

Medical Applications of Visualization

An Interactive Qualifying Project
submitted to the faculty of Worcester Polytechnic Institute
in partial fulfillment of the requirements for
the Degree of Bachelor of Science

Date: May 3, 2006



Christopher Gianfrancesco



Matt Jarmak

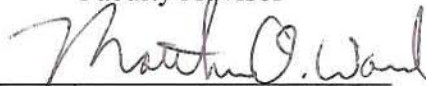


Steve Kolk



Owen Pedrotti

Faculty Advisor



Professor Matthew Ward

Table of Contents

Table of Contents	1
Ownership	2
Preface	3
Abstract	3
Ch. 1: Introduction	4
Ch. 2: Choosing a Field	
2.1 First Pickings	5
2.2 Narrowing It Down	6
2.3 Final Decision	11
Ch. 3: Research & Contacts	
3.1 Searching for Data Sets	13
3.2 Forming a Contact List	14
3.3 Attempts & Results of First Contacts	15
3.4 Further Contact Attempts	16
Ch. 4: Meetings and Observations	
4.1 Prepared Plan for Meetingss	
4.1.1 Xmdv Tutorial	18
4.1.2 User Response Survey	19
4.1.3 Data Conversion	23
4.2 Meetings with MCPHS	25
4.3 Meetings at Umass	30
4.4 Further Meetings	34
Ch. 5: Analysis and Conclusions	
5.1 Evaluating the Tool (results from user input)	36
5.2: Evaluating Ourselves.	39
5.3: Further Works.	40
Appendix A: All About XmdvTool	42
Appendix B: Boilerplate Emails.	46
Appendix C: Outline of Training Program	48
Appendix D: Transcripts of Notes from Meetings.	49
Appendix E: User Survey.	52

Ownership

Christopher Gianfrancesco

- Helped establish contacts by phone.
- Researched and wrote survey
- Attended and took notes at MCPHS meetings.

Matt Jarmak

- Organized and sent mass e-mails.
- Served as contact to answer in-between meeting questions.
- Researched paper format and did majority of write up.

Steve Kolk

- Gathered list of possible phone contacts.
- Primary source of establishing phone contacts.
- Attended and took notes at UMass and Fallon Clinic meetings.

Owen Pedrotti

- Developed expertise with XMDV Tool.
- Prepared and presented for meetings.
- Converted data sets.

Preface

This is our report on our research into the medical applications of visualization. Herein contained are our thought processes at each step we took, as well as some steps that we tripped and fell down on. However, we got up, dusted ourselves off, and continued onward along a different road.

As most Interactive Qualifying Projects do, it seemed a daunting task at times. But we persevered, and achieved our goals. Confidence in ourselves and an unwillingness to give up brought us to this juncture where we now stand, and brought this report to you.

Abstract

Our goal was to evaluate the usefulness of visualization, and more specifically XmdvTool, in the data processing currently being performed in the medical field. We did this by meeting with medical professionals and discussing the data they work with every day, and how the tool could be used to make sense of this work. More specifically, we looked to meet with three kinds of medical professionals; clinicians, educators, and researchers. We then converted some of that data so that they could see for themselves how it looked. By instructing them and answering their questions about the tool, we were able to gain an understanding of their needs, and how they might be able to make use of the tool in their profession. Of course, it is necessary to receive feedback on all of this, and we planned to use surveys to gather this feedback. In this way we could evaluate reactions to XmdvTool.

Chapter 1: Introduction

There are many tools that can be used to help understand a data set, and many ways to think about data. Visualization is an incredibly powerful method of thinking, and we set ourselves the task of evaluating its usefulness in the medical field. The method to evaluate we devised was to show data sets in a particular visualization tool to doctors, medical professors, and medical researchers and collect their feedback. The particular tool we used was XMDVTool, a tool originated in 1993 by Professor Matthew Ward. It has been continually upgraded by him, various students at Worcester Polytechnic Institute, and Professor Elke Rundensteiner since then. We felt that it was an excellent option with which to show people the power of visualization, as well as a good test for the tool itself.

In order to evaluate the tool and the power of visualization in medical applications, we had to devise contact methods, as well as ways to get responses from these contacts. We chose to work with meetings, direct feedback, and surveys in order to get a varied response set. In our tasks, we learned quite a lot about how to manage our time and about how to split a large project into manageable portions. We learned how to design surveys that will give us meaningful results. We learned how to use XMDVTool, of course, and we learned how better to communicate group objectives to each other.

Without further ado, we give you a summary of our experiences, as well as our results. We hope that you enjoy reading it as much as we did performing it, and that you may learn from both our successes and our mistakes.

Chapter 2: Choosing a Field

2.1. First Pickings

The few days between our meetings during the second week of B term were spent brainstorming ideas of what areas would both interest us in terms of applying the visualization tool and have users who would take an interest in spending time with us to try out the program and give us the feedback we need. Also, the particular application should preferably have some sort of societal benefit; this idea was actually a good springboard for us to start narrowing things down.

We started by thinking of the usual organizations one thinks of when considering 'societal benefit' - schools, hospitals or clinics, public works, and general social services. Of those, the applications of visualization seemed to be most prominent in the areas of education and health services, and they were also of greatest interest to us, so we decided to add those two to the list of areas we would potentially be able to work in.

One of the aspects of visualization that we found interesting was its ability to help users detect patterns in data gathered over periods of time, so that led us to the consideration of meteorology as an area of study - researchers in the field generally have a need to find patterns based on raw data, and also the field has societal benefit in that prediction can be applied to things such as disaster prevention.

Another topic that we considered was looking at the patterns related to the

behavioral effects of media on people; i.e., do tv, movies, and video games make people more violent, less violent, or do they not affect them? We could look at patterns such as the amount of time spent with media, and the specific times of the activity as well (whether you spend time with media in the afternoon or at midnight, for example). It might help predict or isolate violent behavior, which of course benefits society.

At our meeting at the end of that week, we presented our ideas, and decided that the next thing we would do would be to think of some pros and cons of each potential application area.

2.2 Narrowing it Down

The major thing we did during the third week of B term was discuss the pros and cons of the topics we had considered. The first one we looked at was meteorology.

Pros:

- Meteorology has a very large data pool to draw from. There are meteorological events all over the world and have been being recorded for hundreds of years. This would give us easy access to extremely significant amounts of data that we could analyze.
- Meteorological data is easy to quantize and compare. It is always of a very specific and rigid form, and it always has the same set of potential variables. As a result, it is easy to work with it.

- Analysis has a significant societal benefit. Thousands die in an average year due to natural disasters; in cases of particularly destructive meteorological events, even more can die. Prediction of such disasters could save a high percentage of those needless deaths.

Cons:

- Meteorology tends to have a lack of reachable analysts. There is not an available pool of meteorologists in or near the Worcester area. Also, most meteorological analysts are well-entrenched in their position and may be unwilling to look at a new technique from someone with few to no references from their colleagues.

The second application we looked at was education.

Pros:

- Large amount of available data from nearby sources. We are students at a university, and there are many more universities in the immediate area. Each of these universities has thousands of students, and data pertaining to each student.
- Many nearby end users to look at the data. Again, we are enrolled at a university, with many more close by. Admissions people would more than likely be willing to see a new way to help them sift through their data.
- Educational statistics pertain to us personally. It is a field that interests

us due to the direct and unsubtle impact it can have on us.

Cons:

- Data is not easy to quantize or gather accurately, as surveys are unreliable. Students themselves are unlikely to flock to us and give us their relevant information.
- There may be privacy issues with the data. Admissions data may not be available to us as students, due to privacy concerns. Any data we could get may be too generic to be of use.

The third application we looked at was medicine.

Pros:

- Fairly easy quantification of data. Medical data is usually well-formatted, neat, and regulated. As a result, we decided that it would not be terribly difficult to be able to look at it alongside apparently unrelated data.
- Many potential local end users to look at the data. There are many pharmacies, medical schools, and personal physicians in the immediate area. That gives us many people to approach with our tool.
- Useful information, not just beneficial to society, but also to people on an individual scale. Tracking the spread of diseases is a valuable thing to a society, as is tracking obesity over the years. Individual people may themselves be interested in statistics such as the likelihood of

heart disease based on their lifestyle.

Cons:

- There are potential privacy issues with the data collection. Medical data is not something that is often freely released to the public, even in an anonymized form.

The fourth application we looked at was the effects of media on behavioral states.

Pros:

- Data is everywhere and very easy to find. People interact with media all the time, perhaps even too much. As a result, we would be able to simply walk down the street, point at anyone, and they would be a potential source of data.

Cons:

- Fairly boring. It is not an incredibly interesting topic to us.
- Has been done many times before. This is a topic that has been studied back and forth with differing conclusions literally dozens of times over decades, since media became more interactive than books.
- Difficult to establish a causal relationship either way. It would require a large amount of data gathering to show how people's behavior was affected, especially since it would be near-impossible to find a person who had never interacted with media of any kind, and thus it would be complicated to find a control.

At our weekly meeting, we presented the pros and cons we had developed, along with further exposition as to the reasoning behind them (included in the bullet points).

Along with input from Professor Ward, we decided to narrow ourselves down to only three areas; meteorology, education, and medicine. For the next week, we decided that we would narrow ourselves down to only two areas of interest, and start brainstorming some specific end users within those areas.

Over the fourth week of B term, we brainstormed ideas for end users for meteorology, education, and medicine. After butting our heads against the virtual brick wall of attempting to think of end users for meteorology, we removed it from the list, thus achieving our goal of narrowing ourselves down to two topics. We then continued with our brainstorming.

The first idea that we came up with for medicine was a clinic; using the visualization tool, doctors could better evaluate and compare patient's health data to find patterns in their health. They could also show patients individually a visual representation of their health as compared to other patients. The second idea we came up with was hospitals. They could use the visualization to find patterns for diseases and epidemics; that could help save many lives. They could also track trends over time, watching such things as the average rate of weight gain over time. The third idea that we came up with was watching trends in personal health. The end user for this is a single person, so it is extremely easy to find. It can be used for dieting; keeping track of different meals and exercise

routines to evaluate lost weight. It can also be used for bodybuilding; the same as dieting, but in reverse. Weight loss groups could be a potential source of dozens of end users. Finally, it could be used by researchers, as they present their data to the target audience of their research.

Education had four end users that came to mind. It could be used for admissions; they could plot standardized test scores/class rank vs grades. That way they could decide how much they want to value those variables. It could also be used to plot pre-college vs post-college grades, to see what the relation is (if any). Another source of end users would be teachers; they could plot different variables against the average grade and see how they stack up. Some variables they might use include hours of homework, hours of lab, hours of class, and hours of conferences. A third group of end users would be student advisors; they could track such things as hours of sleep, studying, clubs, and leisure, and their effects on grades. A fourth group of end users would be the students themselves. They might be interested in seeing their own rank against other students, or how they improve (or perhaps decline) over time, or how their sleep and study habits have changed over time.

2.3 Final Decision

At our weekly meeting, we discussed our ideas, and decided that medicine would be the best choice for a topic. We had looked at all of the facts and thoughts outlined above; medicine came out as the winner for several reasons.

Firstly in our opinion, medicine has a much greater societal benefit than

education. While a greater understanding of educational data is an excellent thing to have, it is unlikely to save lives in the way that a greater understanding of medical data can. Secondly, we found that medical data was more likely to be available to us than educational data. Medical data can, in our opinion, be anonymized more easily than educational data while still retaining all relevant variables. Finally, we concluded that it would be easier to find end users for medicine than it would be for education. Both topics have official channels that we could go through; universities and clinics, researchers and admissions; but if that failed, for medicine, we could talk to weight loss groups and other private study organizations.

We then discussed how we would start thinking about finding the end users we listed and what we would do once we found them. We also determined that during the following week, we would find some sources of data to use with the program to present to potential users, and to begin learning more about surveys and user studies so that we could draw meaningful conclusions.

Chapter 3: Research and Contacts

3.1: Searching for Data Sets

We needed to find at least one data set with which we could draw potential users in. If we could show them something already related to their field, without any effort on their part, it would be a great help towards convincing them of the usefulness of XMDVTool. In addition to that, it would be a strong selling point when attempting to contact users that we have a demonstration fully prepared and ready to show them. It is an even stronger selling point to tell them that we have a data set highly relevant to their field ready to show them whenever they would like to meet with us.

Our initial thoughts of traditional research were dashed when we realized that even should we find something in a journal, encyclopedia, or other printed source, it would take a disproportionate amount of work to add it into a format that XMDVTool could read; that is, onto a computer. Thus we knew that we would have to find it on the Internet; a seven-week term does not leave much room for unnecessary work.

Fortunately, the Internet came to our rescue. After hours of sifting through results of search engines, we had found two large data sets.

http://seer.cancer.gov/csr/1975_2002/sections.html , a cancer statistics review from 1975 to 2002; and <http://www.cdc.gov/cancer/npcr/uscs/index.htm> , containing statistics on the United States incidence and mortality of cancer. These two finds gave us all the data we would need to present prior to meeting

with users, and then we could work with their data for them.

3.2: Forming a Contact List

The challenge facing us now was to find people in the area who might be interested. We examined phone books, online phone books, maps, and search engines, and came up with three initial groups of people who might be interested.

The first such group we found was the nearby UMass Medical Center faculty and students. It is the medical branch of the University of Massachusetts. We thought that specifically the Infectious Diseases branch might be interested in this, since the data sets we had found were related to cancer. While cancer is not an infectious disease per se, it follows many patterns that are similar to certain diseases. We found their contact information online at their website; the contact information for general faculty is located at

<http://www.umassmed.edu/generalmed/faculty/>. The contact information for the infectious diseases faculty is located at <http://www.umassmed.edu/infdis/faculty/>.

The second such group we found was the Massachusetts College of Pharmacy and Health Sciences. The Massachusetts College of Pharmacy and Health Sciences (MCPHS for short) is a private collegiate institution that offers degrees in both graduate and undergraduate medical fields. They also perform medical research. MCPHS has branches in Worcester, Boston, and Manchester; we chose to contact only Worcester faculty, due to their proximity. The contact information for MCPHS can be found at

<http://www.mcphs.edu/contacts/index.html>.

The third such group we found was the Worcester Medical Center. The physicians working there pride themselves on their usage of advanced tools for diagnosis and treatment, and so we hoped they would find XMDVTool useful. The Worcester Medical Center is in general a highly advanced place for both inpatient and outpatient care to take place. The contact information for their physicians can be found at <http://www.worcestermedcenter.com/physdir.html>.

3.3: Attempts and Results of First Contacts

The next step ahead of us was simple in theory and complex in execution; to contact people who may be interested out of the three groups that we found. A boilerplate email to be sent was devised for each group (copies of these boilerplates can be found in Appendix B). The boilerplates were revised with the input of Professor Ward, and then they were ready to be sent out. Each boilerplate was slightly customized based on the information we could find on each recipient, with simple things such as their name and differing fields of interest to them.

Of our initial “batch” of emails sent out, we received one reply within a week. Our respondent told us that he was not interested, but that he knew someone who might be; unfortunately for us, that other person was another one of our recipients, and he had not answered anything. We did not receive a single contact from our first “batch” of emails; it was clear that we needed to continue attempts.

3.4: Further Contact Attempts

The first thing we tried after the emails had failed was to simply call everyone who we had sent an email to. While we got many voicemails or secretaries, the phone was a much more effective tool than the virtual pen; one person from the Massachusetts College of Pharmacy and Health Sciences said that she would be very interested, and that she would arrange for us to present at a staff meeting. We also found another respondent, who spoke for the UMass Medical Center Infection Control Team, and said that they would also be interested. In essence, we had hit pay dirt.

However, those two groups were not enough. So we went to find more. Our initial tactic was again email, sending emails to a list of cardiologists at the UMass Medical Center. As we learned, emails did not rapidly get a response, and unfortunately, we found no one interested with follow-up phone calls there.

On the advice of Professor Ward, we sought out Professor Carolina Ruiz, to see if she knew anyone who might be interested. After meeting with her, and explaining what the tool was and what our project was, she said that she might know some people interested, and that she would give them our contact information. Unfortunately, we never received anything from any of them, so we assume they were not interested.

Finally, we had two more potential contacts to try, one had been suggested to Professor Ward by Doctor Peter Lucas, and the other to him by Doctor John Hosey. After struggling with email daemons, we realized that we were running out of time and skipped that step. Fortunately, the first of those, Doctor Garber, a

physician at Fallon Clinic, was interested, and a meeting was arranged with him.

We had our three contacts.

Chapter 4: Meetings and Observations

4.1.1: XMDV Tutorial

In order to ensure that the professionals we met understood the potential uses of the tool, we needed to give them a broad overview of its primary functions. The first step to doing this, of course, was to learn it ourselves. Some time was used simply interacting with the program, using various data sets and looking for trends and correlations. It was important that we had enough knowledge about the tool to be able to identify its most important functions, those that are likely to be the most commonly used. We also had to be able to answer any questions that might be asked during our meetings. As we are not medical professionals ourselves, we were not able to fully predict what aspects of the tool they might find important. Because of this, it was most important that our understanding of the tool was as broad as possible, rather than focusing on particular details of its functions. With this in mind, a general outline of the tutorial session was laid out.

Although the tool clearly is best suited for complex data sets, we felt that a simple example was necessary for giving people a good understanding of how the tool represents data. The default sample of iris flower measurements proved adequate for this purpose. The resulting uncluttered views allowed us to give examples of interaction that were easy to see and understand. Early in the project, we had found a few medical data sets with the intention of using them in a

tutorial. However, we decided that it would be better to make the demonstration as simple as possible. We also considered that no matter what medical data we chose for the demonstration, it is unlikely to be a good representation of the types of data used by the people we are demonstrating to. In the worst case, they might not understand the data being shown (which we would be even less capable of explaining), making the learning process more difficult. And so, we chose to use a sample data set that anyone would be able to quickly understand for the tutorial.

The tutorial was composed of a explanation of each of the four main views in the toolkit, with each building upon the last. It was emphasized that manipulating the data in one view would also influence each of the other views. By allowing them to see how the different views of the same data related to one another, they would better understand how each view worked individually and as a whole. At each stage, the various ways to manipulate the data would be demonstrated. While we did our best to explain possible uses for each part of the tool, it was clear that we would have to rely on their feedback to determine those uses. And so the question and answer portion of the tutorial was of key importance.

4.1.2: User Response Survey

We believed that we would be meeting with a number of users all wanting to give us their thoughts on XmdvTool, so when we were preparing our plan for the process of these meetings, we decided some standard way to gauge user

response would be necessary. A survey seemed to be the clear solution - they are relatively easy to create and produce, and they can provide a standardized format for input while still having some room for open-ended responses. To go about creating a survey, though, we needed a starting point, so we found a guide to forming surveys that was both understandable to us non-statisticians yet complete enough to ensure we had a good enough grasp of the topic. Our resource of choice was the Sampling and Surveying Handbook (found at <http://www.au.af.mil/au/awc/awcgate/edref/smpl-srv.pdf>).

The Handbook details a few key steps in the creation of a user feedback survey, including statement of objectives, forming a hypothesis, and establishing the survey plan (including data collection plan, data reduction/reformatting plan, and analysis plan). Here we present our progress through these steps leading to the creation of our survey.

Step 1: Statement of Objectives

Our aim with this project was to determine a) whether or not multivariate data visualization is applicable and useful to the field of medicine, and b) to assess the performance of XmdvTool as a multivariate data visualization application. Thus, we can state our objectives as follows:

1. To determine the utility of visualization tools in general for aiding in the work of medical researchers, educators, and practitioners, and
2. To evaluate XmdvTool as a multivariate data visualization tool, based on criteria such as power and versatility to accomplish tasks necessary for

usefully analyzing data sets and user compatibility.

Step 2: Forming a Hypothesis

The challenge in forming a hypothesis is making certain to create an appropriate and scientifically provable one that is not based on subjectivity or moral judgement. Even though ‘We think people will like XmdvTool’ is technically a hypothesis, it is not a very good one, as there is no scientific basis for which we can determine the truth of the statement. With that in consideration, our main hypotheses are as follows:

1. The majority of subjects will find data visualization techniques useful and preferable to other forms of data analysis.
2. The majority of users who undergo user studies will be able to perform tasks within specified guidelines of time and ease.

(Note: We had planned to create the guidelines after we had a stronger understanding ourselves of the learning curve of the program, but that plan was changed after we found that the structure of the meetings and prohibitively short time frames we could get from users meant that performing in-depth user studies was highly impractical. Thus we came to rely on the survey as the quick yet sufficient feedback option.)

Now that we had clearly stated our goals and our predictions/hopes for the outcome of the study, we could begin to get an idea of what form our surveys needed to take and what questions we would need to ask. The questions and answers, like the hypotheses, should be clear and definite. Open-ended questions

can sometimes give interesting and insightful answers, but when looking to prove the validity of a hypothesis, open responses can be difficult to glean meaningful value from - and much time can be wasted trying to make sense of or quantize open responses into appropriate categories.

Step 3: The Survey Plan

The survey plan - essentially, the form of the survey and how it will be carried out, start to finish - was comprised of three parts. The data collection plan includes the physical form of the survey and how it was be distributed to participants. The data reduction and reformatting plan involved the aggregation of responses and putting the results into an easy-to-work-with form. Finally, the analysis plan detailed how we would examine those results in relation to proving or disproving our original hypotheses.

Data Collection Plan

Our data collection plan for this project was relatively simple - we would be going out and meeting with our subjects, providing simple paper surveys for opinion analysis, and conducting user studies with individual subjects. This would allow us to both access a wide range of people easily as well as get a more in-depth analysis from some subjects.

Data Reformatting Plan

We felt at this point that our studies would not likely reach a large enough

audience that we would need to employ formal data reformatting techniques, such as scanning sheets and the like. If the response was larger than we predicted, however, we could have always re-evaluated this as necessary.

Data Analysis Plan

Thanks to our well-made hypotheses, we had a clear aim for a data analysis plan: to categorize responses and tabulate results to see if and how our hypotheses were proven or disproven. Thus, we had to be careful to design questions and answers that would clearly help us quantitatively analyze the validity of our predictions.

With all these in mind, the final survey was created, and can be seen in Appendix E. Things went mostly as planned when it came to the implementation of the survey, with only a couple of notable exceptions:

- As mentioned before, user studies were dropped entirely due to time constraints.
- Some users were given the survey via email rather than paper copies, mostly due to time and availability limitations.

4.1.3: Data Conversion

At the conclusion of the initial meeting, we asked each group that we met

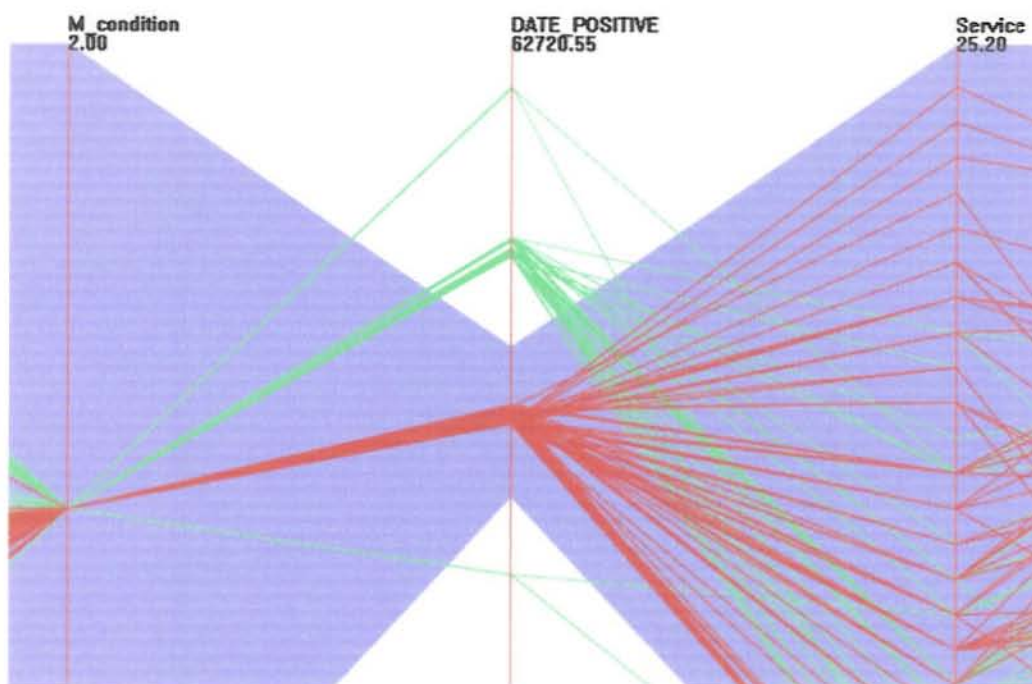
with to give us a simple set of data so that we could convert it into the OKC format used by the tool. We would then meet with them again and allow them to see and manipulate their own data in the tool. Due to the limitations of the program, it was necessary that the data be modified in order to be correctly processed.

One requirement was that every data point must have a corresponding value in every dimension. The data sets we received did have some holes, but this was easy to remedy by simply replacing all empty fields with a zero. The second requirement was that each data point could have only one value per dimension. This was an issue that we chose to avoid when it came up, as the data sets we received did not depend heavily upon it being fixed accurately. The most likely solution would be to replicate a data point any time it had multiple entries for any given dimension. The limitation that made the data most difficult to represent in the tool was the fact that it could only use numeric fields. This meant that every text field in the data sets had to be somehow enumerated. Reading the data in the toolset then required looking up this numeric code in a table.

Dates were also difficult to represent intuitively within the toolset. When we converted the data from UMass, dates were represented in YYMMDD form, so as to be chronologically consistent and easy to decipher. This did not, however, facilitate visualization, as can be seen below. The alternative would be to represent the date as the number of days since 1990. This would provide an even distribution of numbers, but would require a calculator when reading the data. Inserting the data into the toolset itself from this point was trivial, only

requiring that it be exported from Excel into CSV form and that input into a simple command line conversion program. There were, however, some compatibility problems with the conversion program and it did not run on all machines. This was easy to work around, but it may be an issue for those who work in an office environment and do not have access to different machine configurations.

Figure 4.1



Data set illustrating problem with representing date information. Brushed region represents dates in 2004. The user must zoom in very far in order to see individual months or days.

4.2: Meetings with MCPHS

Massachusetts College of Pharmacy and Health Sciences

First Meeting: February 21, 2006

For our first meeting at MCPHS, we had an audience of three - two were professors, and the third was not a teacher but did work at the college, mainly in the realm of research. For convenience, we will reference the two professors as persons A and B, while the researcher will be person C.

To start our meeting, Owen simply loaded up XmdvTool and began his planned introduction to the program, a simple walkthrough of its major features and functions using very simple sample data. In the first view presented, parallel coordinates, there seemed to be some confusion about what exactly was being represented and what it meant, but a little further explanation mostly cleared that up. Throughout the demonstration, there were very few questions raised about how certain things were being done, and it seemed that they were “getting” it and picking up what was going on fairly quickly. Once the walkthrough was concluded, comments were made from all three saying they recognized that the program could be a powerful tool and appeared that it would be user-friendly enough for them to pick up, given some time. However, while the functionality

and usability of XmdvTool was easily seen, what they were most interested in was finding out how they could apply this to their work. They did work on some research projects where they felt using the tool to help examine their data might be a good idea, but the specifics of doing so didn't come quite as easily. So, what we decided to do at that point was get some sample data (provided from a past study done by person A), do some slight reformatting to have it be usable by the program, and come back for a second meeting and hopefully shed some light on how the program can help them specifically.

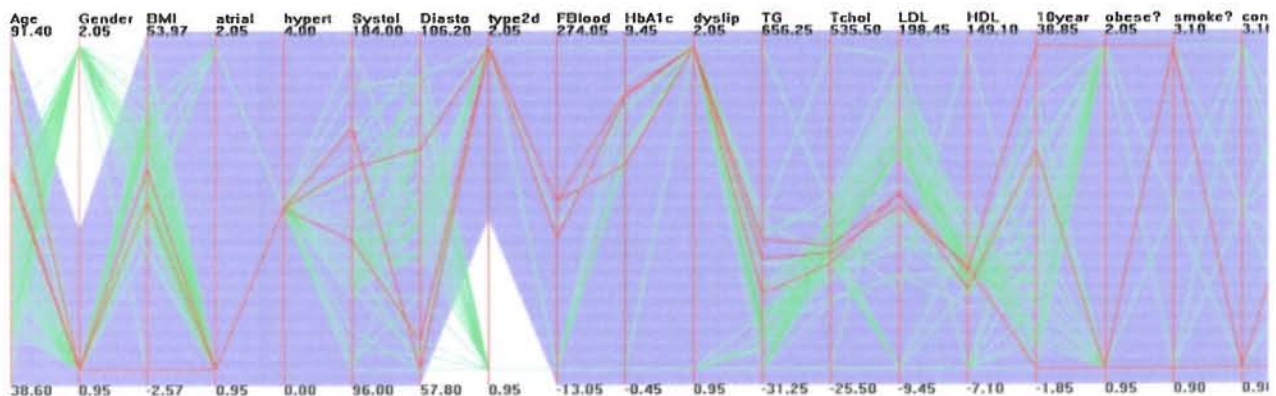
As mentioned before, not many questions were brought up during the demonstration of the tool, but there were a couple worth noting. While showing the variety of forms of visualization the tool could implement, person A asked if there was a way to get back to the plain numerical data underneath it all, in order to do further statistical analysis where needed. Owen was able to show her that the program had an option to not only show the original data, but it displayed only the data currently in selection on the visualizations, which our audience agreed seemed highly useful. Also, person C asked how this program compared to others currently in use by her and others doing research, specifically SPSS and SAS. We replied that, truthfully, we had no experience with those programs and thus could not offer a comparison, but she herself added that XmdvTool seemed to have overall more versatility than what she usually works with in terms of visualization of data.

Second Meeting: March 29, 2006

We walked into this meeting with the data provided to us the previous time, modified for use with XmdvTool and ready to be loaded up. This time, our audience consisted of person A from the previous meeting, as well as person D, a graduate student who had experience with research projects that might benefit from the tool. At the very start of the meeting, A gave a recap of what the tool was about and what we went through in the first meeting, which not only brought D up to speed on everything but showed us that she had gotten a solid understanding of the program from our last meeting and retained what she had learned.

To get things going with this meeting, Owen loaded up the new data and showed off the various elements of the tool again, with their data set (seen below in parallel coordinate view):

Figure 4.2



Visualization of patient data for stroke prevention. The brushed region represents male patients with type 2 diabetes.

Almost immediately, we could tell that person A was understanding what was being represented in the different views much more readily when she was familiar with the data that was behind the views. The parallel coordinates view, which seemed to be the most confusing in the original meeting, made a lot more sense when she could recognize that each individual line represented a single patient, and that the line crossed different axes to represent the values of different variables. Additionally, we provided a more specific brushing example by showing how one could select points based on the gender trait to view either all males or all females at once. This was more understandable than previous examples with the other sample data, and the application of brushing when it came to working with real data became more readily apparent - person A seemed to think it would be quite useful.

Overall, the tool became seen as not just a powerful program, but one that could directly benefit research projects in the future. Person A commented that on top of having a clear application for her work now, it seemed overall fairly user-friendly and she was eager to try it for herself. She also let us know that of all the different views, the parallel coordinate view was still a little tricky to understand, but she felt very comfortable with the scatterplot view. Because of this (and the fact that many researchers also have worked with scatterplots before), she commented that she could see herself using the scatterplot view not only to help conduct her research, but also to present her findings to her peers in presentations and the like.

Since the basic functionality of the program had already been seen, there

was little brought up in regards to that. Person A again wanted to see how to view the numerical data that the graphs were based on, and we showed her how it was done. Person D was curious as to how to calculate correlation coefficients and the like, to which we replied that the tool was centered more around just visualization than statistical calculation, so numbers would still have to be crunched to get something like that.

At this point, person A desired to work with the program and her data independently to make sure she understood everything well enough on her own and could more accurately evaluate its user-friendliness, as well as to discuss the results of this meeting with her colleagues (particularly B and C of the first meeting). We left her with a copy of the program and the data we had shown her, and agreed to set up a meeting after she had a chance to spend some time with it. As of the time of this writing, her schedule in the previous few weeks did not allow for her to evaluate the program or find a time for another meeting, so we will be receiving a completed copy of the survey from her and may be able to refer to her for any projects that continue in the vein of this one.

4.3: Meetings at UMASS

This section covers our meetings with the Infection Control team at UMass Medical Hospital. They are a group of researchers that track patients who get infected with a contagious disease and try and find possible causes within the hospital. It is very detrimental to a hospital's efforts if they are helping spread a

contagious disease, so this team tries to compare data of patients to try and see if different factors of their hospital visits contributed to their infection. They keep track of what rooms patients were in, their doctor, nurses that were present, and what days they came in. They then try and find patterns to determine if a problem exists with the way the hospital treats patients with different contagious diseases.

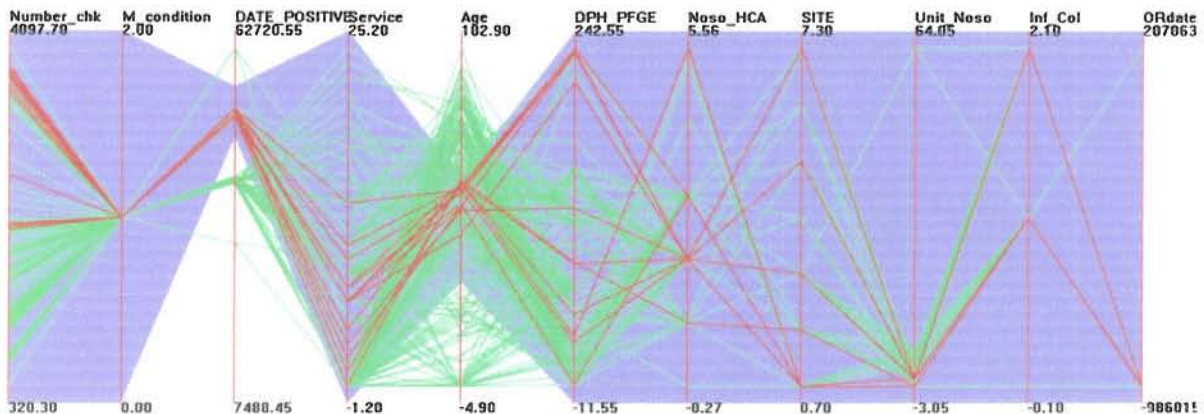
The team had not used any computer visualization programs for data analysis before. Instead, they mostly worked with spreadsheets, which they tried to look at many at a time attempting to decipher the patterns they were looking for. They also actually attempted their own form of visualization by creating a graph of data on a huge sheet of paper on a wall, using tacks to determine data points. However, this method was not terribly effective for them as it took them an entire month to create a graph for a single week's data set.

On March 16th we had our first meeting with the Infection Control team. After giving a brief introduction about what XMDV tool is capable of, they were all very curious to see it in action. Just by hearing the phrase 'finding patterns' you could see them light up as they told us how that is exactly what they need and if XMDV tool could assist them with that purpose it would be a great help. We started off with a simple data set and started showing them the different graphs. They all thought that the way to select variable ranges and selections was very intuitive and easy to use. One function that they suggested be added is the ability to edit variables within the program if they want to input a change or make a correction. The team was very quick to understand how to use the scatter plots for their own analysis, but when we showed them the glyphs they had a much

harder time understanding it. Those who were not entirely computer literate seemed a bit lost on the concept of the glyphs. However, despite some slight confusion they were definitely interested and wanted to see their own data in the program to analyze. We set up another meeting date and they planned to anonymize some data to e-mail us so we could convert it to the XMDV tool format.

On April 11th, we met again with the Infection Control team at UMass Medical Hospital. This time we came with their data prepared ready to be analyzed by XMDV tool, seen below in parallel coordinate view.

Figure 4.3



Visualization of patients who have contracted infections. The brushed region represents patients between the ages of 30 and 60 who were tested positive in 2005.

After changing some variable ranges and trying out the different graphs, they voiced some of their suggestions for improvements of the UI. They found it

relatively difficult to view individual data points and asked for an easy way to increase their size. They also had some problems viewing data points because some of the color schemes made it difficult to see. Specifically, the dimensional stacking graphs were particularly difficult to look at.

One feature that they really warmed up to was the ability to easily remove and re-add variables from analysis. They said it allowed them to use as big data sets as they wanted to start with and could then trim out all the clutter as they saw fit and could easily add or subtract variables if they deem them necessary. For that same reason they liked the feature of being able to save a selected data subset and then re-opening it to only work on that particular selection. One option they thought would be nice would be one to add a new data point from within the program without having to go out and edit the original data file itself. They also asked for a print option to be available with an easy to use button on the UI at the top if possible.

One big hurdle that the team seemed to have was understanding the specifics for converting their data to the proper format. Not being extremely computer literate, they asked for an easy way to convert an excel worksheet, which may contain things such as empty variable values and other things not allowed by the file format. Whether it was an explicit instruction guide, a separate tool, or even a built-in function, they made it clear that converting their data was the biggest hurdle from them trying to use the program on their own. After finishing our meeting, they expressed interest in continuing to work with

XMDV tool. They said that if any future projects could work with them they would definitely be interested.

The meetings we had with the Infection Control team were very helpful in our analysis. First and foremost, we saw that XMDV tool is definitely useful for a medical researcher. Finding patterns and interactions between variables is what they do and this program is a means to achieve those ends. There is room for improvement in the various points they mentioned, including readability, ease of converting data, print options, and possibly tips or instructions for some of the less intuitive graphs, such as glyphs, in the program. Overall, it would appear that XMDV tool would be extremely useful to medical researchers in general assuming that the problems mentioned are addressed.

4.4: Further Meetings

This section covers our meeting with Dr. Larry Garber, a doctor who works at Fallon Clinic. He works with individual patients, as well as looking at whole populations to try and find patients who have similar characteristics, called cohorts, so that he can predict symptoms or vulnerabilities before they happen. Dr. Garber has a lot of computer experience in general including working with other visualization tools such as SAS and SPSS.

We began our presentation by showing him the different graphs with a simple data set and he immediately took a liking to it. He mentioned that all his previous visualization methods were much more static. The other programs couldn't interact as much as is possible in XMDV tool. After going over the basic

graphs he showed a distinct interest in the glyphs, as they would be very useful in finding cohorts as detailed earlier. Glyphs are great at matching similarities across multiple variables and that is exactly what the process of finding cohorts is.

After going over all of the functions Dr. Garber had some suggestions for the program. He had problems viewing the dots on some graphs and thought that an adjustable dot size would definitely be a must. He also mentioned that the glyphs mirroring of variables was redundant and suggested that they instead use more variables in their place.

Overall, it is apparent that XMDV tool could be useful to a clinical doctor. It would not help them in all of their duties, but it could help them in at least one form of their analysis.

Chapter 5: Analysis and Conclusions

5.1 Evaluating the Tool

With both the direct feedback given at our meetings and the completed surveys from our test users, we got a good idea as to the general feelings on XmdvTool, and saw what is most appreciated as well as what might stand to see improvement.

From the discussions we had at meetings, many subjects seemed to understand the basic concept and power of the tool right from the start. The scatterplot view in particular was well-received, seeing as many had worked with that type of graph in some form before. However, while the strength and versatility of the program was understood easily, a reinforcement for its application to their specific purpose was important in warming them up to the tool. Overall, it can be said that familiarity was key - most people seemed interested in finding a more efficient way to do things they already knew how to do or had been doing, rather than jumping into something completely new. The tool was viewed as a way to enhance their existing work methods, although admittedly, learning completely new ways to do what they do would take far more time and effort than what we spent with them.

To get a more in-depth look at what our users really thought, we turn to the survey. Each different view in XmdvTool, as well as other aspects and overall opinion, were asked to be rated on a scale of 1-5 for effectiveness, learning curve,

and accuracy (see survey for more details, Appendix E.). These results are compiled here, separated into the groups we worked with, and the average across the groups is also provided.

	MCPHS	UMass	Fallon	Average
Scatterplots: Effectiveness	4	5	5	4.67
Scatterplots: Learning Curve	5	4	4	4.33
Scatterplots: Accuracy	5		4	4.5
Star Glyphs: Effectiveness	4	2	4	3.33
Star Glyphs: Learning Curve	3	2	4	3
Star Glyphs: Accuracy	3		4	3.5
Parallel Coord.: Effectiveness	4	4	5	4.33
Parallel Coord.: Learning Curve	4	4	4	4
Parallel Coord.: Accuracy	3		4	3.5
Dim. Stacking: Effectiveness			3	3
Dim. Stacking: Learning Curve	2		2	2
Dim. Stacking: Accuracy			4	4
Data Entry: Effectiveness	3	3		3
Data Entry: Learning Curve	2	1		1.5
Brushing: Effectiveness	4	5		4.5
Brushing: Learning Curve	3	4		3.5
Brushing: Accuracy				
Overall: Effectiveness	4	4	4.5	4.17
Overall: Learning Curve	4	3	4	3.67
Overall: Accuracy	4		4	4

Figure 5.1 Survey results

Looking at the results here, we can get a pretty solid idea of which parts of the program were received well and where some improvement might be needed. Scatterplots and parallel coordinates ranked highly in effectiveness and learning curve, meaning that our subjects found these views both easy to use and useful for their work. Star glyphs ranked more moderately, as some seemed fairly comfortable with them while others had difficulty understanding them conceptually and practically. Dimensional stacking rated even lower, although we knew going in that it was a more complex view that not everyone would be able to get.

Data entry was another area that looked as if it could use improvement; the user-provided data often needed to be modified considerably to work it into a format accepted by XmdvTool, and determining what changes needed to be made and how to get them done might have been a bit too much for someone to figure out without some time and effort.

Brushing was an interesting case in that pretty much nobody we met with had worked with anything similar, yet they seemed to understand what its purpose was and how to use it for themselves quite readily.

Other responses to the survey included which types of similar programs had been used by our subjects beforehand, what functionalities were found useful and which were missing or could use improvement. Programs in use seemed to be in the domain of SAS, SPSS, Access and Excel, so keeping that in mind when designing with the goal of user-friendliness would be a good idea. Most respondents cited the parallel coordinates view as particularly useful, even though

they had worked more with scatterplots in the past. The biggest idea for improved functionality was a better interface or program for converting existing data into the appropriate format, which definitely was a sticking point for new users.

All in all, everyone we met with expressed a definite interest in trying to incorporate use of the tool in their work, which we can take as a great sign of the tool's potential use in the medical field. Continued work with the people we met with during our project (as shown in Ch. 5.3) would definitely be possible, and the outlook seems good for expanding the scope of this sort of project to many other users within the medical community.

Section 5.2: Evaluating Ourselves

This section covers what we learned about our methods of establishing contacts and conducting meetings during our project. One lesson we learned from this project is that e-mail can be very limited when trying to reach people who we have not made contact with before. Obstacles like spam filters and the inability to properly catch someone's interest are large factors that contribute to e-mails being an ineffective tool for establishing contacts. We sent out nearly seventy e-mails during our project and only got two responses, both of which were saying that they were not interested in working with us.

On the other hand, making phone calls was a much better approach for establishing contacts. Being able to hear what we were saying and asking any

questions they might have had went a long way towards establishing a successful contact. It was much harder for people to ignore someone talking to them as opposed to a random e-mail that happened to appear in their inbox. It should be mentioned as well that voice mails, while not as effective at provoking a response as reaching the person you wish to talk to, are much more effective at establishing contact than e-mail. One of our contacts started with a returned voice-mail with the others being phone conversations.

If you have any connections or references to the person you are trying to contact, it makes a significant difference. Sometimes they may not understand exactly what you are trying to explain to them, but if they know that someone they trust recommend that you contact them, they would be more inclined to listen and set up a meeting.

Our meeting plans went very well at catching people's interest in the tool and then creating interest to see their own data with the program. For the most part the participants had a relatively easy time following the presentations and whenever they had questions or something they did not understand they were not afraid to speak up and clarify what they needed to. We also feel that all of the information that we wished to convey in these meetings got across to the intended recipients.

Section 5.3: Future Works

The members of the Infection Control team at UMass Medical have demonstrated

a definite interest in continuing to work with XMDV tool. They would most likely be interested in a project that would help them apply the tool to their specific research and teach them how to use the tool successfully on their own.

Richard Ellison, the team member who we worked with most, can be reached by e-mail at EllisonR@ummhc.org, or by phone at 508-856-5166

Larry Garber was happy to work with us, but instead of wishing to work again with us he recommended us to Dr. Sri Emani. Dr. Emani does a lot of work with computer tools and was interested when we told him about XMDV Tool. Unfortunately, he did not have time to meet us during the duration of our project but he said that he would be interested in working with members of any future projects with XMDV Tool. Dr. Emani can be reached by e-mail at srinivas.emani@fallon-clinic.com or by phone at 508-852-0600 (Fallon Clinic)

Professor Michele Matthews did not specifically say that she was interested in a continued project, but is interested in using XMDV Tool. Future projects could contact her for either further work with the tool or references to others who she may think are interested. She can be reached by e-mail at Michele.matthews@wor.mcphs.edu, or by phone at 508-354-0522.

Appendices

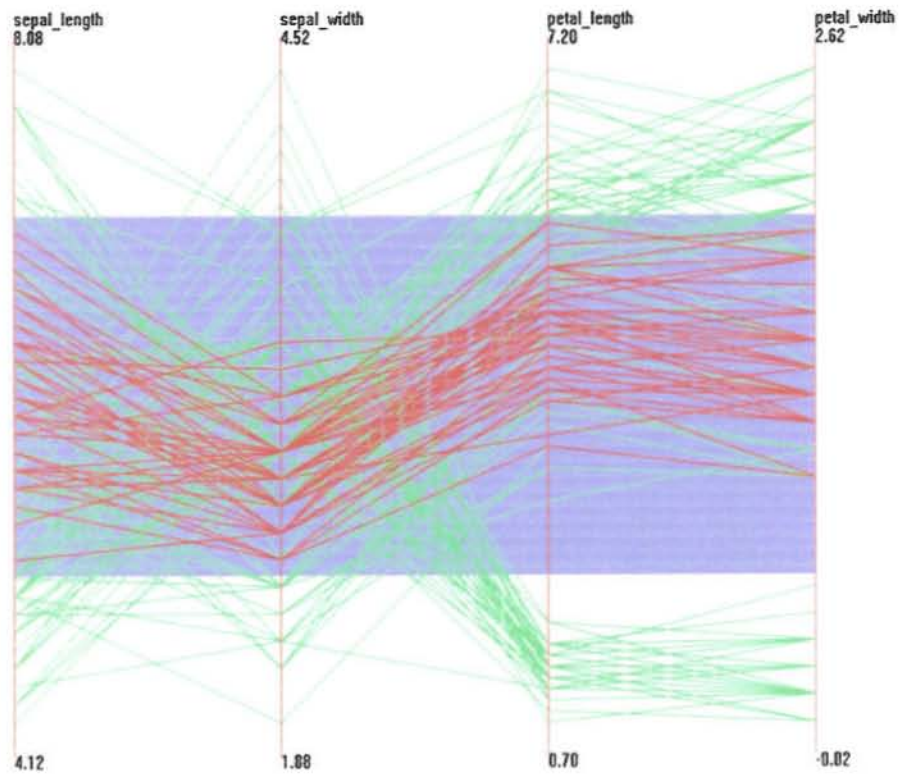
Appendix A: All About XMDVTool

The XmdvTool software kit is a data visualization system intended for use with large, multivariate data sets. It allows the user to identify and investigate trends in their data visually, as well as interactively. It provides a set of four primary views which are each inter-related. This allows users to make changes in one part of the toolkit and see those changes reflected across all other parts. Additionally, there are a number of analytical tools included which assist the user in finding information of interest. Once a particular subset of data is identified as important, it is then possible to view a detailed summary of that data, or simply isolate that subset for further analysis. By allowing the user to effectively shrink their data set by isolating these areas of interest, it allows them to break even the largest and most complex data set down into manageable portions.

This task of identifying key features in a data set might normally be done using complicated command-based software, but can often be done in minutes interactively using an intuitive visual interface that can be learned quickly by non-technical personnel. It is our assertion that this may be of particular importance to those in the medical profession, who have large quantities of data to manage and little time to learn or use new software. Idealistically, the time saved with a tool such as this could save lives.

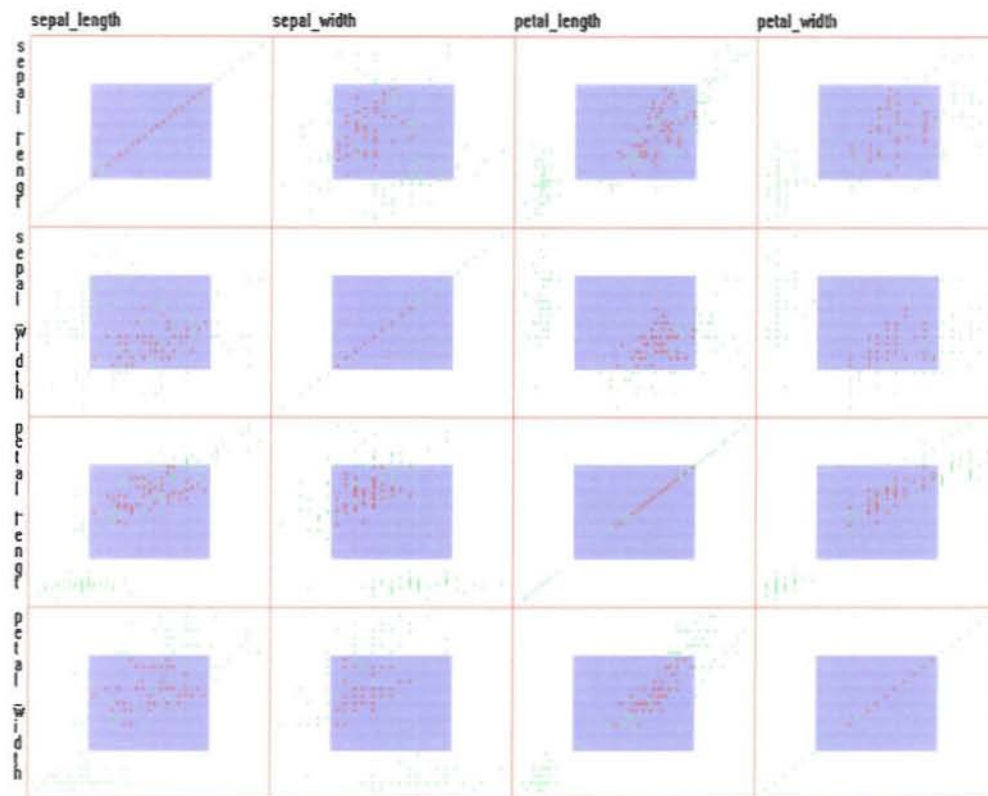
XmdvTool is a public-domain project, free for anyone to use or modify as suits their needs. It is also the subject of ongoing research and development at Worcester Polytechnic Institute. Following are some images of the tool's results.

Parallel coordinate view:



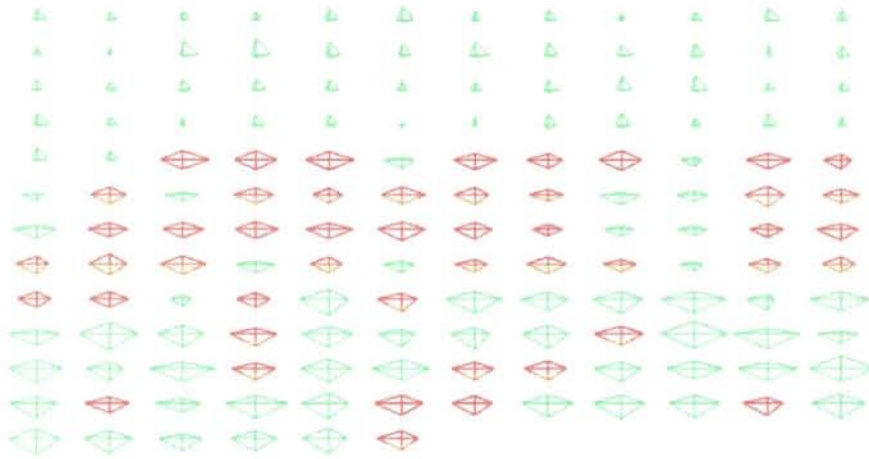
Each dimension is represented as a vertical axis, with data points represented as lines connected across each axis. Only lines that are completely contained within the brushed region are highlighted.

Scatterplot view:



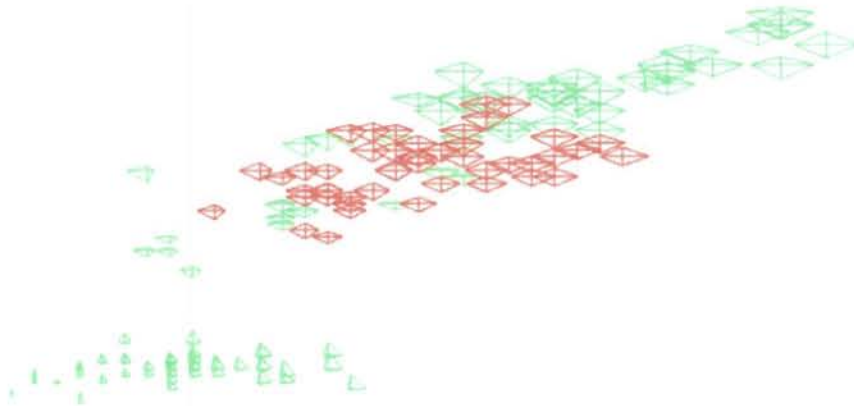
The data is displayed as a grid of scatterplots, with one for each pair of variables.

Glyph view:



Each data point is represented as a polygonal glyph, with each point on the polygon representing a single dimension.

Glyphs with derived positions:



This is a glyph view where the positions of the glyphs are derived from the data, forming another sort of scatterplot.

Appendix B: Boilerplate Emails

Below is the boilerplate used to contact doctors at the University of Massachusetts Infectious Diseases department.

Dear Doctor:

We are undergraduate students at Worcester Polytechnic Institute, and we are conducting an evaluation of the usefulness of a particular data visualization tool, developed at WPI, in a variety of data analysis tasks within the medical domain.. If you are willing to participate in this evaluation via demonstrations and questionnaires, we would like to see if this visualization tool can aid your research with infectious diseases and immunology. We also believe that it may be interesting or helpful to your students, and would be willing to provide a demonstration for them. The tool supports the interactive visual analysis of multivariable data sets using a number of different visualization routines. This can be useful for analyzing data sets such as a population, detecting and observing patterns such as the spread of a disease, and finding trends such as the growth of a specific virus over years. If you would like to look at the official web page of this tool, it can be found at <http://davis.wpi.edu/~xmdv/index.html>.

If you are interested, or would like more information, please contact us at mjarmak@wpi.edu. If you know of any colleagues who would be interested in participating, we would greatly appreciate their contact information as well.

Thank you for your time,
Matt Jarmak
Chris Gianfrancesco
Steve Kolk
Owen Pedrotti

Below is the boilerplate used to contact professors at the Massachusetts College of Pharmacy and Health Sciences.

Dear Professor:

We are undergraduate students at Worcester Polytechnic Institute, and we are conducting an evaluation of the usefulness of a particular data visualization tool, developed at WPI, in a variety of data analysis tasks within the medical domain.. If you are willing to participate in this evaluation via demonstrations and questionnaires, we would like to see if this visualization tool can aid your research. We also believe that it may be interesting or helpful to your students, and would be willing to provide a demonstration for them. The tool supports the interactive visual analysis of multivariable data sets using a number of different visualization

routines. This can be useful for analyzing data sets such as the reaction of different people to a medicine, detecting and observing patterns such as the effectiveness of treatments on average people, and finding trends such as the usefulness of a drug over time. If you would like to look at the official web page of this tool, it can be found at <http://davis.wpi.edu/~xmdv/index.html>.

If you are interested, or would like more information, please contact us at mjarmak@wpi.edu. If you know of any colleagues who would be interested in participating, we would greatly appreciate their contact information as well.

Thank you for your time,
Matt Jarmak
Chris Gianfrancesco
Steve Kolk
Owen Pedrotti

Below is the boilerplate used to contact doctors at the Worcester Medical Center.

Dear Doctor:

We are undergraduate students at Worcester Polytechnic Institute, and we are conducting an evaluation of the usefulness of a particular data visualization tool, developed at WPI, in a variety of data analysis tasks within the medical domain. If you are willing to participate in this evaluation via demonstrations and questionnaires, we would like to see if this visualization tool can aid your research and possibly treatments. The tool supports the interactive visual analysis of multivariable data sets using a number of different visualization routines. This can be useful for analyzing data sets such as patient records, detecting and observing patterns such as the spread of diseases, and finding trends such as the average statistics of patients across the years. If you would like to look at the official web page of this tool, it can be found at <http://davis.wpi.edu/~xmdv/index.html>.

If you are interested, or would like more information, please contact us at mjarmak@wpi.edu. If you know of any colleagues who would be interested in participating, we would greatly appreciate their contact information as well.

Thank you for your time,
Matt Jarmak
Chris Gianfrancesco
Steve Kolk
Owen Pedrotti

Appendix C: Outline of Training Program

General Practice in Training

* At all points in the training, keep coming back to the point of how narrowing down individual variables can give the user a clearer view of how their variables are related.

* When manipulating the brushes, try narrowing one variable, and then two while leaving the others wide, until the number of highlighted data points is small. This will emphasize that the tools can be used to break up large amounts of data into manageable pieces.

* Mention at some point that the tool provides a function for saving pictures of the data. Probably best to reiterate this point at least once.

Parallel Coordinates Training

1. Explain to the user that they are looking at information taken from a simple CSV file, with each line representing a single data point intersecting a set of axis.

2. Demonstrate how to manipulate brushes manually with the left and middle mouse buttons, indicating how only items that fall entirely within the brush are highlighted.

3. Explain that the above concept is seen throughout the toolset, and how it allows the user to isolate individual variables and see how their distributions relate to other variables.

4. Demonstrate Data Values for Brushed Region menu to show how one may read the actual values.

5. Demonstrate the brush toolbox, and how multiple brushes can be created, with rules for how they should behave if they overlap.

6. Demonstrate Dimension On/Off/Reorder, and how it can be used to eliminate extraneous dimensions, or to place related dimensions close to each other for the sake of comparison.

7. Demonstrate Time Series Animation and how it may be used to determine whether relationships are positive or negative by showing how they change over time.

Scatterplot Training

1. Demonstrate how brushes may be manipulated in this view in much the same way as the parallel coordinate view.

2. Demonstrate basic zooming and panning, also mention that things such as colors may be changed but do not demonstrate. Mention that these things may be done in any view.

3. Demonstrate that the Parallel Coordinates Brush Toolbox may also be used to modify the brush without leaving the scatterplot view.

Glyph Training

1. Demonstrate the Glyph Brush Toolbox, and explain how each axis represents a dimension, creating a two dimensional polygonal shape equivalent to the parallel coordinate brush wrapped into a loop. Show how this view can be manipulated the same way. Leave this view open and in view for the remainder of this section. Also mention that this toolbox may be used in any view.

2. Demonstrate that the glyph key can also be used as a quick reference for how the dimensions are represented.

3. Demonstrate how clicking on a glyph in the main view will shift the values of the brush to encompass that glyph, the result being that similar glyphs to the one clicked are highlighted, within a tolerance defined by the thickness of the brush.

4. Demonstrate glyph placement, first with basic ordering. Then show how the data can be used to create a scatterplot of glyphs by using derived placement by sepal and petal length as axis. Show again how clicking on a glyph can highlight similar groups of glyphs.

Dimensional Stacking Training

1. Show how the dimensional stacking view shows nested grids to represent data.

2. Use dimension stacking key to show what the dimensions mean.

3. Show dimension stacking brush, note that it can be used in any view.

Appendix D: Transcripts of Notes from Meetings

First meeting with MCPHS: 2/21/06

- Some confusion about what was being represented (mainly in parallel coord)
- Recognized software's capabilities, thought it looked powerful yet relatively user-friendly
- Saw potential usefulness with large datasets

? How does it compare to SMSS/SAS?

A: Unfamiliar with their capabilities

? How to view original number data

A: Found option to view selected entries

? How could this be of use for them?

A: Return with example using their own data

Second meeting: 3/29/06

- Understood meaning of visualizations better with own data
- Saw helpfulness of brushing/selection (M/F example)
- Seems user-friendly, eager to try for herself
- Scatterplot more understandable, parallel least
- Use in presentations (scatterplot)
- Seemed ok with 60-point dataset, figured would be even better with more
- Again interested in viewing numbers

? How to perform statistical calculations (correlation etc)

A: Not done by itself

First meeting with UMass on March 16th:

- contagious disease tracking
- reduce risk of spreading
- seem interested, engaged
- very intrigued at idea of "finding patterns"
- related to data mining
- couple starting to lose interest, staring
- seemed slightly confused at glyphs
- some quite confused, not understanding
- interested in excel use specifically
 - looking at raw data
- interested in scatter plot for their data
- considered valuable
- asking questions about how crowded it gets
- like the simple functions
 - said "That is really neat" for selecting variables
- wants to see overlapping variables/ordering variables
- some seem more comp. literate, more interested comparatively
- sorting by various, catching
- want to merge data
- having trouble with confidentiality
- wanting to work with changing vars
- like the pictures
- This program helps determine what to analyze

Second meeting with UMass on April 11th:

- want to show pattern of exposure, cause?
- would like way to enter fields easily
- colors are hard to see
- dots are too small
- trying to find patternsn easily

- using key (numerated data) seems slow
- dimensional stacking colors (green small dots on white)
- some error in converting access to excel
- if error, easy to correct
- want larger dots
- like easy way to select and change in scatter plot view
- still seems excited, “intriguing”, think it would be useful
- clean data = good
- too many variables can be confusing to look at
- way to isolate variables
- features are good
- clusters at 5-10 data points
- make patterns different colors (easier to look at)
- make line size bigger
- easy way to import data!
- outline of how to data format
- zoom options not 100% intuitive
- runs slow on old machines
- can make use of program
- easy way to manually add data
- the save file with narrowed field good
- medmine, redbat
- modify display (easier to read/show)
- print option
- state health syndromic surveillance

Meeting with Doctor Garber on April 14th:

- made comment about Owen’s laptop (computer literate)
- interested in glyphs
- cross pop, want to look at disease patterns
- very valuable for cross pop
- not as valuable for single patient (easy to do already)
- want to change size of dots
- really interested in glyphs “very cool” “this is neat”
- very useful for data mining
- do research study, find cohorts as research patients, want to find similar
- sees it as very useful for doing that setting parameters to find people
- should not reflect same variables “redundant”
- use remaining variables
- analyzing pop opatient + cohorts
- putting in electronic records
- whole dept. on data mining

Visualization & XmdvTool Analysis

Survey

Worcester Polytechnic Institute

Application of Visualization IQP

Steve Kolk, Matt Jarmak, Chris Gianfrancesco, Owen Pedrotti

Advisor: Prof. Matt Ward

Purpose

This survey is designed to help us understand your thoughts regarding your experience with XmdvTool and visualization techniques in general. We are hoping to use this information to draw meaningful conclusions about the role of visualization in data analysis, including what aspects of visualization are most effective, how it compares to other methods of data analysis, and so on. Furthermore, we are seeking to determine the strength of XmdvTool itself as a visualization application that can aid in data analysis.

Questions

1) What data analysis and visualization tools do you currently use?

2) If you have worked with one or more visualization tools, what types of visualizations have you worked with before?

- Scatterplots
- Star glyphs
- Parallel coordinates
- Dimensional stacking
- Other (please specify)

For each of the following, you will be asked to rate certain features of the tool based on certain criteria.

If you **have** worked with other visualization tools, please rate XmdvTool in relation to previous tools you have used, with a 1 meaning XmdvTool is very weak compared to other options and a 5 meaning XmdvTool was very strong compared to other options.

If you **have not** worked with visualization tools, please rate XmdvTool in relation to other sorts of relevant data analysis techniques, with a 1 meaning XmdvTool is very weak compared to other options and a 5 meaning XmdvTool was very strong compared to other options.

3) Rate the effectiveness of these aspects of the tool. (How well would they aid in your everyday work?)

- Scatterplots 1 2 3 4 5 N/A
- Star glyphs. 1 2 3 4 5 N/A
- Parallel coordinates 1 2 3 4 5 N/A
- Dimensional stacking 1 2 3 4 5 N/A
- Data entry 1 2 3 4 5 N/A
- Brushing 1 2 3 4 5 N/A
- Overall. 1 2 3 4 5 N/A

4) Rate the learning curve of these aspects of the tool. (How difficult was it / How long did it take to learn to perform essential tasks?)

- Scatterplots 1 2 3 4 5 N/A
- Star glyphs. 1 2 3 4 5 N/A
- Parallel coordinates 1 2 3 4 5 N/A
- Dimensional stacking 1 2 3 4 5 N/A

- Data entry 1 2 3 4 5 N/A
- Brushing 1 2 3 4 5 N/A
- Overall. 1 2 3 4 5 N/A

5) Rate the accuracy of these aspects of the tool. (How do the results and observations using these match up with other tried-and-true methods?)

- Scatterplots 1 2 3 4 5 N/A
- Star glyphs. 1 2 3 4 5 N/A
- Parallel coordinates 1 2 3 4 5 N/A
- Dimensional stacking 1 2 3 4 5 N/A
- Brushing 1 2 3 4 5 N/A
- Overall. 1 2 3 4 5 N/A

6) Would you be interested in incorporating this tool into your own research or work?

- Yes
- No

7) Were there any functionalities that you found particularly useful?

8) Are there any important functionalities that you thought were missing?

Thank you for taking the time to participate in our demonstration and survey. Your help is greatly appreciated, and will contribute much to our project's success. If you have any further questions or concerns about XmdvTool or our project, feel free to email us at mjarmak@wpi.edu anytime.