spent exploring the interface. If a user were taking their time, they would be better able to complete tasks and solve problems using the tools, than a user that was trying to just get to the answers without having a complete understanding of the tool’s functionality. Based on the minimum/maximum completion time observations, it seems that users spent a bit more time using Tree Graph than the other methods, which may be a factor in the higher average scores. Neither Indented List or Tree Graph had a statistically significant correlation between survey completion time and survey accuracy, however the treemap method had a very strong positive correlation, possibly a pointer to the steeper learning curve. As suggested in the research hypotheses, the treemap may be less efficient (longer response times) for completion of cognitive tasks due to the more complex spatial layout, so this correlation indicates that a user who spent longer familiarizing themselves with the treemap is more likely to be able to answer the multiple choice questions correctly.
### Table 2: Survey Accuracy and Completion Time by Method

<table>
<thead>
<tr>
<th>Method</th>
<th>Average Percent Correct</th>
<th>Average Completion Time</th>
<th>Min Completion Time</th>
<th>Max Completion Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indented List (25)</td>
<td>77%</td>
<td>14.95</td>
<td>5.37</td>
<td>30.78</td>
</tr>
<tr>
<td>Tree Graph (25)</td>
<td>82%</td>
<td>14.45</td>
<td>6.18</td>
<td>43.15</td>
</tr>
<tr>
<td>Tree Map (25)</td>
<td>79%</td>
<td>12.47</td>
<td>3.13</td>
<td>29.83</td>
</tr>
</tbody>
</table>

### Table 3: Independent Sample T-Tests to test if the differences between the means are statistically significant between methods

<table>
<thead>
<tr>
<th>Method</th>
<th>N</th>
<th>Mean</th>
<th>Independent Sample T-Test Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indented List</td>
<td>25</td>
<td>77%</td>
<td>0.549</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>82%</td>
<td>.325</td>
</tr>
<tr>
<td>Tree Map</td>
<td>25</td>
<td>79%</td>
<td>.681</td>
</tr>
</tbody>
</table>

### Table 4: Correlation between Total Survey Time and Average % Correct on Multiple Choice portion

<table>
<thead>
<tr>
<th>Total Time/Survey Accuracy (% Correct on Multiple Choice)</th>
<th>Correlation</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Surveys (75)</td>
<td>0.368</td>
<td>0.001</td>
</tr>
<tr>
<td>Indented List (25)</td>
<td>0.266</td>
<td>0.198</td>
</tr>
<tr>
<td>Tree Graph (25)</td>
<td>0.165</td>
<td>0.43</td>
</tr>
<tr>
<td>Tree Map (25)</td>
<td>0.588</td>
<td>0.002</td>
</tr>
</tbody>
</table>
2. *Is one spatial layout significantly more or less effective than others for completing a particular cognitive task?*

MC1 and MC4 (Table 5) test the ability to understand the structure of the BoK by asking questions about relationships between elements. Tree Graph and Indented List performed best of the three methods for answering questions related to the structure of the BoK. (Table 6). For MC1, Indented List and Tree Graph both averaged 96% correctness- 11% higher than the treemap. Independent T-tests confirm the statistical significance of this difference at the 98% confidence interval (Table 7). For question MC4, Tree Graph and Indented List again had higher scores than treemap, however only the difference between Tree Graph and Treemap was statistically significant (Table 9). We can infer based on these questions that Tree Graph and Indented List yielded better results than Treemap in questions related to understanding the structure and relationship between BoK elements.

MC3 and MC5 (Table 5) test the ability to navigate the interface to locate a particular term or concept, which can be done in a few different ways (manual expansion of nodes, text query). It also requires the user to become familiar with the legend and aggregation functionality in order to view the answers at the Knowledge Area level. Looking at the Average Scores by method, Treemap seemed to perform best of the three methods for completing this type of task, followed by Tree Graph, then Indented List (Table 6). Independent T-tests were run for both questions and yielded interesting results. For question MC3, none of the results were statistically significant, meaning we cannot make any meaningful conclusion about a particular method performing better than the other for solving this question (Table 8). The Average Scores on MC3 were notably lower than the rest of the multiple choice questions (in the 50%/60% range), which could be the product of the users not understanding how to solve the question. While observing the subjects taking the test, there did seem to be some overall confusion with how to solve for this question. Many subjects were doing “trial and error” clicking through the individual nodes rather than using the query tools in the toolbar. This general confusion is also reflected in the time stamp for these question types which was higher than the questions about the structure of the BoK. (Figure 15) Further statistical analysis of the variation in time stamps between questions showed that these differences were not statistically significant, however. (Table 11)
Scores increased significantly between the first question of this type (MC3) and the second (MC5), perhaps indicating that the user had familiarized themselves more with the interface and thus was able to come to an answer more effectively. Independent T-tests for MC5 concluded that Indented List underperformed for answering this question in comparison to Treemap (statistically significant at the 98% confidence level). A potential reason for this score deficit could be the bugs in the scroll bar functionality, which may have been discouraging to users who were relying on expanding individual nodes to find answer. The difference in means between Treemap (92%) and Tree Graph (84%) was not significant, so no assumptions can be made about their functionality in completing this type of task.

Table 5: Multiple Choice Questions and ID (In the order in which the subjects saw the questions)

<table>
<thead>
<tr>
<th>ID</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC4</td>
<td>How many Units does the Knowledge Area &quot;Organizational and Institutional Aspects&quot; contain?</td>
</tr>
<tr>
<td>MC1</td>
<td>Which element is the parent element of the other: Database Design or Conceptual Model?</td>
</tr>
<tr>
<td>MC3</td>
<td>In which knowledge Area can one find the most information on accuracy?</td>
</tr>
<tr>
<td>MC2</td>
<td>How many learning objectives deal with economics?</td>
</tr>
<tr>
<td>MC5</td>
<td>In which of the following Knowledge Areas can you find out about the role of economics?</td>
</tr>
</tbody>
</table>

Table 6: Survey Accuracy- Percent of subjects with the correct answer, by question

<table>
<thead>
<tr>
<th>Method</th>
<th>MC1</th>
<th>MC3</th>
<th>MC4</th>
<th>MC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indented List</td>
<td>96%</td>
<td>56%</td>
<td>80%</td>
<td>76%</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>96%</td>
<td>60%</td>
<td>88%</td>
<td>84%</td>
</tr>
<tr>
<td>Tree Map</td>
<td>84%</td>
<td>64%</td>
<td>76%</td>
<td>92%</td>
</tr>
</tbody>
</table>
Table 7: Which is the parent element of the other: Database Design or Conceptual Model?

<table>
<thead>
<tr>
<th>Method</th>
<th>N</th>
<th>Mean</th>
<th>Independent Sample T-Test Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indented List</td>
<td>25</td>
<td>96%</td>
<td>1</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>96%</td>
<td>0.004</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>84%</td>
<td>0.004</td>
</tr>
<tr>
<td>Indented List</td>
<td>25</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>84%</td>
<td></td>
</tr>
</tbody>
</table>

Table 8: In which knowledge Area can one find the most information on accuracy?

<table>
<thead>
<tr>
<th>Method</th>
<th>N</th>
<th>Mean</th>
<th>Independent Sample T-Test Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indented List</td>
<td>25</td>
<td>56%</td>
<td>0.587</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>60%</td>
<td>0.574</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>60%</td>
<td>0.574</td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>64%</td>
<td>0.292</td>
</tr>
<tr>
<td>Indented List</td>
<td>25</td>
<td>56%</td>
<td></td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>64%</td>
<td></td>
</tr>
</tbody>
</table>

Table 9: How many Units does the Knowledge Area "Organizational and Institutional Aspects" contain?

<table>
<thead>
<tr>
<th>Method</th>
<th>N</th>
<th>Mean</th>
<th>Independent Sample T-Test Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indented List</td>
<td>25</td>
<td>80%</td>
<td>0.128</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>88%</td>
<td>0.028</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>88%</td>
<td></td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>76%</td>
<td>0.028</td>
</tr>
<tr>
<td>Indented List</td>
<td>25</td>
<td>80%</td>
<td>0.505</td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>76%</td>
<td></td>
</tr>
</tbody>
</table>

Table 10: In which of the following Knowledge Areas can you find out about the role of economics?
**Table 11: Paired Sample Correlations between Multiple Choice Question Time Stamps**

<table>
<thead>
<tr>
<th>Method</th>
<th>N</th>
<th>Mean</th>
<th>Independent Sample T-Test Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indented List</td>
<td>25</td>
<td>76%</td>
<td></td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>84%</td>
<td>0.084</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>84%</td>
<td></td>
</tr>
<tr>
<td>Tree Map</td>
<td>25</td>
<td>92%</td>
<td>0.165</td>
</tr>
<tr>
<td>Indented List</td>
<td>25</td>
<td>76%</td>
<td>0.002</td>
</tr>
<tr>
<td>Tree Map</td>
<td>25</td>
<td>92%</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 16: Average Minutes per Question - Multiple Choice**
<table>
<thead>
<tr>
<th>Pair</th>
<th>Comparison</th>
<th>N</th>
<th>Correlation</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td>MC1_time &amp; MC3_time</td>
<td>73</td>
<td>.139</td>
<td>.239</td>
</tr>
<tr>
<td>Pair 2</td>
<td>MC3_time &amp; MC4_time</td>
<td>73</td>
<td>-.170</td>
<td>.150</td>
</tr>
<tr>
<td>Pair 3</td>
<td>MC5_time &amp; MC4_time</td>
<td>73</td>
<td>-.030</td>
<td>.802</td>
</tr>
<tr>
<td>Pair 4</td>
<td>MC3_time &amp; MC5_time</td>
<td>73</td>
<td>-.011</td>
<td>.926</td>
</tr>
<tr>
<td>Pair 5</td>
<td>MC1_time &amp; MC4_time</td>
<td>73</td>
<td>-.126</td>
<td>.288</td>
</tr>
<tr>
<td>Pair 6</td>
<td>MC1_time &amp; MC5_time</td>
<td>73</td>
<td>.103</td>
<td>.384</td>
</tr>
</tbody>
</table>

3. **Is the user able to utilize the BoK as a basemap or reference system and make inferences based on BoK scorecard overlays?**

This question is intended to determine the impact of the visualizations for uses beyond exploration of the structure or content of the BoK by testing the user’s ability to make inferences about mappings onto the BoK. The questions are posed as “real-life scenarios”, requiring the participant to obtain answers using the BoK interface. The first scenario requires the user to copy and paste text (a job ad and resume) into the query boxes and compare the two via the visualization (Figure 17). The user is then prompted to infer in which Knowledge Areas the user is overqualified and underqualified in the position. The user is able to check the box of all or any of the 10 Knowledge Areas. Measuring accuracy or correctness for these questions is slightly subjective. An answer was deemed correct if the user had checked the boxes for the Knowledge Areas with the darkest hues. Allowing the user to check more or less boxes, however, provided more robust results because we are able to see the thought process behind the answers. For example, some users selected the “yellow-ish” hues in addition to the dark red or dark blue, indicating a potential lack in understanding of the meaning of the legend. This is addressed further in the qualitative feedback.
Figure 17: Visual results of copying/pasting the job ad (query 1) and resume (query 2) onto the BoK

For the first real-world scenario, the three methods yielded a wide variety of scores (Table 12). Tree Graph resulted in the highest scores (88%), followed by Indented List (76%), then Treemap (60%). The difference in these means was statistically significant across all methods at the 98% confidence interval. This is surprising because at the Knowledge Area level, Tree Graph and Indented List look quite similar, and would seemingly yield similar results, however the differences between all three method’s scores were notable. Interestingly, for the second question (Unqualified) regarding the same subject matter and requiring the same steps to solve, there were no statistically significant differences between the visualization methods. Regardless of method used to answer the questions, 86% of all of the subjects were able to answer at least one of these questions correctly, and 64% answered both parts correctly, indicating an ability to utilize the interface to solve a real world problem to some degree. However, throughout the testing, many users were observed experiencing issues interpreting the visual results which could account for the variance.
between the scores. Potential issues with these real-world scenarios are addressed further in the qualitative feedback analysis.

**Table 12: "In which Knowledge Area(s) is the candidate likely to be most qualified for this position?"**

<table>
<thead>
<tr>
<th>Method</th>
<th>N</th>
<th>Mean</th>
<th>Independent Sample T-Test Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indented List</td>
<td>25</td>
<td>76%</td>
<td>0.028</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>88%</td>
<td></td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td>Indented List</td>
<td>25</td>
<td>76%</td>
<td>0.024</td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>60%</td>
<td></td>
</tr>
</tbody>
</table>

**Table 13: "In which Knowledge Area(s) is the candidate likely to be least qualified for this position?"**

<table>
<thead>
<tr>
<th>Method</th>
<th>N</th>
<th>Mean</th>
<th>Independent Sample T-Test Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indented List</td>
<td>25</td>
<td>76%</td>
<td>1</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>76%</td>
<td></td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>76%</td>
<td>0.076</td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>64%</td>
<td></td>
</tr>
<tr>
<td>Indented List</td>
<td>25</td>
<td>76%</td>
<td>0.076</td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>64%</td>
<td></td>
</tr>
</tbody>
</table>

The question that proved most problematic was the second real-world scenario, which required the user to use two pre-scored Scorecards from the dropdown list, effectively comparing two Courses to one another. The user was asked to determine which course would be most useful if the student wished to learn more about two particular Knowledge Areas. The mean scores for correctness on this question were quite low (in the 40%/50% range), and there were no statistically significant differences between the methods in terms of correctness, indicating an overall lack of understanding of how to go about solving for this answer (Table 14). In observing the subjects, the most common question was how to get to
the scorecards, because the dropdown menu wasn’t apparent in the interface. Once the users were able to access the scorecards, however, they ran into issues with the aggregation of the scorecard (since it was scored at the Topic level and needed to be propagated up through the tree to be viewed at the Knowledge Area level). The combination of these two issues was likely the cause of the low scores across the board for the users.

Figure 18: Visual results of two scorecards (course 1 and course 2) mapped onto the BoK

Table 14: Independent Sample T-Test: Scorecard Question

<table>
<thead>
<tr>
<th>Method</th>
<th>N</th>
<th>Mean</th>
<th>Independent Sample T-Test Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indented List</td>
<td>25</td>
<td>40%</td>
<td>0.352</td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>52%</td>
<td></td>
</tr>
<tr>
<td>Tree Graph</td>
<td>25</td>
<td>52%</td>
<td>1</td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>48%</td>
<td></td>
</tr>
<tr>
<td>Indented List</td>
<td>25</td>
<td>40%</td>
<td>0.352</td>
</tr>
<tr>
<td>Treemap</td>
<td>25</td>
<td>48%</td>
<td></td>
</tr>
</tbody>
</table>
This scenario also prompted the user to type in a written response for the second portion, in order to elicit a reason for choosing a particular answer. This was not scored as part of the total correctness for this question but served as a sounding board to see the rationale behind the interpretation of the visualization.

About 10 users demonstrated a solid understanding of how to interpret the scorecard:
- “Course 1 would be more useful for this job applicant because the focuses of Course 1 include the two topics she needs to learn more about. I can tell this because they are red and orange which indicate course 1”
- “I think Course 1 would be most helpful because it contains more blocks with deeper colors than the other one for the areas of Organizational and Institutional Aspects and GIS&T and Society”
- “Course 1 because it would help her learn more about GIS&T and Society and Organizational and Institutional Aspects”

About half of the feedback listed Course 1 as the most useful course but lacked a detailed or complete explanation as to why:
- “Course 1 because it contains the most of what she is looking for”
- “Course 1 because it covers her topic area”
- “Course 1 meets more of the required qualifications than what she already knows.”

13 users left this portion blank, about 10 listed the Knowledge Area/Topic rather than the Course, and a few were unable to find an answer:
- “GIS and T and Society, Organizational and Institutional Aspects; This applicant lacks knowledge and/or competency in these two areas.”
- “I have no idea”
- “I’m really confused”
This confusion regarding how to use and analyze the scorecards is apparent when looking at the time stamps for the overall survey. Users took on average 2.5 times longer on the scorecard question than on the Multiple Choice or first Open Ended question. (Figure 19)

![Average Minutes per Question](image)

**Figure 19: Average Minutes per Question- Entire Survey**

4. *Which design aspects of the interface assist in carrying out the survey objectives? Which design aspects of the application detract from fulfilling the objectives?*

   The last portion of the survey invited the user to leave some feedback on what they liked best and least about their experience with the application. From the very diverse range of feedback offered, there were a few recurring themes which were extracted and are listed below in. They are listed in order of most common to least common. For qualitative feedback in its entirety, please see the appendix.

**Overall feedback (all Methods):**

1. Would have liked a demonstration of tool use prior to beginning survey.
2. How to expand and un-expand nodes? Shift-click is unintuitive.
3. How to use/find scorecard, sum/avg toggle
4. How to interpret legend
5. Liked using the text query
6. Confusing GUI  
7. Steep learning curve  
8. Liked the tree structure/organization  
9. Helpful to have two screens

*Indented List:*  
1. Scrollbar hard to use  
2. Easy interface to interact with  
3. Boring to look at

*Tree Graph:*  
1. “Use of the term map is confusing conceptually”  
2. “Program is a bit buggy-scrolling, overlapping text”

*Tree Map:*  
1. Cool design  
2. Hard to read text  
3. Takes time to get used to  
4. Overwhelming/cluttered  
5. Need some sort of status to show how deep in the tree

Besides the qualitative feedback, the users were asked to rate on a scale of 1-5 how easy they found the application to use. The results are shown in Figure 20 below. 3, or ‘moderately easy’ was the most common answer for all methods. There was no distinct pattern between methods on which one was easiest to use, most likely due to a variety of factors including the diversity of the test subject group. Only one user found this program to be “very easy” to use, and they were using the Indented List method. Interestingly, there was a very strong positive correlation (.425) between the perceived “ease” of using this program and the survey accuracy score, confirming that the users who did better on the survey thought the program to be easy to use.
Prior to beginning the Multiple Choice portion of the survey, the subjects were asked some questions about themselves, in order to get an idea for the composition of the test subject pool and better understand and interpret the results. They were asked to provide multiple choice answers to the following questions:

1. Level of expertise/understanding in Geography
2. Level of expertise/understanding in GIS
3. How often do you interact with a computer?
4. Age Group:
5. I am a(n): Undergraduate/Graduate Student/Professional/Other
6. Please Select your gender: M/F/Decline to respond
7. Have you ever been diagnosed, or do you think you could be color-blind or vision-impaired?
8. Did you know about the GISTBoK prior to taking this survey?

Pearsons Correlation scores were computed for all of these variables to see if any of them had an effect on a subject’s performance on the survey. Of the participants, 45 were male and 30 were female. The male’s average survey accuracy was 81% and the female survey accuracy was 75%, however this variation was not statistically significant between genders. There was a medium strength positive correlation between the subject’s level of
experience/understanding in GIS and their total score (.362 correlation- significant at the 99% confidence level), which intuitively makes sense because the user would already have a familiarity with the terminology used in the BoK, and may possibly have experience with cartography/visualization techniques. There was also a slight positive correlation with the age group of the user and their overall performance on the survey (.231 correlation- significant at the 95% confidence level). The user’s status (Undergraduate/Graduate Student/Professional or Other) also was an indicator of how well they would perform on the survey. There was a notable difference in Average Score from Undergraduate to Graduate, and after running a T-Test, it was deemed to be statistically significant difference in the means of the two groups (Significance of .022). The Average Score for the “Professional/Other” user is an impressive 100%, however because of the small sample size for that particular group, it was not a statistically significant variation.

Table 15: Human Subject Status

<table>
<thead>
<tr>
<th>Status</th>
<th>Avg Score</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate</td>
<td>72%</td>
<td>61</td>
</tr>
<tr>
<td>Graduate</td>
<td>87%</td>
<td>10</td>
</tr>
<tr>
<td>Professional/Other</td>
<td>100%</td>
<td>2</td>
</tr>
</tbody>
</table>

Colorblindness/vision impairment did not have a statistically significant effect on user’s performance on the survey, however that could be due to a small sample size. The 6 users that said they had been diagnosed as colorblind/vision impaired, and the 9 that said they were unsure scored an average of 70% on the survey compared to the overall average of 75% . Prior knowledge of the GIS&T BoK also didn’t have a statistically significant effect on total score, but is likely also due to small sample size. Intuitively, the results make sense, that a user with prior knowledge of the program/content would perform better overall.

Table 16: Human Subject's Prior Knowledge of the BoK

<table>
<thead>
<tr>
<th>Prior BoK Knowledge?</th>
<th>Avg Score</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>83%</td>
<td>17</td>
</tr>
<tr>
<td>No</td>
<td>74%</td>
<td>51</td>
</tr>
<tr>
<td>Unsure</td>
<td>54%</td>
<td>5</td>
</tr>
</tbody>
</table>
An additional sample of participants was recruited from various outside professional GIS/Geography societies and those volunteers were asked to take the test remotely. Of approximately 50 emails sent out, 8 additional surveys were completed and submitted. Unlike the main pool of users, 5 of the 8 subjects were “professional/other”. Due to this small sample size, data compared to the larger testing group did not produce statistically significant results, yet uncovered some interesting perspectives and perhaps will encourage further testing within the larger GIS community. This is a sample of some of the useful feedback from Professional/Other users from the remote testing:

**Professional/Other User using Treemap:**

"Works quickly to maps keywords/concepts to the BOK areas. Able to show several levels of hierarchy (and quickly toggle between) in the treemap. Interesting to consider as a tool for comparing text documents, as with the examples. Some additional guidance on navigation and menu options (Start Here) or Help option would be useful. The survey instructions guided through without problem. Some 'overplotting'. The Area headings plot over some of the Unit descriptions. Other visualization methods (tree graph, sim graph, indented list) were not available. Sometimes the toggle between Navigation levels with both queries seemed to have some quirks in display (why don't they show at the Area level-unless cumulative?)."

"The interface is nice, but restricted to the screen size. It would be great if one could 'zoom out' and pan around at full detail, with a much larger canvass, and with more flexibility in labeling and text size. At the Unit, Topic and Objective level- some of the hierarchy is lost, it is difficult to retain the original Area distinction- bolder lines?"

“I think that there is a lot of possibility for this tool and it could be a lot easier but there really wasn't a tutorial so it took quite a bit to get the hang of what I was doing. Additionally, right from the start we were asked really specific questions and I didn't even know what I was doing at first. A short tutorial on what each button does and how to interact with this tool would completely solve this issue and I can see the utility of the tool but at the moment it is really hard to interact with considering no instruction. With more support I think this will be great. Train users more. Give more examples. Then I think there will be a lot of utility in this.”
Professional/Other user using Indented List:

“I liked NOTHING about this tool. The layout is very basic & reminds me of using old UNIX based inventory software. Unix based software, just as it seems this BoKVis software is, requires very specific, verbal, instructions because the program has limited functionality & tasks must be performed in an exact order of operations. There is no 'Help' function to learn on your own. The selection: 'Visualization Method,' only allowed me to see the Indented List & not the Tree Map and Graph which might have been easier to understand. There seems to be no simple Search function to help you narrow down your criteria if you do not know exactly what you are searching for or how to search for it.”

“The ability to enter text and have it mapped out is pretty amazing. There are so many different applications for this. Without a prior understanding of some of the options at the top, it can be a little hard to navigate. The organization/visualization of the bok was very good”

Because of the diverse range of perspectives shown here in this small sample of subjects from the GIS community, further development of the application to address some of the feedback presented in this research and further testing would be an anticipated next step.
CHAPTER 4

DISCUSSION AND CONCLUSIONS

The statistical analysis helped to obtain meaningful conclusions regarding the performance of the three spatial layouts, thus developing an understanding of the relative value of different approaches to BoK hierarchy visualization.

4.1 FINAL RESULTS

We are unable to reject the Null Hypothesis (HoA) that there are no statistically significant differences among the three visualizations in terms of survey accuracy. This was confirmed with Independent T-tests based on subject’s percent accuracy on the multiple choice questions. Tree Graph and Indented list had significantly higher scores than the Treemap for questions about the structure of the BoK specifically. No conclusions could be made about which of the three spatial layouts was the most effective for locating a particular term or concept, likely due to an overall confusion with the functionality of the external controls. For the real-world scenario questions, Tree Graph returned the highest survey results for the first question asking about the applicant’s qualification for the open position, but for the remaining questions, none of the three spatial layouts proved to be better or worse than the others for correctly answering the questions. Qualitative feedback confirmed that these questions were confusing to the user, and the interface seemed unintuitive for helping the user solve for the answer.

We accept the Null Hypothesis (HoB), that there were no statistically significant differences between the three visualizations in terms of recorded time to correctly answer questions (time stamp). No concrete inferences can be made from this hypothesis, which was intended to determine which spatial layout was most efficient and effective by examining how long it took users to complete tasks. Although we accept this Null Hypothesis, it should be noted that there was a significant positive correlation between total overall score on the survey and total time taken to complete the survey. The relative strength of the correlation
varied between spatial layouts, however; the Treemap method had a strong positive correlation between survey completion time and total score (Table 4), indicating that a user who spent longer familiarizing themselves with the treemap is more likely to be able to answer the multiple choice questions correctly. The other two methods had positive correlations as well, but they were not statistically significant.

We cannot accept Alternate hypothesis 1 (HA1) that the treemap proved less efficient (longer response times) for completion of cognitive tasks due to the more complex spatial layout, because the response times were not significantly longer or shorter than the other methods. Treemap had higher average scores for questions requiring users to locate specific terms or concepts. These results were just short of statistical significance, but may be worthwhile to look into further in future work.

The qualitative feedback from this study revealed that the BoKVis interface needs some redevelopment to better assist the user. There were no statistically significant differences between the visualizations in terms of which was considered “Easiest” to use. However a positive correlation between perceived “ease” and total score showed that users who found the program to be easy to use obtained higher scores relative to those who did not. All three spatial layouts received roughly equal scores across the board. Qualitative feedback tells us that the external controls built into the user-interface need to be redesigned to be more intuitive, and some sort of introductory tutorial should exist for the user to watch/read should they choose to do so.

4.2 LIMITATIONS

During the testing process, many limitations were made apparent, which must be considered when assessing these results. The statistical results are limited due to small sample size. Many of the statistical analyses were not statistically significant, but may have been should the sample size have been larger, particularly in the case of the remote user testing component. Also, the majority of the subjects were selected from Geography courses at SDSU which leads to a limited, and potentially biased sample population. The remote testing portion attempted to account for that limitation, but due to low numbers of participants, it didn’t help to diversify the sample group as planned. Remote testing participation may also have been limited by compatibility issues, such as the necessary
Java/Flash requirements. Also, the language used in the survey was considered to be confusing by some subjects, which may have had an effect on results.

The BoKVIs applications also have limitations in their functionality, which will guide plans for future development of the program. In the treemap, it is difficult to display many of the lower levels without some textual content becoming overcrowded and perhaps altogether compromised. Although it’s possible to display all of the Learning Objectives at once, the response time of the application becomes much slower. In the Tree Graph and Indented Hierarchy, it is difficult to see all of the content when the lower levels are expanded, so the user may have to scroll very far to find particular elements at the lower levels. This accounted for a multiple choice question being thrown out of the analysis due to the difficulties finding that answer. Much of the other feedback that subjects offered while interacting with BoKVIs helped to identify aspects to be considered for future development.

Also, feedback from subjects indicated that users have a desire to be able to choose between the various visualization methods, and felt restricted by only having the option to use one spatial layout during the test. Future tests could be designed such that the users have the option of selecting which layout they preferred for various tasks, rather than testing the spatial layouts independent from each other.

4.3 Conclusions

From this study emerged some possible future research avenues. Future research stemming from this project could compare content based visualizations of the BoK to the structure-based visualizations, test various GUIs for use with the BoKVIs and perform more user testing within the professional GIS community. In future tests, web-based data collection methods, such as online surveys, would be instrumental in targeting a larger audience to avoid issues with statistical insignificance due to small sample size. This thesis, as situated within the larger BoK2 project, provides one of many interfaces in which to interact with the BoK. BoK2 is meant to be inclusive, and allows for a more diverse creation of the BoK ontology, rather than the original expert-generated BoK which could not be interacted with in a dynamic or collaborative fashion. The BoKVIs project will facilitate this move towards a more accessible GIS&T BoK. To this end, user-testing incorporating the various components of BOK2, such as the dynamic wiki, would help to guide visualization
efforts going forward on a broader scale. It would be useful to see if users are able to not only understand the BoKVis tool, but the integration with the larger Body of Knowledge project. The results of this study suggest that the Tree Graph is the most effective means in which to interact with the BoK of the three spatial layouts presented here. Further testing is needed to establish which visualization method – including others not investigated here – may be preferable for particular tasks.

BoKVis is valuable as a model for other bodies of knowledge and also in the professional realm for visualizing educational standards and qualifications, which exposes the potential to be used beyond academia. Upon further development, the BoKVis tools and associated computational procedures could enable a wide range of applications for the GIS&T BoK reference system. With the domain of GIS&T rapidly changing and evolving, now more than ever is this research necessary for capturing and utilizing these dynamic changes in this rapidly growing community.
REFERENCES


Skupin, A. (2014). *Making a Mark: a computational and visual analysis of one researcher’s intellectual domain*.


APPENDIX A

FLASH SURVEY
The following research study is being conducted by Marilyn Stowell of the San Diego State University Geography Department in fulfillment of her Master’s Thesis research.

Note the following points about the study:

- The study seeks participants over the age of 18 with background/training in geography, either enrolled in a geography course at SDSU or involved in the professional field to complete a short computer-based exercise.
- Each participant will be asked to complete a short pre-test, engage in the interactive visualization in order to answer questions regarding the subject matter, and then provide feedback and reactions about their experience using the interface. The entire process should take approximately twenty to thirty minutes.
- Participation is 100% voluntary for both students and outside volunteers.
- Participation in this study is anonymous. Your personal information, including your name, will not be collected or linked to study materials. Additionally primary documents collected after the study will be destroyed once the statistical analyses are complete.
- For student participants: any extra credit offered will be assigned at the professor’s discretion. Should you discontinue the survey, your professor will offer an optional extra credit assignment if he/she has offered any.

If you have any questions about this study please feel free to contact:
Marilyn Stowell
Geography Annex, 108
San Diego, CA 92182-6220

stowell@rohan.sdsu.edu or mrstowell@gmail.com

Or, if you have questions about your rights as a participant in this study you may contact the:
Institutional Review Board
Graduate and Research Affairs
Student Services East # 1410
San Diego, CA 92182-6220
(619) 594-6622

Figure 21: Participant Agreement/Qualification
Survey Information

Which survey ID number is at the top of your instructional document?

- [ ] 1
- [ ] 2
- [ ] 3

(This number can be found on the instruction sheet titled ‘instructions.pdf’, and is located at the top of the document, shown circled in red in the example below)

Survey ID: 1

Welcome to the Body of Knowledge Visualization user testing!

Some quick background about the Body of Knowledge:

Published in 2006, the Geographic Information Science and Technology Body of Knowledge (GIS&T Body) has become one of the most influential documents in the area of GIS&T education, both in the United States and internationally. It was created by the University Consortium for

This data will only be used in analysis of survey results and is confidential.

Figure 22: Spatial Layout associated with survey
Personal Information
Please answer the following questions about yourself:

Please select your level of expertise/understanding in Geography:

- [ ] I've never taken a Geography course before
- [ ] I've taken a few courses
- [ ] I feel confident in my understanding of Geographic concepts and principles
- [ ] I believe I can work as a professional in some aspect of this field

This data will only be used in analysis of survey results and is confidential.

Figure 23: Experience/Expertise in Geography
Personal Information
Please answer the following questions about yourself:

Please select your level of expertise/understanding in Geographic Information Science and Technology (GIS&T)
- [ ] I've never taken a course within the domain of GIS&T before
- [ ] I've taken a few courses in this area
- [ ] I feel confident in my skills and understanding of GIS&T
- [ ] I believe I can work as a professional in some aspect of this field

This data will only be used in analysis of survey results and is confidential.

Figure 24: Experience/Expertise in GIS
Personal Information
Please answer the following questions about yourself:

Which answer below best describes how often you interact with a computer?

- ☐ Never
- ☐ Monthly
- ☐ Weekly
- ☐ Daily
- ☐ Multiple times a day

This data will only be used in analysis of survey results and is confidential.

Figure 25: Computer Interaction
Figure 26: Age Group
Personal Information
Please answer the following questions about yourself:

I am a(n):
- Undergraduate Student
- Graduate Student
- Professional/Other

This data will only be used in analysis of survey results and is confidential.

Figure 27: Educational/Professional Status
Personal Information

Please answer the following questions about yourself:

Please select your gender:

- Male
- Female
- Decline to Respond

This data will only be used in analysis of survey results and is confidential.

Figure 28: Gender
Personal Information
Please answer the following questions about yourself:

Have you ever been diagnosed, or do you think you could be color-blind or vision-impaired?

☐ Yes
☐ No
☐ Unsure

This data will only be used in analysis of survey results and is confidential.

Figure 29: Color-blind/Vision-Impaired
Figure 30: Prior GIS&T Knowledge
Body of Knowledge Background Information

The Geographic Information Science and Technology Body of Knowledge (GIS&T BoK) has become one of the most influential documents in the area of GIS&T education, both in the United States and internationally. In order to develop “Model Curricula”, or a vision of how higher education should prepare students for success in the variety of professions that rely upon geospatial technologies, a comprehensive Body of Knowledge was created by GIS experts, that specifies what aspiring geospatial professionals need to know and be able to do. The interactive visualization you are about to use gives you access to all of the contents of this Body of Knowledge.

The BoK consists of a four-level hierarchy of:
1. Knowledge areas
2. Units
3. Topics
4. Learning objectives

Figure 31: GIS&T BoK Introductory Information
Survey Procedure

1. First, you will have an opportunity to take your time and familiarize yourself with the interface.

2. Then, you will be asked a series of questions which will require you to interact with the visualization in order to obtain the answer. You can take as much time as you like to answer the questions.

Figure 32: Survey Procedure
If you haven't done so already, open the BoKVis application or applet and position it in the most easily accessible location on the screen. Please take some time to familiarize yourself with the interactive interface. When you are ready to proceed with the survey, press "Next". Thanks!

Note: Left mouse click will expand an element to reveal its children, and Shift+Click will close the element.
BoKVis Survey

Please answer the following question by using the BoKVis interface:

How many Units does the Knowledge Area "Organizational and Institutional Aspects" contain?

☐ 5
☐ 6
☐ 7
☐ 8
☐ 10
☐ 12

Figure 34: MC4
BoKVis Survey

Please answer the following question by using the BoKVis interface:

Which element is the parent element of the other: "Database Design" or "Conceptual Model"?

- Database Design
- Conceptual Model

Figure 35: MC1
BoKVis Survey
Please answer the following question by using the BoKVis interface:

In which of the following Knowledge Areas can you find out about the role of economics?

- Data Modeling
- Conceptual Foundations
- GeoComputation
- Cartography and Visualization
- GIS&T and Society
- Data Manipulation

Figure 36: MC3
BoKVis Survey
Please answer the following question by using the BoKVis interface:

How many Learning Objectives deal with economics?

- 1
- 3
- 5
- 7
- 11
- More than 11

Figure 37: MC2
BoKVis Survey
Please answer the following question by using the BoKVis interface:

In which Knowledge Area can one find the most information on accuracy?

- Conceptual Foundations
- Data Manipulation
- Geospatial Data
- Cartography and Visualization
- Organizational and Institutional Aspects
- Design Aspects

Figure 38: MC5
One use for the Body of Knowledge visualization is the ability to "map" different entities onto it and use it as a basemap or frame of reference. You can map one item, such as a text query or scorecard, which will appear in just blue or red, or two items to see the difference between them, using a red/blue diverging color scheme.

The next few questions introduce real-world scenarios involving entities that can be mapped onto the BoK. Use the interface to help answer the following questions.

Figure 39: Introduction to real-world scenario questions
**Scenario:** A GIS professional found a job ad online and is interested in seeing if her qualifications (as indicated in her resume) would meet the needs of this potential employer. Copy (Ctrl-C) and paste (Ctrl-V) the contents of job ad and resume into the two text query boxes, respectively. After pressing "Enter", you should see a map with a mix of red and blue tones. Proceed to the next page to answer the questions based on the displayed results.

**GIS Professional: Resume**

**Skills and Experience:**

**GIS Applications:**
- Able to recognize and define geographic problems, to apply methodologies that permit analysis of the problem, to design a series of analytical steps, and to implement a solution using GIS software through prescriptive and descriptive modeling techniques. Able to manage geographic data input, storage, analysis, and output. Knowledge Visualization Seminar: Can conceptualize both vector and network spaces, utilizing various tools to produce visualizations. Familiar with Self-Organizing Maps, Dimensionality Reduction techniques, and network creation based on various data sources.
- Spatial Demography Seminar: Integrated mixed methodology using demography, spatial statistics and GIS/remote sensing to analyze spatial phenomena.
- Spatial Data Analytics: Capable of interpreting, understanding and interpreting underlying patterns and trends in spatial relationships as defined by calculation of spatial autocorrelation, interpolation and creating maps.

**GIS Specialist Position: Job Ad**

This is a technical position serving as the GIS resource for spatial data processing and reporting, including transportation network analysis, thematic mapping, database management, data collection and data input. Job tasks: Develops and maintains the Geographic Information System (GIS) in cooperation with other public safety agencies, counties and private entities; provides digital maps, geographic databases and data files. Designs, creates, maintains and disseminates maps and databases. Maintains all GIS databases for accuracy and data integrity. May perform fieldwork to verify accuracy of both spatial and attribute data. Performs complex spatial data processing involving geodatabase management, data collection, detailed sitting reporting, and transportation network analysis. Develops and implements standards, processes and procedures for data input and maintenance. Develops and maintains the fire agency performance reporting system for extracting, processing, storing, manipulating, and analyzing data from the CAD (Computer Aided Dispatch) system and RMS (Records Management System). Writes queries, prepares forms and reports for agencies allowing customer access to reporting system information. Provides training and technical assistance to agencies in the use of GIS data and computerized mapping programs. Collects and edits geographic data themes such as building footprints, street centerlines, trails, addresses points and other various layers. Performs other duties related to GIS for Public Safety dispatch and other responsibilities as assigned. The candidate must have a comprehensive knowledge of GIS, including computers and related GIS tools, equipment and software.
BoKVis Survey: Open Ended
Please answer the questions based on the results, then click Next to continue:

In which Knowledge Area(s) is the candidate likely to be overqualified for the position?
- Analytical Methods
- Cartography and Visualization
- Data Modeling
- Geocomputation
- Conceptual Foundations
- Design Aspects
- GIS&T and Society
- Organizational and Institutional Aspects

In which Knowledge Area(s) is the candidate underqualified for the position?
- Analytical Methods
- Cartography and Visualization
- Data Modeling
- Geocomputation
- Conceptual Foundations
- Design Aspects
- GIS&T and Society
- Organizational and Institutional Aspects

Reminder: An element's "score" is calculated and visualized based on the selection you choose from the toggles.

- displays an individual element's BoKscore

- displays either the calculated sum or average scores of an element and its descendents in the tree structure

Figure 41: Multiple Selection Interface 1
Scenario: The same GIS professional decides that she wants to take a class to learn more about GIS&T and Society and Organizational and Institutional Aspects in order to better qualify for this position. She found two courses that seemed to be a good fit. Both courses have an asserted scorecard, indicating the degree to which each covers particular Topics from the GIS&T BoK. Select the scorecards for Course 1 and Course 2 from the dropdown lists in the interface. Note that the courses were scored at the Topic level.

Use these results to answer the question below:

According to the visualization, which Course would be most helpful to the applicant and why?

Which knowledge area(s) does Course 2 seem to focus on more than Course 1?

- Analytical Methods
- Cartography and Visualization
- Data Modeling
- Geocomputation
- Conceptual Foundations
- Design Aspects
- Data Manipulation
- Geospatial Data
- GIS&T and Society
- Organizational and Institutional Aspects

Figure 42: Multiple Selection/Open Ended Answer Interface for real-world scenario 2
BoKVis Survey: Feedback
Thank you for participating! Please take a few moments to offer some feedback about your experience.

On a scale of 1-5, how easy was this tool to use? (1 = not easy at all & 5 = easy)

- [ ] 1
- [ ] 2
- [ ] 3
- [ ] 4
- [ ] 5

What did you like best about this tool? What did you like the least? Please be as specific as possible.

Please feel free to share any additional comments or suggestions in the space below.

Figure 43: User Feedback Interface
Thank you so much for participating in the survey! Please feel free to contact me with any questions, comments or concerns.

Marilyn Stowell  
Geography Annex, 108  
San Diego State University  
5500 Campanile Dr.  
San Diego, CA 92182-8220  
Email: bokusertest@gmail.com

Figure 44: Survey Submission
APPENDIX B

SURVEY DOCUMENTS
San Diego State University  
Consent to Act as a Research Subject

**Examining a Knowledge Domain: Interactive Visualizations of the Geographic Information Science and Technology Body of Knowledge**

You are being asked to participate in a research study. Before you give your consent to volunteer, it is important that you read the following information and ask as many questions as necessary to be sure you understand what you will be asked to do.

**Investigators:** Marilyn Stowell, Geography Department San Diego State University. This study is being supervised by Dr. Andre Skupin, Geography Department San Diego State University.

**Purpose of the Study:** This study will compare four interactive visualizations of the Geographic Information Science and Technology Body of Knowledge, to determine which spatial display is most effective for acquiring knowledge about a knowledge domain. To participate, you must be at least 18 years of age.

**Description of the Study:** If you choose to participate in the study, you will be asked to answer survey questions about visualization, and complete a few tasks related to the visualization’s content and structure. Answers will be submitted through the computerized survey. The entire process should take no longer than 45 minutes.

**Risks or Discomforts:** If there are any questions that you do not feel comfortable answering, you have the option to skip those questions. You also have the option to stop your participation in the study at any time.

**Benefits of the Study:** All participants will have the opportunity to explore the GIS&T Body of Knowledge and gain insight into knowledge domain visualization and the structure and content of GIS&T. You additionally may receive extra credit from your Professor (determined by your Professor).

**Confidentiality:** Confidentiality will be maintained to the extent allowed by law. No personal information will be collected. Data will be stored in a locked file that is only accessible to the investigator.

**Voluntary Nature of Participation:** Participation in this study is voluntary. Your choice of whether or not to participate will not influence present or future relations with San Diego State University. It will not influence your grade, standing in class, or relationship with your Professor. If you decide to participate, you are free to withdraw your consent and to stop your participation at any time.

**Location/Environment:** The survey will take place in the Spatial Analysis Lab in Storm Hall (SDSU), room 338 A. Participants will have access to a computer to use, and a quiet environment in which to complete the survey.

**Questions about the Study:** If you have any questions about the research now, please ask. If you have questions later about the research, you may contact Marilyn Stowell (Storm Hall office 338A) at 253-225-6039 or through e-mail at stowell@rohan.sdsu.edu. If you have any questions about your rights as a participant in this study, you may contact the Division of Research Administration San Diego State University (telephone: 619.594.6622; email: irb@mail.sdsu.edu).
Survey ID: 1

Thanks for participating in this study!

Published in 2006, the Geographic Information Science and Technology Body of Knowledge (GIS&T BoK) has become one of the most influential documents in the area of GIS&T education, both in the United States and internationally. It was created by the University Consortium for Geographic Information Science (UCGIS) as a key component of their Model Curriculum Project, which aimed to better prepare students for GIS related professions. The BoK has been primarily distributed as a printed document, created with contributions from professionals and scholars from over 80 institutions. It consists of a four-level hierarchy of 10 knowledge areas, 73 units, 329 topics, and over 1600 learning objectives, designed to support the development and assessment of GIS&T curricula (Prager & Plewe, 2009).

GIS&T Body of Knowledge

10 Knowledge Areas

73 Units

329 Topics

1600+ Learning Objectives

The interactive visualization (BoKVis) lays out the content of this document in its entirety and allows you to interact with the BoK to answer questions related to the GIS&T Body of Knowledge.

Accessing the BoKVis Application:

Below is the URL to the Dropbox that will grant you access to the Body of Knowledge Visualization:

https://www.dropbox.com/sh/o475aa4jlfz1fh5/nDvZlFnExP

In this dropbox you will see that you have two options with which to access the application. Select the folder for either the Mac (application.macosx(1).zip) or the Windows Application (application.windows(1).zip), depending on the machine you are using.

• If you select the Windows Application: Click the folder titled application.windows(1).zip, and choose to download the .zip file to your local computer. Once the file has been downloaded, right click and extract the files to the destination of your choice. Open the folder where you chose to save the files and open the folder called “application.windows”. Once inside this folder, double click the Application called BoKVis to open the BoK Visualization.

• If you select the Mac Application: Click the folder titled application.macosx(1).zip, and choose to download the .zip file to your local computer. Once the file has been downloaded, unarchive the files to the destination of your choice. Open the folder where you chose to save the files and open the folder called “application.macosx”. Once inside this folder, double click the Application called BoKVis to open the BoK Visualization.

Accessing the User Test:

To take this survey you must have an internet connection, and Flash enabled within your browser. Should you encounter any accessibility issues due to these compatibility restrictions, please email me at stowell@rohan.sdsu.edu. Access the user test using the URL below:

http://geography.sdsu.edu/Research/Projects/Stowell/UT_724.html

Then you may begin! Follow the instructions on the user survey from this point forward.
APPENDIX C

QUALITATIVE FEEDBACK
**Indented List:**

- the drop down menu was easy to use and show the difference and precise information per category
- It is good that you get to interact with it, and having two screens really helps. However, it was confusing and I did not really get what was going on
- The material seems to be well organized which is great. I don't like some of the aspects in regards to control. Firstly, the user is able to expand an area or unit to see the subcategories, but then cannot un-expand it. Also, the scroll bar is very touchy. If the mouse goes outside of the scroll bar it won't move. I enjoy the overlay using the queries the most. It seems to be the most practical application of the software
- Best: Being able to quickly see areas where most matches and least matches; Least: Scorecard overlay was difficult to understand how to use Individual/Cumulative and Sum/Average; also could not easily find out how to get to Course
- It is very well organized in the various areas, units, topics, and objectives that are connected to one and another. I do not see anything that I least like, but with more practice
- I like how it can search for key words throughout all layers of the text
- Easy interface, going through material was very self-explanatory. Ridiculous amount of information in the program. Explanation of different shades of colors would have been helpful. I didn't get to choose other types of visualization methods. The indented list was a bit boring to look at, perhaps other visual options were more appealing to the eyes. I
- i didn’t like that if a word was in all caps that it wouldn’t search correctly, i felt like the tool did everything it was supposed to.
- I like that how it base on my query search, it easily tell me where to look. In addition it have some descriptions that tells me about the role of a particular position
- The tool is very easy to use. Make the print larger. The topics should have some type of description after them
- I liked the vocabulary recognition, so instead of having to read through every category, the computer does it for you by identifying key words. And the information was presented well and organized with the different branches. And it was easy to close them and open them without getting confused on what category you were studying.
- Navigating was a bit confusing and some of my buttons were very sensitive which made my navigation through the site not as simplistic.
- Very confusing to find the courses and such. No key for what red and blue mean
- It was useful for telling what a person was capable of doing for a job. Without the second computer I wouldn't have been able to understand any of what was on the first computer. However I don’t have the best experience with geography. I also didn't like how the modules wouldn't close after you open them
I like the color coding, but maybe be more clear about what the colors mean. I didn’t really know what over and under qualified meant, so i had to guess.

I didn’t understand where to copy and paste the resume/job ad info.

I liked best the power of using queries in the tool. The indented tree list is nice but it becomes cumbersome after many trees are opened. It is against my intuition to use shift+click to close a tree list. I would prefer to simply click the upper branch again to close the list (open and close should work the same). The use of the query system is a bit hard to follow at first and I am not sure that the popup tips explain things in the most easy to understand way, such as INDIV vs CUMUL. I think the text is a bit hard to read, especially with the color over it. I would play around with different typefaces or a more subtle colorization. I do like the pixelated font in the menu items but I wish it was monospace!

I have no basic understanding whatsoever.

The color coding made it easy to scale the relevance.

I liked the level of interlay and cross-examination it provided, which is useful to high level corporations and experienced student-placement type programs. My first time using an interface as such was difficult, however.

I liked the way your searches are color coded for easy determination and comparison.

This tool is very visual. However, i did like the idea of ctrl-shift to close the tab. it would be easier to go back on the tab to close.

I like the idea of it. The way it is set up I think would appeal more to a geographer and GIS student. It can be a little confusing at first if you don't know how to use the toggles. A simpler back button too I think would be a good idea.

"Like the Best: The function of this tool seems pretty amazing for identifying subjects of interest, and I liked the way terms were organized into many groups and subgroups."

Like the Least: After pressing the plus button to open up a new group of terms, I was unable to collapse the opened terms which made it very cluttered and I had to scroll a lot. "

it makes no sense. i have no idea what i have to do.

It might need a little bit translate based on that I didn’t know some of the information in the category in a way i mean put in layman term.

I was extremely tired, and I have had a very long day, so that may be a factor that affected my performance.

"Shift-click to collapse is non-intuitive; click should toggle visibility."

Difficult to quickly navigate through the expanded objective list; Text size too small; had to cut/paste into WordPad so I could read the Resume and Ad.

Instead of having to press shift+click to close an area, maybe (if possible) change it to right-clicking the mouse only.

The in between color of the queries is somewhat confusing to me because is hard to tell which weight more than the other. You can extend the bars to get more detail.
within a area, but I can't click the area again to hide the extend. Until I got little more familiar with the program, I didn't know how to hide the text and I got lost within a sea of text. The scroll bar is difficult to use, if the mouse got off-center from the bar, I can't move it

- I don't understand what the terms even mean to apply it to example
- At times it was confusing but after a little bit of practice I was able to catch on
- Loved the color interlay theme, it provided contrast and easier identification.
- maybe have broader categories for your home screen and then go from there
- After pressing the plus button to open up a new group of terms, I was unable to collapse the opened terms which made it very cluttered and I had to scroll a lot.

**TreeGraph:**

- "The webs created from the areas to the objectives were easy to use and the colors of the displays were well used. The color ramp was also very helpful. It would be nice to be able to remove lists of subsections by clicking the parent sections (e.g. clicking geocomputation a second time should remove units listed under Geocomputation from the screen.) Also, the instructions for choosing an ideal course were unclear. I was unsure about what criteria to base my decision off of. I did not understand what the course settings did"
- The information domains and sub-domains are well organized and easy to access. I least liked the query function because it was hard to use. The drawdown menus were hard to use, and a table of contents would have been more helpful in choosing texts to compare
- The Text Query function was straight forward and easy to use. The crowding and organization of data made it difficult to manipulate
- I liked the query tool the most. It was easy to understand and the color scheme was intuitive. The conceptual map also offers a helpful visualization. Working the query was somewhat challenging as someone not used to this type of technology
- I liked the menus on the left of the screen and the way that each item mapped to other multiple items
- "Would be very helpful in finding what skills one has and which skills one needs to work on for a particular job. Parts of the GUI are confusing at first but fairly easy to figure out with a little time. After expanding and exiting a couple levels lots of scrolling was needed as the "Areas" were not centered any more. Really interesting to see what objectives and topics are under each area of study"
- I'm really bad at technology, absolutely terrible, so I liked the color coating.
- I like how powerful the query function is and how they grade them accordingly. I do not dislike anything about it because it seems to still be a prototype and it seem to work properly.
- "I like the concept of searching by text to narrow focus of relevant topics to the example of a job search/qualification situation."
Some of the terminology is confusing such as using the term map to refer to a color scheme. Map is appropriate in the sense that it gives an indicated direction, but confusing conceptually. The box style interface is not immediately intuitive as to the function of each and again terminology could make it more direct and clear as to the function and intended use of the tool. The product should be a great tool since the body of knowledge and the applications of GIS are extremely varied and educational requirements also very divergent.

- Search queries simplified the process of finding the required data but I did not like the UI as it seemed to complicate thing
- The program was a bit buggy in responding to scrolling commands. The boxes should be easier to see in terms of space, and there should be a way to close each section so that it makes it easier to open other sections that need to be opened
- I like the color coding that corresponds to the query fields and the fact that you can compare two searches at a time
- I enjoyed how clearly the colors visualized comparisons between queries. It really helped show which query was related to which topics using the sliding scale, and coding the shades to each subsequent unit or topic made it easy to figure out where further research was necessary. The lines linking the tree graph were confusing to follow at first, but once I got used to the highlighting and zooming system I was able to clearly connect the fields I was looking at.
- "I liked that it was able to color coordinate the desired answer using the text query system with two distinct contrasting colors: Red and Blue.

The collapsing method of boxes can be a bit difficult to manage as I am use to pressing the "+" sign again to close an expanded box"

- I disliked the small text query area and the lack of description of what everything did. It was fairly easy to figure out, though
- Like the GIS and technology aspect
- I was really confused the whole time, there wasn't much that I actually understood. I wish it was a little bit more simple
- The drop downs were easy to use. It was a little too much information in each drop down to understand exactly what you wanted
- The data tree was easiest part to use. The text query and scorecard overlay confused me. I understood that it was matching the data I pasted into the text boxes, but I wasn't sure how it was making those matches or how the two texts I pasted in related to each other
- I probably just missed this earlier in the instructions, but it was not obvious to me how to get rid of unit and topic levels I no longer wanted to view. More instructions on the text query would have been helpful.
- Overall I liked the simplicity of the interface. I just needed more direction on how to navigate through it.
- Lots of information
• It was easy to use, the Colorization of red and blue way very helpful, the shift click was annoying and the "copy and paste" wasn’t "right clickable"
• I found the tool confusing. Each item maps other items, but it is difficult to recognize which parent element each "child" element maps too. Also, it wasn't clear to me how the course drop down boxes acted as filters on the data that was presented. There were two drop down menus but I wasn't sure if both menus were simultaneously filtering the data or if one had priority over the other.
• once I started clicking on things, more and more popped up and I couldn’t figure out how to make them go away, so it got a little disorganized for me
• I am very impressed with this tool. I have never seen another one similar and I like the design.
• The interface was quick to learn for someone with plenty of experience with computers, but I feel that some older folks might have trouble with the visualizations. The plus/minus buttons are particularly small
• It was very informative in how each areas in the BoK1 would show a breakdown of the topics and detailed information below as each expansion box is opened.
• I know I did not read the directions very careful which is my fault, but I still wish it could be a little bit more simple
• information is difficult to navigate through

TreeMap:
• I like the fact that it has a clean framework to select smaller sub-sections.
• I liked that it broke down sections into a deductive way of thinking about concepts
• Pretty cool design and it was easy to locate everything
• "1. I did like the query scenario with copy and paste resume and job ad and compare the two with the knowledge base.
2. I like the least is the clutter of all the information within the knowledge base from the unit down to the objective, maybe include a status bar"
• It's easy to visualize the tree structure. It needs to have more native computer functions, e.g. Undo or "Back" to go back to a previous set of data and display. The toolbar GUI could use usability improvements, the buttons, menus, etc. are far to small and are unintuitive
• It is a little foreign to new people so some things may seem hidden like where the drop down menu even was or where course 2 was within the drop down menu. It took a lot of digging say to find analytical boxes but once I cognitively got what it was I was looking for and digested what each term meant I could easily find what I was looking for. I wasn't exactly sure on how the key worked color wise but it made perfect sense once I thought about it for a little while. The tool was effective and it just takes some time figuring things out which isn't a bad thing.
• A little confusing at first. It seems that it would take some time to become familiar with the tool and use it to its full potential. I was confused on the course 1 and course 2 comparisons. While i understood the hot and cold areas, i wasn’t sure exactly what
I was being shown with the comparisons. I think the strong point is that the tool has so much information and data to share. But this positive can also be a negative because it can appear overwhelming to the user. That is why time and instruction may be required. But I can see how this can be a great tool for quick information reference and to better understand your own skill set.

- "I liked the hierarchical design and simplistic display for seeing the hierarchy. I was not sure how to view higher levels while I delved into the more detailed, lower levels. For instance, it would be nice to have a reference of what knowledge area you are in while you look inside the very specific learning objectives. That said, I could have missed where this information is displayed, so maybe make it more prominent."

- I liked best how it correlates the specific and general. I liked least the drop down menu and its link with the color coding system.

- "I like the function of finding the most appropriate job at section2. It's really useful when the company has many requests for their candidate. However, sometimes I don't know the meanings of each small blocks and I cannot understand the practical function of the dropdown list."

- I felt like some of the text boxes were a little small, making it a little hard to navigate. I am also color blind so not good with colors, as indicated earlier in the survey.

- It is kind of confusing in the beginning, but you get used to it as you use it, although you have so many options it takes you too much time to read them all.

- I liked how interactive the tool was. It was fairly easy to use as well. The thing I liked least was the time it took to get back to the "home" screen. Each time I wanted to exit out of an individual section I had to push the shift key multiple times. Having a button that would take you directly to the home screen would be much easier. Also, the title of the section covers some of the titles inside the section and it makes it hard to see. For example, if I was looking in the Geospatial Data section, and my mouse was in the box, it is very difficult to read the subsection that says "Earth's Geometry."

- I like the design of having an original broad topic and being able to go deeper and deeper into the subject you want knowledge on. However, I found this did get a little overwhelming in the sense that it would keep dragging you into more and more boxes. I did also like the short cut ctrl+shit and click to get out of the box and back to the main menu.

- I like how it is divided and each Area that you click on "breaks down" into further sub-units, then each unit into topics and topics into objectives. I didn't like how there was no back button, so I would have to click on AREA to go to the beginning. Also, the last question asked me to choose a COURSE 2 from the dropdown scorecard, however both Query 1 & 2 did not have a COURSE 2 option...

- Easy to manipulate and get around. Some of the open ended questions were confusing.

- Tons of information can be found with relatively minimal searching.

- I like how there is a search function in order to expedite access to the different areas and their subcategories. I also like how things are color coded and labeled.
• this tool was a little hard to understand how to use
• THERE IS A LOT OF INFORMATION; HOWEVER, I WAS UNSURE WHAT WAS GOING ON, WHY I WAS DOING IT, AND WHAT THE OUTCOME WAS SUPPOSED TO LOOK LIKE
• It is a little confusing at first
• I like how all the information is divided up but altogether as well. It is easy to move back and forth from one topic to another
• the navigation and visualization methods were both great, but for a beginner who has no knowledge on the subject, it can be slightly confusing. possibly more directional tools would be useful,
• I think this would be fairly easy to use for people, because it looks professional and clean. This could not only just be applied to Geography related topics
• It was very hard to navigate the program without proper training on the subject.
• maybe some different fonts or styles to make it even easier to locate thing
• One suggestion to help alleviate the clutter of all the information within the knowledge base in the treemap would be creating a status bar in the tool bar indicating how far within the tree graph you are in. For example, in windows explorer, if you were in the java folder, you'll notice the status bar computer > local disk > program files > common files > java
• Great work. I look forward to using this tool in the future.
• A quick demonstration before starting would be helpful.
• It is great :)
• It was a little annoying that, when selected, the knowledge are title was in the way of viewing the subtitles with in the box.
• I liked the organization of this tool, it was easy to use once I figured out what each button did. If an brief overview of how to navigate the BokVis would have been provided, this would have been much easier.
• The resume and course section was a little difficult to figure out. The rest was pretty straight forward.
• It would be great if the software contained brief prompts that explain what the different components of the software mean. For instance, with Microsoft Word or InDesign, when you roll your mouse over an icon, a brief message appears to explain the function of the command.
• THE INTRODUCTION AND EXPLANATION OF WHAT IS GOING ON NEED TO BE MORE THOROUGH
• need more time to understand it