Project Number: MXC-I004 - 46

Project-404: Results of Electronic Information Exchange in a Free Society

An Interactive Qualifying Project Report
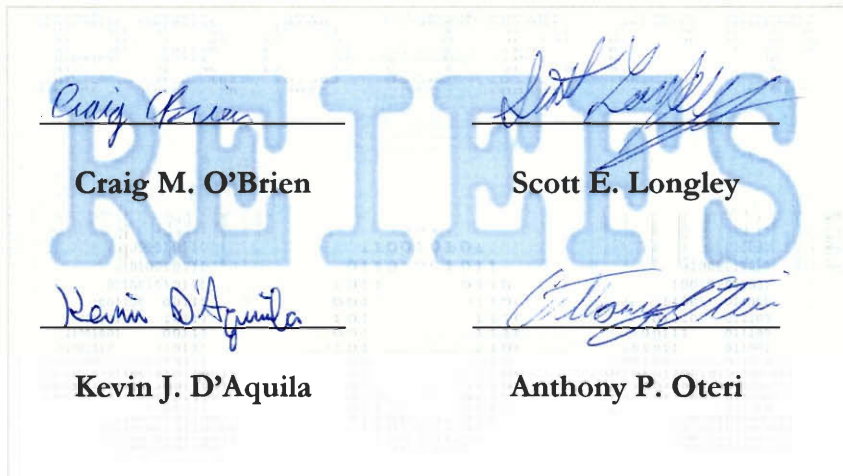
Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

Craig M. O'Brien   **Scott E. Longley**

Kevin J. D'Aquila   **Anthony P. Oteri**

Date: May 2, 2003

Approved:

Professor Michael Ciaraldi, Project Advisor

Abstract

In the late 20th century, the rapid advances in digital technology became a tremendous influence on people's lives. The outlook and reactions of youths to these changes, whether positive or negative, will guide the course of information technology in the future. The ways the Internet influences and interacts with young individuals deserve a serious examination. This project reveals the changing nature of the current usage of computers and the Internet.

Table of Contents

Introduction

Report Overview
Purpose

Report Overview

In Project 404, we used a survey that targets young people, and gathers information about young technologies. This study is designed to update existing ideas about the realities surrounding users of the Internet and computers. We believe that the present time is unique in a few ways that may prove important. Computers are finding many new uses, and Internet access is spreading. However, computers are still a young technology, and the utility they find is often user-specific. Some regard Internet access as a necessary household utility, while others regard it as an eccentric luxury. These differences span income ranges, geography, and technical proficiency. Computers are also being introduced into classrooms more than ever. These facts have led us to believe that we live in a particularly interesting time for the developing use of these technologies. We decided to focus our study on developing people, so that we could more readily see the changing trends in our society's use of technology.

Purpose

The foremost purpose of Project 404 is to expose truths. Instead of using thousands of test subjects to prove existing theories, we have chosen to focus on exposing interesting trends at work. We intended to expose a broad range of conclusions that reveal the nature of the times in which we live. Data mining allowed us to explore our results at a level that surpasses our predictions. By reporting accurately on the current state of affairs, we can show a societal reality While no single person can fully experience this reality, it affects all who live within it, so the information we provide should be beneficial to those that look forward to technological progress.

History

What is an IQP?

Worcester Polytechnic Institute places heavy emphasis on its Projects Program. Each student must complete three projects to graduate. One of these projects is the Interactive Qualifying Project, or IQP:

> The Interactive Qualifying Project (IQP) challenges students to identify, investigate, and report on a self-selected topic examining how science or technology interacts with societal structures and values. The objective of the IQP is to enable WPI graduates to understand, as citizens and as professionals, how their careers will affect the larger society of which they are part (About the WPI Projects Program).

"Results of Electronic Information Exchange in a Free Society" is a unique IQP, yet students before us have completed projects that share a resemblance at one level or another. The "KMUTT Internet Survey" studied the use of the Internet in academic work by faculty at King Mongkut's University of Technology Thonburi, Bangkok. The KMUTT Project analyzed why and how faculty members were using the Internet for their academic work. "Napster: Impact on Intellectual Property Law" presented a study about the Napster file sharing program, how it spread across the Internet, and how it affected the music industry.

"Introducing Young Children to Basic Computer Skills" studied the technology gap between technologically developed countries and those that were behind. The project placed focus on introducing younger children to basic computer skills at the Duang Prateep Foundation of the Klong Toey slum in Bangkok, Thailand. Lastly, "Analysis of Online Societies" asked players about their social experiences in a virtual world game, and in the real world. This study made conclusions about how social participation in a virtual world affected a person's social behavior in the real world.

History of the Internet

The massive networking infrastructure and communications system known as the Internet took many years to develop into what it is today. Many individual technologies and concepts encompass the whole of the Internet. J. C. R. Licklider of MIT wrote a series of memos in August of 1962, discussing his "Galactic Network" concept. This was the first recorded description of enabling social interactions through networking. Licklider was the first head of the computer research program at DARPA, the Defense Advanced Research Projects Agency. A man at MIT named Leonard Kleinrock published a paper in July of 1961 about packet switching technology, and convinced an MIT researcher at DARPA named Lawrence G. Roberts of the feasibility of using packets instead of circuits (A Brief History).

Four years later, Roberts worked with others to connect a computer in Massachusetts to one in California using a low speed dial-up telephone line. This was the first wide-area computer network. After this success, Roberts went to DARPA to develop the computer network concept. In 1967, he published his plan for the "ARPANET". In 1970, the Network Working Group (NWG) finished the Network Control Protocol (NCP), which was to serve as ARPANET's initial Host-to-Host protocol (A Brief History).

The first public demonstration of ARPANET was at the International Computer Communication Conference in October of 1972. Earlier that year, the concept of electronic mail was developed. Bob Kahn, one of the DARPA team, introduced the idea of open architecture in 1972. This developed from a radio packet program he was working on, called "Internetting". NCP was restrictive; it was more of a low-level device driver than a communications protocol. To make an open networking architecture possible, Kahn went on to develop what became the Transfer Control Protocol/Internet Protocol (TCP/IP) as a replacement for NCP (A Brief History).

TCP/IP started as one protocol, called TCP. Host systems automatically re-sent packets dropped in transit. Usage of certain applications, such as voice over IP, showed that this was not conducive to optimal operation in every application. IP became a distinct entity, and the User Datagram Protocol, or UDP, served as an alternative that would not re-transmit lost packets (A Brief History). These protocols serve as the primary conveyers of Internet traffic today. It was not until a decade after TCP/IP's inception that ARPANET transitioned from NCP to use it. This transition took several years of planning, but was quite successful (A Brief History).

Many implementations of network technology have developed over the years to carry packets around the world. Perhaps the most popular, currently, is Ethernet technology. Bob Metcalfe developed Ethernet technology at Xerox in 1973. The Internet was starting to spread at this point, and the number of hosts was increasing beyond the capability of the naming systems that were in place. Each system had a hostname, so people did not have to memorize numeric addresses. With a small number of hosts, it was viable to simple store each hostname and corresponding numeric address in a lookup table. As the networking paradigm shifted toward increasingly numerous independent Local Area Networks, the simple lookup table did not suffice. Paul Mockapetris of USC/ISI invented the Domain Name System, or DNS. DNS still serves today as a scalable means of resolving hierarchical host names across the Internet (A Brief History).

Out of this evolving network rose the development of Tim Berners-Lee's creation: the World Wide Web. A coordination organization formed, called The World Wide Web Consortium, or W3C. Tim Berners-Lee led the W3C, along with Al Vezza, from MIT's Laboratory for Computer Science. The W3C is responsible for many of the evolving protocols and standard associated with the web (A Brief History).

As the Internet continued to develop over the years, the Federal Networking Council finally passed a resolution to define the term "Internet" on October 24, 1995:

RESOLUTION: The Federal Networking Council (FNC) agrees that the following language reflects our definition of the term "Internet". "Internet" refers to the global information system that -- (i) is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons; (ii) is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP-compatible protocols; and (iii) provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure described herein (A Brief History).

According to Gene I. Rochlin, author of *Trapped in the Net*, "The Internet came into being almost by accident" (Rochlin, 44). ARPANET was split into two networks in 1983, the research-only side now being called ARPANET, while the military part became MILNET. By 1989, ARPANET was disbanded, because commercial networks had become the backbone of the Internet, taking over its functions (Rochlin, 44). Since then, the commercial and military backbone of the Internet has grown tremendously, from about 2000 users and 213 host computers in 1981, to tens of thousands of computers in 1989. In 1993, the Internet became available to civilians, and had a total of over 1.5 million computers (Rochlin, 44). The number of people on the Internet today is difficult to determine. Estimates on the Web cite numbers ranging from 300 million to over 630 million.

It is hard not to think of the World Wide Web as being one and the same as the Internet. The WWW, though, was not the only interactive content proposal for the Internet. The Xanadu project, pioneered by Ted Nelson, was more complex. The Xanadu project embodies dissatisfaction with the hypertext system that is in wide use across the WWW, calling the web merely "an imitation of paper" (Project Xanadu History). The Xanadu proposal claims features such as bi-directional links, and automatic version and rights management. The Xanadu project also claims to be the inspiration for Tim Berners-Lee's WWW proposal, Ray Ozzie's Lotus Notes, and Bill Atkinson's HyperCard, in addition to others (Project Xanadu History). While Ted Nelson developed his original vision of Xanadu in 1960, to this day it has not largely developed, instead only spawning off side-projects.

Previous Research in our Topic

The Internet of today is composed of seemingly endless information and content. As a resource, there is no limit as to how useful it may be. Many journalists and authors, though, focus on the dark side of the Internet. Like any information medium, the Internet contains illicit materials. It seems that children's use of the Internet, especially, raises strong concerns among parents much more than children's exposure to other mediums. Don Tapscott, author of *Growing Up Digital*, thinks that part of the problem is that "many…adults feel they don't really understand the new technology themselves. Exacerbating these anxieties is a lot of negative media coverage, often focusing on the 0.5 percent of online material that is violent, racist, or sexual in nature" (44).

The Internet, though started in the United States, is an amalgamation of cultures. When Halsey Minor, CEO of CNET Inc., was asked if American culture will dominate the Internet, he replied, "Basically I've always looked at the Internet as an entirely different country—a country without rulers, a country without borders, and a country that's very, very difficult, if not impossible, to regulate. And slowly but surely we're all getting co-opted in this country and spending more and more of our time there" (Sikes, 249).

In 1997, an author named Theodore Roszak wrote an article entitled, "Internet as teacher makes students stupid," stating that technology should be kept away from children. Roszak goes on to explain, "The World Wide Web is primarily an advertising medium," and that "used as a teaching device, the Web is an expensive way to distract attention and clutter the mind" (Tapscott, 45). In a nationally syndicated column in 1998, Cokie Roberts and her husband write, "[The Net] is yet another

influence in…children's lives over which [parents] have no control" (Tapscott, 45). Inventions such as radio and television were the norm for the parents of today's generation of kids. Children's viewing of television and magazines is fairly easy to control. In reference to parents, Tapscott proclaims, "For the first time in history, children are more comfortable, knowledgeable, and literate about an innovation central to society. … They are a force for social transformation" (Tapscott, 1).

It is no wonder that many parents are concerned; illicit material on the Internet comes in many forms. After the Oklahoma City bombing, television reports condemned the availability of information about bomb making on the Internet. "While two New Jersey teenagers are injured in their attempts to make a pipe bomb, they tell reporters that they learned they could access the construction instructions via the Internet from watching television news" (Tapscott, 45). A *Baltimore Sun* columnist tells parents, "If you have a teenage boy and you are on the Internet, he is downloading pictures of naked women. … If you are thinking, 'Oh no, not my boy,' give it up because he is counting on that" (Tapscott, 45).

Yet other things on the Internet are bittersweet. Paul Matteucci, CEO of HearMe, writes:

> I've been in chat rooms where I've observed, for the first time in my life, African-Americans and white supremacists talking to each other. I'm not sure that's a good thing or a bad thing, but if you go through the threads of the conversation, by the end you'll find there's less animosity than there was at the beginning. It's not pretty sometimes, there's shouting and obscenities, but these groups of people would never even talk to each other in real life. They would talk through the press at each other or they would talk through hate sites. But here they are online, actually talking to each other. I find that fascinating and uplifting (Sikes, 15).

The Internet does not adhere to the same trends and tendencies of previous media. It is a new paradigm for information conveyance, for expression, for fun, and for profit. There is dangerous and questionable material in flux with potentially endless informational resources and stimulating educational opportunities at any surfer's disposal. Because of this, exposure to the Internet may be socially and cognitively affecting children as they grow up.

Another concept that has captivated young minds is that of the video game. Tapscott draws a parallel between video games and computers, "Conventional wisdom says that because boys spend more time playing video games than girls, they develop sharper computer skills and are more comfortable working with technology later in life. However, studies conducted since the advent of the home computer, video console, and hand-held Game Boy have noted that girls have been playing as enthusiastically as their male peers" (166).

Television spread throughout the world quickly after its inception. Kids were increasingly enthralled by it, and spent much of their time on passive observation of the glowing screen. Now many kids sit in front of a glowing computer screen instead. The difference, Tapscott claims, is that "TV is controlled by adults. … In contrast, children control much of their world on the Net. It is something they do themselves: they are users, and they are active. They do not just observe, they participate" (25).

Tapscott claims that in 1998, 2 out of 3 kids used a computer either at home or at school (3). He further states, "… in the United States, there is a direct relationship between family income and access to computers and the Net. This correlation also exists between the higher- and lower-income schools" (11). He also claims that at the time, kids watched less television than five years prior to then, and much less than their parents did as children (2). Instead of a decline, "children appear to be developing their cognition, intellects, and abilities better" (Tapscott, 104). Pertaining to social development, Tapscott claims further that children growing up with the Internet today are "smart, fluent, social, analytical, self-reliant, curious, contrarian, creative, articulate, media-savvy, bored with television—a child that interacts with his world, and creates and achieves balance" (124).

It is important to remember that the development of technologies such as computers in general and the spread of the Internet are not easy to predict. In his book entitled *The Social Impact of Computers*,

Gerald A. Silver discusses his thoughts about computers and their place in society. Published in 1979, Silver's book portrays a view before the ubiquity of the Internet existed:

> An enlightened population is in the best position to deal with rapid and fundamental changes that the computer is bringing. On graduation from college, most people have been exposed to the basic disciplines of English, literature, humanities, history, mathematics, and science. These and other subjects are part of the necessary course work in becoming an 'educated' person.
> The principles of computers must be integrated into grade school, high school, and college curricula. Students should view the computer as a general purpose problem-solving tool, as useful as mathematics or English.
> Computers are dominating the data processing activities of most large organizations. As such, citizens will need to understand basic concepts, programming languages, and computer logic. An understanding of these elements will be necessary in order to function in a world of increasing complexity and sophistication (Silver, 327-328).

Alvin Toffler, author of *Future Shock*, *The Third Wave*, and *Power Shift*, has a different view. His first computer, in 1976, was a custom-built system with 16K of memory, two floppy drives, and a Panasonic television for a monitor. The system cost $3000. Toffler and his wife, Heidi, went online through ARPANET. Even back then, e-mail "could be crushing." "We went away one weekend leaving the system on and came back so avalanched and overloaded with material we turned the damn thing off." Regardless, Toffler believed that the technology was "an extraordinary phenomenon that would spread to a much wider audience." He explains further that the Internet takeoff came when the media discovered the Web in the nineties. "I would have expected it to have developed sooner than when it did" (Sikes, 244-245).

3

Hypotheses

Social Aspects of the Internet
Economic and Technological Impacts
Correlations from Data Mining

Before any information can be gathered about computer and Internet use, some solid hypotheses must be formed, and the topics of interest clearly defined. The scope of our hypothesis can be broken down into three sections: social aspects of the Internet, economical factors, and technological impacts of the Internet. Each of these factors relates to people's proficiency with computers, as well as what they use them for, and for how long. Our Hypotheses apply to Americans, but were only tested in eastern Massachusetts. Hypotheses appear in underlined text.

Social Aspects

Since computers have become more mainstream, the old stereotype of a nerd sitting at a computer has been replaced by the reality that ordinary people use computers for important tasks. People are using computers at younger ages, and for greater lengths of time than ever before. This represents a social shift with youths using computers at a young age that was not present in previous generations. Why would computers and the Internet be more popular now? Computers offer several benefits, of which even young people can take advantage. The major reason younger people are suspected of using computers more often is the entertainment value one can derive from the use of a computer and the Internet. One day, computers may even replace television as the primary source of entertainment in the average household. The scholastic responsibilities of children in high school, middle school, and grade school are such that children have enough free time to spend using a computer extensively.

A very important topic concerning youths using computers is that of parental rules and restrictions. The Internet can be a dangerous place, as stalkers, hackers, child molesters, and other questionable characters can associate freely with impressionable young people. It also contains uncensored content, such as pornography, pirated software, and gambling, and is a medium for many types of fraudulent activities. Informed parents may have many concerns about their children being exposed to this kind of content on the Internet. These concerns generally translate to household rules. The types of rules parents assert on their children with Internet use are important because they give clues about what the parents perceive, think, and feel about their children. A parent that has several strict rules for their children shows a significant lack of trust that their child will be able to not only behave, but to deal with the content.

Girls are generally targeted more often on the Internet than boys are, so informed parents will likely be more protective of their daughters. The last thing a parent wants to happen is to have their daughter be the target of a less than honorable stranger on the Internet. While it true that they would have the same aspirations for their son(s), there is considerably less risk involved. Boys have certain habits that girls do not often have to lead to rules targeted at them as well. Boys, especially ones that are experiencing puberty, tend to be very curious about the female anatomy. The vast majority of pornography is marketed towards men, and boys are far more likely to indulge in it than girls are. The age of puberty coincides with 7th and 8th grade for most boys. Therefore, it would not be surprising if boys around this age had the most restrictive rules on content. For these reasons, boys should be more likely to be restricted with what sites they are allowed to visit, whereas girls would be more restricted in whom they are allowed to talk to on the Internet.

The other important factor with regards to rules is age. Generally, as children grow older, parents tend to be come less protective of them. They are more willing to give them space and assume they will be responsible. With increased responsibility, older youths are granted greater freedom from restraints that would be imposed on younger children. Higher maturity levels allow for parents to ease rules and constraints on children because of the increase of maturity, and ability to handle obscene content maturely. As students get older, their online rules become less severe.

Rules are an important aspect of how someone uses the Internet. However, these issues are not as important as a person's outlook and attitude on computers. One of the most significant influences on someone's life is their upbringing. This especially relates to an individual's computer use. People are different, and demonstrate significantly differing abilities and skills with Internet use. It is only natural that some of this ability and knowledge would rub off on younger generations, especially if parents are active and involved in their children's lives. As with just about anything, if learning starts at a young age, knowledge and involvement will be more permanent. For example, if someone is taught a second language at 5 years of age, it tends to be a lot easier and stick better than if that person was taught at 50 years of age. This same principle applies to proficiency with computers. Children who have parents that show them at a young age how and why to use a computer will most likely be better at using computers than peers that do not.

As people get older, their need for a computer increases, especially with academics. Fourth graders do not need to do as many essays, reports, research, etc. as older students do, and consequently, do not need to use a computer as much as a tenth grader. Doing something as simple as word processing may not be very significant, but it still demonstrates that people consider a computer the correct tool to use for such tasks. In the old days, people did research in a library, with books; now it is more popular to get research and information online. People communicated on the telephone, now many people do it through instant messaging systems. This all leads to the same hypothesis: computer related activities are replacing their non-digital counterparts.

Economic factors

In order for a child to become proficient with computers, consistent computer and Internet access is essential. A computer has to be readily available either in the home or in a public setting, such as a library. People with higher incomes tend to have more money available and therefore can afford to spend more on newer computers as well as better Internet connections. Likewise, poor people have a greater probability of having an old/outdated computer, or none at all. Parents who have higher incomes are more able to provide their children with better computer access.

People that have a faster connection tend to use the Internet more because there is less frustration and stress due to less waiting or tying up of phone lines. Dial up connection tie up a phone line, so that it cannot be used to make or receive calls. In those households where Dial-Up is the only means of Internet connection, it is more likely that there will be rules about how much time they can remain online. People that have a broadband connection don't have to worry about tying up the phone line, and have more leeway in regards to how much time they are allowed to use a computer. Furthermore, because of the slower speed which dial-up offers, streaming and downloaded media take much longer to get. This will result in less use of a computer for watching videos, and listening to music.

Technological factors

There are several types of computers available to consumers. These systems have a lot of differences between each other, many of which could result in varying uses and proficiency levels with said systems. These topics will be covered by the survey. It is interesting to note, what kind of affect the computers used in the school systems have on the students' ownership of computers. Will they be the same? A lot of schools use Macintosh computers. How many students will follow by getting a Macintosh themselves? Most computers that children have were purchased by their parents. If PC's are more popular amongst parents, then their children will probably have them as well, regardless of what their school system uses.

An important thing to be ascertained from this research is what people are using computers for. Are they doing school work? Playing games? Chatting? People use computers for many different reasons, and during this project we will be able to compare relative levels of use at different ages, and find differences and correlations. This will be compared among the individual survey takers as well, drawing correlations between time spent on the computer and time spent on other things, like hobbies, friends, sports, etc.

Correlations from Data Mining

Upon conducting the survey, we realize that we will not be able to predict every correlation that is going to present itself in the survey data. From the use of data mining, future results will allow us to draw correlations and make connections so that we can better understand what is really happening. As with any survey, these results will have a small margin for error, but as the number of survey takers increases, the more accurate the results will be.

4

Procedure

Communication
Planning
Survey Design
Choosing Survey Sites
The PR Package
Making the "Sale"
Conducting the Survey
Data Entry
Statistical Analysis
A Note on Software
The Final Report
Issues and Problems

Communication

The uninhibited flow of information is critical to coordinating a group. As a group, we were faced with the challenge of timely communication between ourselves, our advisor, and participants in our survey. Weekly scheduled meetings have been useful from the start as a way of keeping everyone in sync. E-mail was useful for communicating when it is difficult to meet during the rest of the week. For this reason, we set up the following mailing lists at our domain, project-404.net:

info@project-404.net	forwards to all four team members
all@project-404.net	forwards to all project members, and our advisor
reiefs@project-404.net	same as "all@project-404.net"
tony@project-404.net	forwards to Anthony Oteri
kevin@project-404.net	forwards to Kevin D'Aquila
scott@project-404.net	forwards to Scott Longley
craig@project-404.net	forwards to Craig O'Brien
advisor@project-404.net	forwards to Professor Ciaraldi

The list that received the most use was info@project-404.net, which was useful for communication with survey participants as well as communication within the group.

Of course, it is also necessary to keep records. For our private use, we put text files with useful contact information, outlines, ideas, and drafts on a file-serving system in a private network. For more public information such as progress reports, public relations materials, and web resources, we set up a website (http://www.project-404.net) on a separate system. Using two systems allowed a logical division of information to benefit us in many ways including data security, ease of use, and the reliability that comes with redundancy. In the later phases of Project 404, we centralized our data mining software and related files on WPI's new Linux machines, to make use of their superior processing power. We found that keeping files centralized was critical for avoiding version control problems (see Procedure: issues and problems).

Planning

Project 404 began as an idea inside Anthony Oteri's head:

> Social Implications of a technology based society. In particular, pros and cons of Internet dependencies, online shopping, online gambling addictions, and chat programs (ex. Aim, ICQ, MSN messenger), Software piracy, pornography, and legal and ethical issues. In general, we would be looking to find out how the Internet and other forms of modern computer-based technologies affect society (Oteri).

After assembling a team of four interested parties, we were ready to change everything. We discussed many possibilities about how to administer and write a survey. We realized, from an early stage that surveying young people would be, at the very least, a necessary part of our project. This presented a difficult problem: How would we survey children about illicit activities on the Internet in a socially conscientious manner? We discussed ways of asking questions that would not expose children to the potentially objectionable material we would be asking about. When we eventually decided to avoid being specific, the project lost some depth. However, many similar discussions yielded breadth. This breadth led us to the conclusion that some data mining would be useful. At this point, before the official start of our project, we had a vision. The next step would be finding direction.

Choosing a project advisor turned out to be a simple and successful task. We needed a Computer Science professor who was socially conscious and receptive to our topic. Professor Michael Ciaraldi was known to all the members in our group as an engaging character with extensive experience relevant to our topic. He taught "Social Implications of Information Processing" (CS3043) at WPI repeatedly, and has a strong interest in the history of computers. Research that examines changing trends is well served by an advisor who has observed related phenomena in the past. In practice, Professor Ciaraldi has continually shared compatible goals and sound, timely advice with our group. Completing our team with him has proven to be a rewarding decision.

After finding an advisor and officially beginning the project, the planning that remained was a matter of refining existing ideas into working plans. Brainstorms and editing were the elements that helped us make this transformation.

Survey Design

We chose to survey 4th, 7th, and 10th graders, in addition to college freshmen, so that we could get a good cross section of children from different age groups. We intended to survey groups spanning rural, suburban and urban areas. However, circumstances and limited time made comprehensive surveying difficult, so most of our data ended up coming from suburban towns. College freshmen were never included in the survey, because of a similar set of obstacles. This is not particularly bothersome, since our goal was to do preliminary data collection. Future work that focused on bringing our survey to a wider audience would be very beneficial.

In designing our survey, we faced many challenges. We needed a survey that was concise, but also conducive to meaningful responses. Because a goal of our project is to observe how computers and the Internet affect people's development, we needed a survey appropriate for a wide range of ages. Of course, geographical location provides necessary diversity of subjects, but must not interfere with the integrity of data. These factors presented a huge challenge in writing the survey. We had to word each question to be unambiguous and concise, using a vocabulary well within the grasp of most fourth-graders. Many times, the right word to describe something would render a question incomprehensible to younger people, while verbosity would risk confusing or alienating older audiences. Careful compromise, persistent revision, and thorough testing were all important parts of the writing process.

Our survey began with a group brainstorm, yielding descriptions of everything we might possibly want to learn from the survey results. Initial revisions removed redundancy, irrelevancy, and helped us to start identifying less effective lines of questioning for removal or replacement. After some initial organization, we put these descriptions in question form. We made sure we had questions that covered general personal information, general information about computer use, and information specifically related to a hypothesis. Group discussion and revision took us far, in addition to trial runs with individual volunteers. By encouraging volunteers to take a red pen (for constructive criticism) in hand in addition to one of black, we were able to find problems that we had overlooked. This testing culminated in a trial run using an entire WPI class (CS 3043 c'03) as subjects. As in previous tests, we encouraged the takers to make comments in the margins about the effectiveness of the survey. This process provided the invaluable information that allowed us to make final revisions to the survey.

Choosing Survey Sites

A large-scale, nation-wide survey is beyond the scope of a project that had to be planned, implemented and reported on in six months. Instead, we focused on getting information that might be useful to point out possible trends. If our final conclusions serve merely as hypotheses for a larger scale survey, we will have satisfied our goals. Therefore, we set our sights on a variety of schools around

Massachusetts, hoping to get as good a cross section of the state as possible with limited means. Our original plan was to survey 13 classrooms in 7 school systems. The breakdown was to be as follows:

4th, 7th, and 10th graders at:
A rural school system,
A suburban school system,
An urban school system

Freshmen at:
A community college,
A large state university
A technical university
A liberal arts university

Our actual survey sites differed substantially from this initial outline, but we did manage to span several schools, grades and locations. We conducted surveys at the following schools (in chronological order):

Bridge Elementary School, Lexington, MA:
        One 4th grade class
Clarke Middle School, Lexington, MA:
        One 7th grade pre-algebra class
Lexington High School, Lexington, MA:
        One 10th grade History class
Burnell Elementary School, Bridgewater MA:
        One 4th grade class
Olde Rochester Regional High School, Mattapoisett, MA:
        Two math classes with a majority of 10th and 11th graders
Silver Lake Regional High School, Kingston, MA:
        Two 10th grade English classes of different levels
Silver Lake Regional Middle School, Pembroke, MA:
        Two 7th grade homerooms

See Appendix E for a map showing these survey locations.

Our public relations package

Conducting a successful survey requires much more than the survey itself. To be ready to deal with the potential survey sites we had chosen, we needed to be ready with materials that would answer questions and foster good will. For this reason, we prepared ourselves with a brochure, a standard cover letter, a website, and a logo (see Appendix B). We were careful to be consistent and professional, knowing that first impressions have lasting effects. We displayed our URL (http://www.project-404.net), and mailing list address (info@project-404.net) frequently in our PR materials to catalyze the flow of information.

Making the "sale"

In addition to being difficult and time-consuming, actually convincing the administration of various schools to allow us to survey their students was the most unpredictable part of our project. We employed many different strategies to achieve [partial] success.

We decided to survey schools that our team members had previously attended, because often, dropping the name of an alumnus would lend us credibility. Also, knowing whom to talk to was less of a problem with this strategy.

For most of the time we were setting up surveying venues, "phone tag" was the name of the game. Usually, we would get referred to the right person quickly, and then be delayed by answering machines, lack of cooperation, or finding that there was more preparation to do. Almost all of the school systems needed to have the survey authorized at some higher level, but no further generalizations can be made. We made good use of our PR package when we needed to, and learned the importance of being aggressive. We learned this lesson late, but it helped us much near the end. Towards the end of the time that we were surveying schools, we improved at calling schools more often, but we also tried to meet with administrators and individual teachers in person as much as possible instead of relying solely on telephone communications. In some cases, appearing in person to speak with an individual teacher was more productive than calling ahead of time.

Conducting the survey

When we actually arrived in the classrooms with blank surveys in hand, we were usually met with blank stares from students who did not know exactly who we were or what to expect. We began by telling them what school we were from, briefly explaining the intent of our project, and reading the paragraph on the cover of our survey aloud (see Appendix A). We offered to answer their questions at any time, answered any that occurred immediately, and then handed out the surveys. Typically there were a few questions. The 4th graders inquired about the meaning of survey questions with some frequency, but not so often as to cause us concern. We routinely heard: "What does 'online' mean?", referring to the question that asked how much time they spent online on the last page of the survey (see Appendix A). Since this survey question was intended to investigate their perception of the word, we consistently replied with "Whatever you think it means". However, some of the questions we faced were more difficult. Many students did not know how to characterize their parent's professions, so we often had to provide specific suggestions. This may have introduced some inconsistency into the survey, but we kept it close to the necessary minimum with replies that were consistent. If a question came up more than once, we would address the answer to the entire class, which should have helped with consistency. Every class we surveyed behaved very differently before and during the survey, but when we thanked them and distributed candy at the end, nobody complained. The key elements for success in actually proctoring our survey were politeness, succinctness, consistency, and gratitude.

Data entry

Some discussion has arisen about whether this step could be avoided by putting the survey in digital form to begin with. We felt that because our survey dealt with technology, this could significantly skew our results. The people, who consider "AIM" (AOL Instant Messenger) to be their operating system of choice may not be capable of filling out a web form as well as someone who is computer literate. We decided that if the survey was given in digital form, there would be no way to level the playing field, short of supplying each student with a ready-to-go laptop computer. Our final decision was to give our survey in paper form, and expedite the data entry process as much as possible. This decision led us to the Common Gateway Interface (CGI).

We used a simple web form (see Appendix A), coupled with a Perl program (dataent.cgi in Appendix C) using CGI, to output comma separated value (CSV) files. We then manually entered data as fast and as accurately as possible. There is little to say about this task, as it amounts to little more than unskilled, manual labor, but in rare cases, it was necessary to use some of our own judgment. In almost all cases, when a survey question was answered as "other", it was left that way, but there were occasions when this was not the right solution. For example, if a student wrote "myself" in the blank next to "other" to answer the question about what they used e-mail for, we would enter it as "personal", since the synonymous option already existed. Exerting this sort of judgment during data entry happened rarely, and was discussed by at least two people before being finalized. We did this conservatively, so it would not amount to tampering, but some of this was necessary to allow the actual meaning of the data to reflect in our results.

Statistical analysis

Analyzing our data proved to be a large, multi-faceted task with many intricacies. We had three main goals to accomplish. The first goal we had was to support or refute our existing hypotheses. This goal was achieved by using the right software to analyze and display our data in revealing ways. This task was made easier by putting the right questions in our survey (see Procedure: survey design). The task was quite practicable because we were armed with the right data, the right software, and we knew exactly what data to look at

Our second goal was to find conclusions that we had not anticipated with hypotheses using data mining techniques, a nontrivial task. One of the pieces of software we used for this was our own Perl program, "miner" (see appendix C for source), which was used to find and count correlations in our data. In addition, we wrote some command-line utilities to help us make sense of miner's results and find the important parts (see Appendix C). Identifying correlations that implied a significant conclusion was a task in and of itself, but one that yielded some interesting ideas. At this point, our data mining pointed to specific pieces of data that were important, but not expressed in our hypotheses.

The third goal we had that falls within the realm of statistical analysis was to find ways to organize and display our data in a way that made sense and implied meaningful conclusions. Often, advanced tools such as SAS were needed to identify compelling trends in our results, but sometimes a simple spreadsheet application was sufficient. The right tool for this task was a function of the nature of the data, what we were trying to get out of it, and how we wanted to display it. We have attempted to illuminate the important parts of our results in this fashion (see Results and Conclusions).

A note on software

Finding the right software to assist our statistical analysis efforts was a challenge. At first, we were directed towards SAS. SAS has proved to be extremely useful for analyzing data, but we found it to be incomplete for the tasks we had in mind. We implemented our own data mining algorithm in miner (see Appendix C). This took us a long way towards finding more conclusions and analyzing existing hypotheses. The command-line utilities we wrote, such as nix, kr, NIX, num2pr, and genex were essential to dealing with the results miner generated. These utilities helped select important data and format it in ways more useful to humans that comma-separated integers (See Appendix C for details and source code). Existing UNIX utilities like grep, head, and tail were also very helpful. The point of miner was to direct our attention to important details, so that they could be further analyzed. This process left only one gap; displaying results. Sometimes SAS did not measure up in this regard, leaving us without necessary graph archetype options, so Microsoft Excel was used to generate graphs.

The final report

The creation of this document is a straightforward task. We have done what we needed to for the duration of the project, and are left to merely record it. Our goal is to be complete, concise and correct. We divided the task among ourselves, and are proceeding, stopping occasionally to confer, or exchange pieces for editing.

Issues and Problems

Our team encountered several obstacles while completing Project 404. The most detrimental of these was a problem with version control. Due to lacking strictness about file centralization, a preliminary version of our survey was used for all 3 classrooms we surveyed in Lexington. Fortunately, this error affected only two questions on the survey (questions 25 and 26). The two questions that asked about one's attitude towards technology in general, appeared with the wrong options (see Appendix A for details). Since we gave different versions to different schools, our data from these two questions was inconsistent, which prevented us from using it later.

Unfortunately, three more questions found no utility. Questions 5, 6, and 7, about religion, political preference and income respectively, were not well received in the schools we surveyed. To satisfy the policies of these schools, we asked students not to answer these questions. Even though one school did allow the students to answer these questions, using that data gave us little information without data from more schools. We did not use any data from these three questions in our analysis.

Another issue we encountered along the way was getting surveys printed. Student discount rates at one printing outfit were discontinued during our project. After finding a great price for printing surveys, we had to waste time repeating the process. We recommend that other people with similar printing needs for surveys get all printing done initially.

In many of the questions on our survey, varied perception is an important part of collecting meaningful data. However, in some cases, it worked against us. Many students, especially fourth graders, seemed to be unsure about the meaning of several survey questions. We responded to many of their inquiries, but it is likely that unchecked confusion introduced some error into our results. However, we have every reason to believe that this error is within acceptable limits. For the most part, the fourth graders we surveyed did seem to comprehend the survey we gave them, and the data they provided seems meaningful and clear.

Results

Raw Data
Results from Datamining
Data Mining Results

Raw Data

The following section of our report contains the number of occurrences for each option on every question in the survey. We were able to collect a total of 203 surveys from 9 classrooms in 4 different towns.

Age Distribution

The first set of data is a distribution analysis of the ages of our subjects. This shows the spread of ages for all of the survey takers. Figure 1 is of a box and whiskers plot. This illustrates the overall range from age 8 to 18, and where the majority of the data lies in relation to that range. The large section of the plot is the range from the 25% quartile at age 12 to the 75% quartile at age 16. This area represents the middle 50% of the data. The line at age 13 represents the mean age of all subjects. Since the majority of the data is to the right of the mean age, it shows that the data is slightly left skewed. In other words, the main mode of the distribution is slightly to the right of the mean.
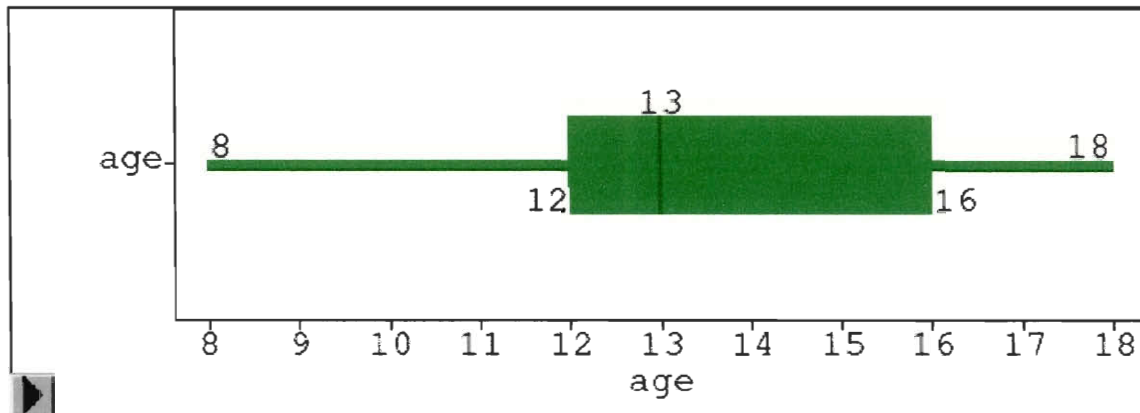
Figure 1: Box and Whiskers Plot.

This chart illustrates the number of students at each age. This graph shows three distinct modes in the ages. The modes correspond to the three different grades of student of which we surveyed. We surveyed slightly more people in the higher grades than the lower grades.

Age Distribution
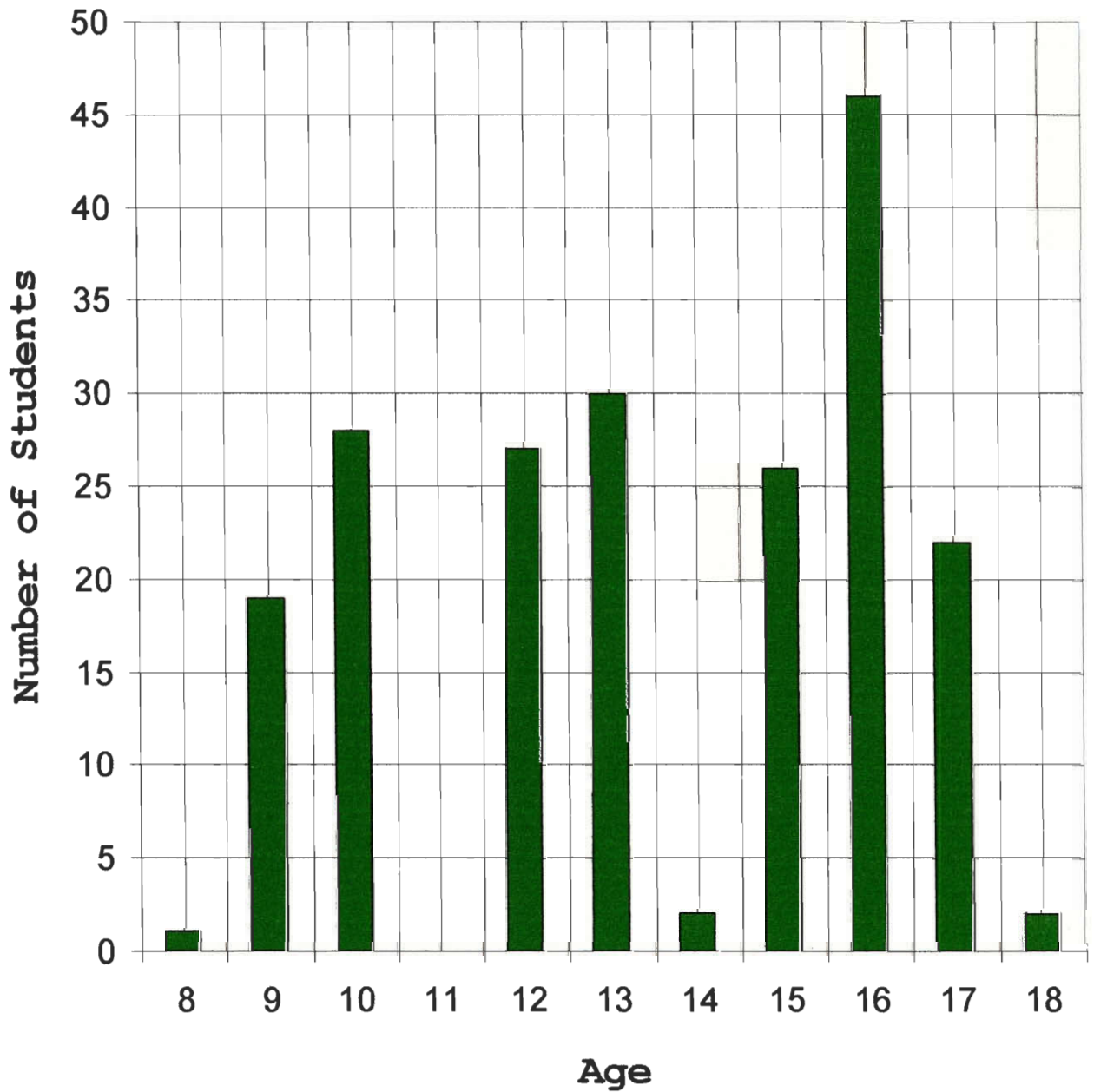
Parental Occupations or Activities

The next chart illustrates the spread of occupations held by the parents of our subjects. It shows a majority of parents are currently involved in the fields of engineering and business, and a very limited set of parents holding jobs pertaining to data entry, arts, government, and physical labor. The following chart represents the many fields that parents are involved in.

One of your parents or guardians is involved in...
(Check all that apply) (as a job or otherwise)
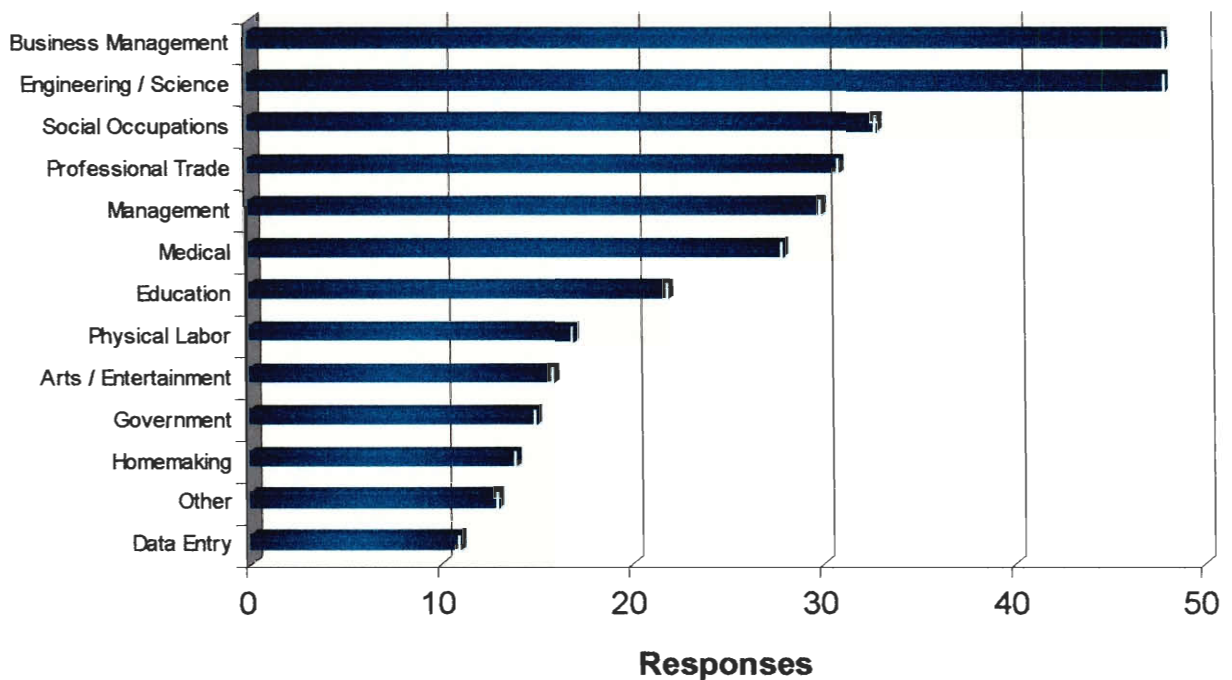
Responses

Number of responses out of 203 surveys. Categories are not mutually exclusive.

Academic Subjects

What is your favorite academic subject in school?
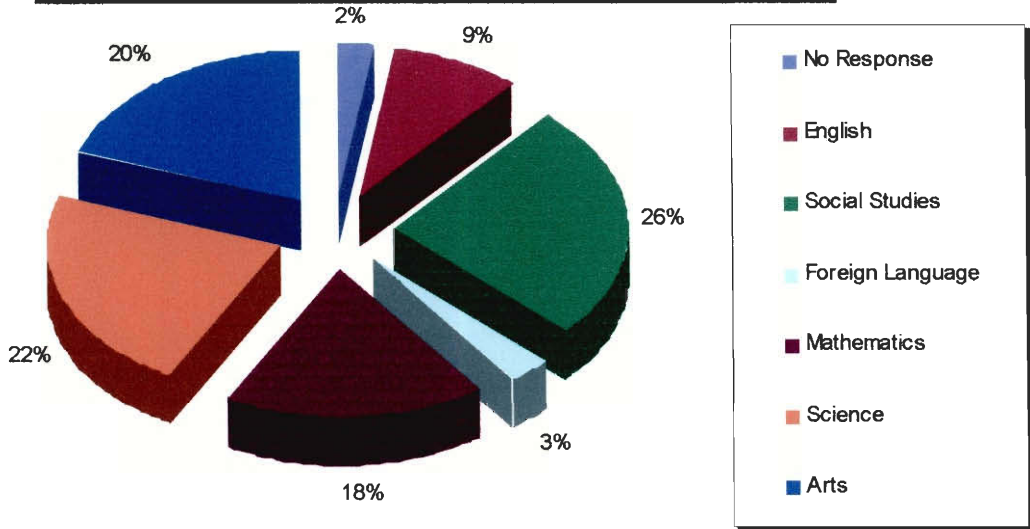
- No Response
- English
- Social Studies
- Foreign Language
- Mathematics
- Science
- Arts

2%
9%
20%
26%
22%
3%
18%

What is your least favorite academic subject?

- No Response
- English
- Social Studies
- Foreign Language
- Mathematics
- Science
- Arts

3%
3%
12%
18%
31%
12%
21%
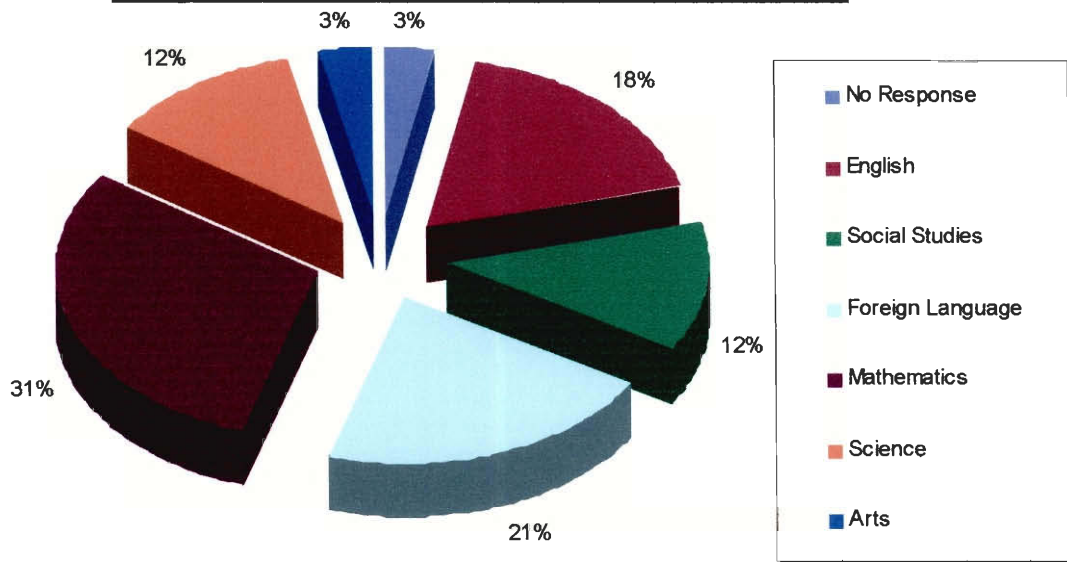
See Appendix A for details about each survey question.

Town Size

Our survey data came from a variety of different areas. We were able to survey nine classrooms in four schools. Two of the schools that we surveyed were regionalized: Olde Rochester Regional High School serves three towns, while Silver Lake Regional Middle School serves four. Most of the schools with the exception of Olde Rochester Regional High School were located in suburban towns, but because of the regionalization of some of our schools, it was possible that some of the students live in rural and urban areas.

The area that you live in is...

2% 0%

34%

64%

- No Response
- Rural
- Suburban
- Urban
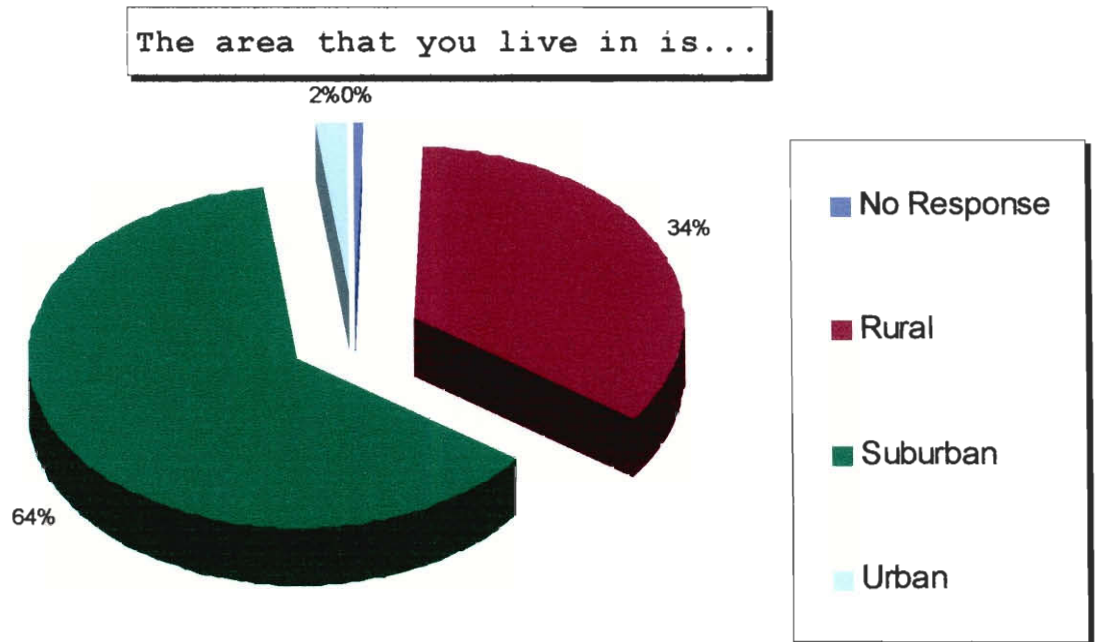
See Appendix A for details about each survey question.

Length of Residence

The next question in our survey is related to how long our subjects have lived in the same home. An interesting fact that the data shows us is that 18% of students have moved rather recently.

How long have you lived in your current home?

0%   5%

13%

31%

No Response

Less than 1 Year

1 - 3 Years

More than 3 Years

Lifetime

51%
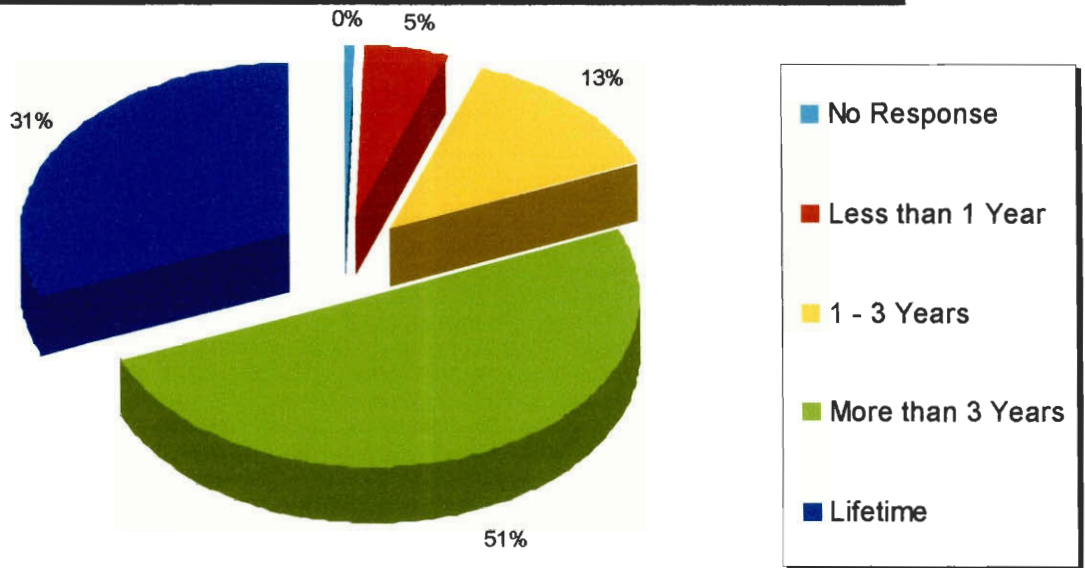
See Appendix A for details about each survey question.

Age of First Computer Use

One of the most striking observations that can be made from the results of this question was that only 1% of the people whom we surveyed do not use computers. Another observation was that around 1/5 of the students began using computers under the age of 5.

How old were you when you first started using computers?

0% 1% 1%
15%
21%
62%

- No Response
- Under 5
- 6 to 10
- 11 to 15
- Over 15
- No Computer Use
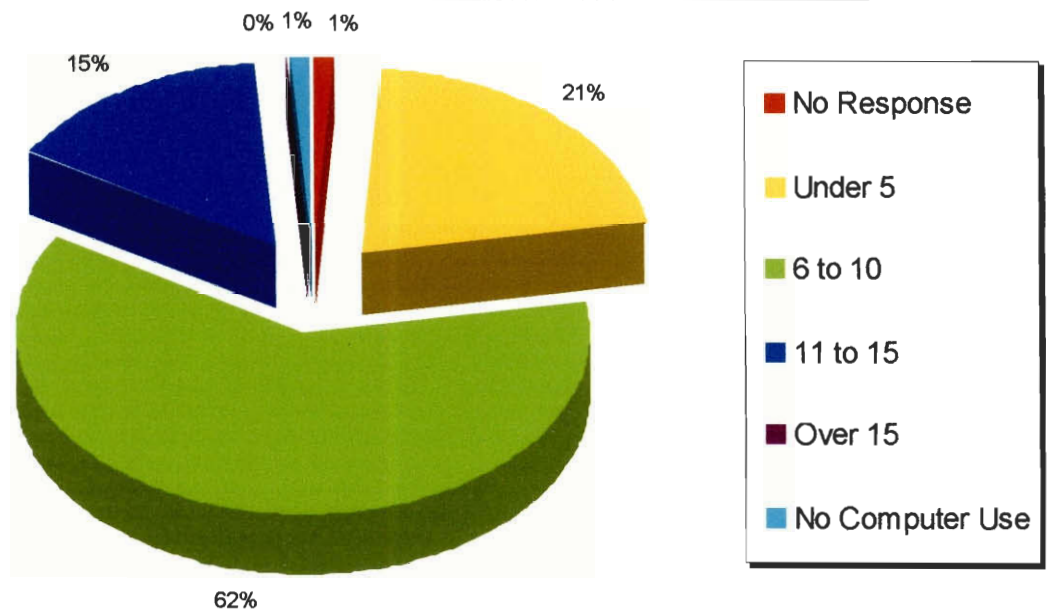
See Appendix A for details about each survey question.

Primary Access to Computers

A very interesting observation that is clearly illustrated in the following chart is that 1% of the students surveyed claimed to not have any access to computers. This represents the possibility that there are some people who are unaware of public computer terminals located in many libraries and schools.

How do you primarily access a computer?

0% 1%

1% — 0%

35%

63%

- ■ No Response
- ■ Own Computer
- ■ Family Computer
- ■ School / Library
- ■ Other
- ■ None
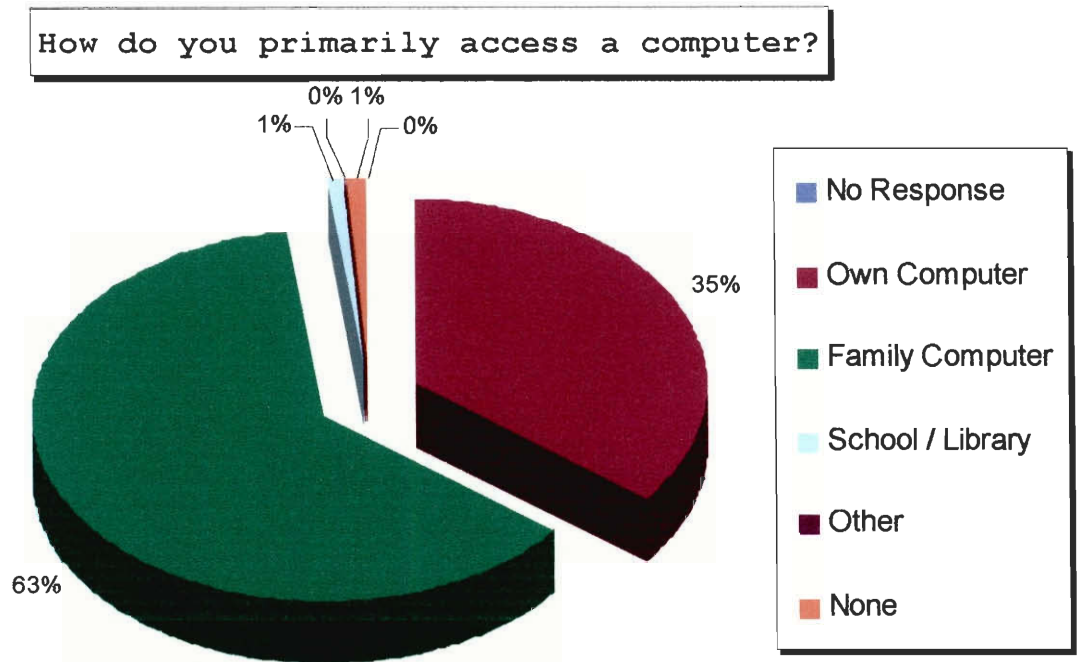
See Appendix A for details about each survey question.

Number of Computers in the Household

A majority of households seem to have one or two computers while 1% of families do not own a computer. An interesting fact shown in the following pie chart is that nearly 1/3 of the households in our survey audience have three or more computer systems.

How many computers are in your home?
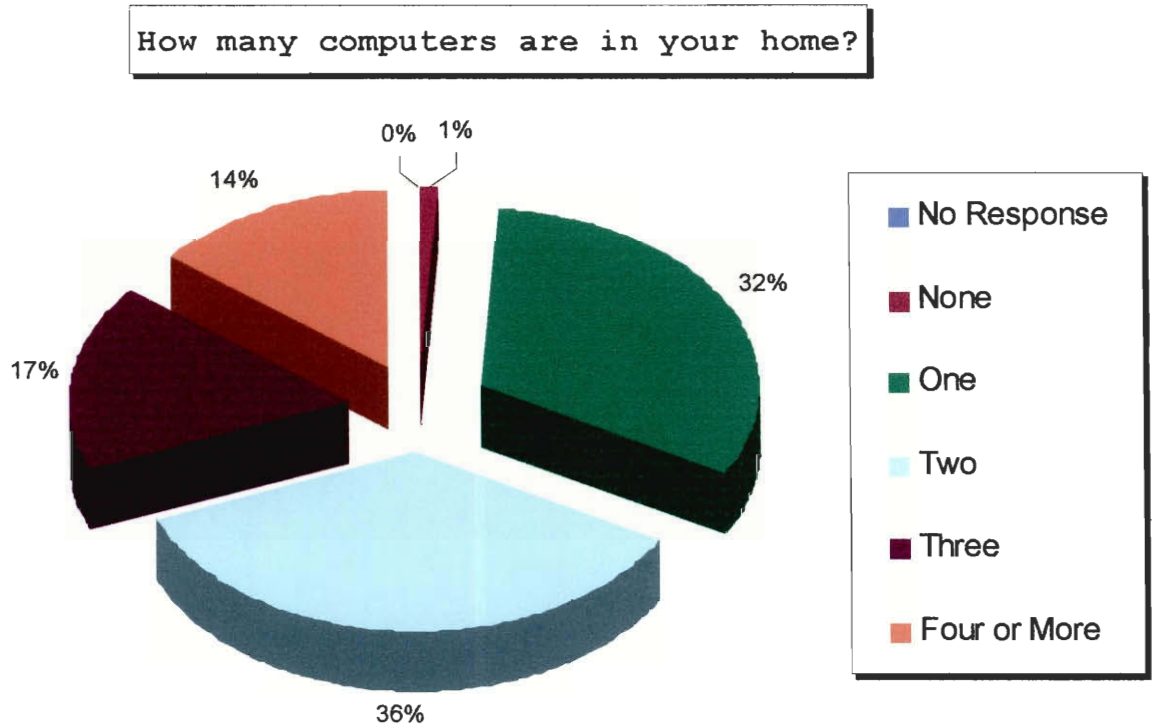
See Appendix A for details about each survey question.

Age of Newest Computer

One interesting observation from our data on this question was that forty-five percent of households have a computer system less than 18 months old. Nearly half of our subjects are using older hardware.

How old is the newest computer in your home?

1% 2% 1%

13%

25%

38%

20%

- No Response
- Less than 6 months
- 6 to 18 months
- 18 to 36 months
- 3 to 6 years
- Older than 6 years
- N/A
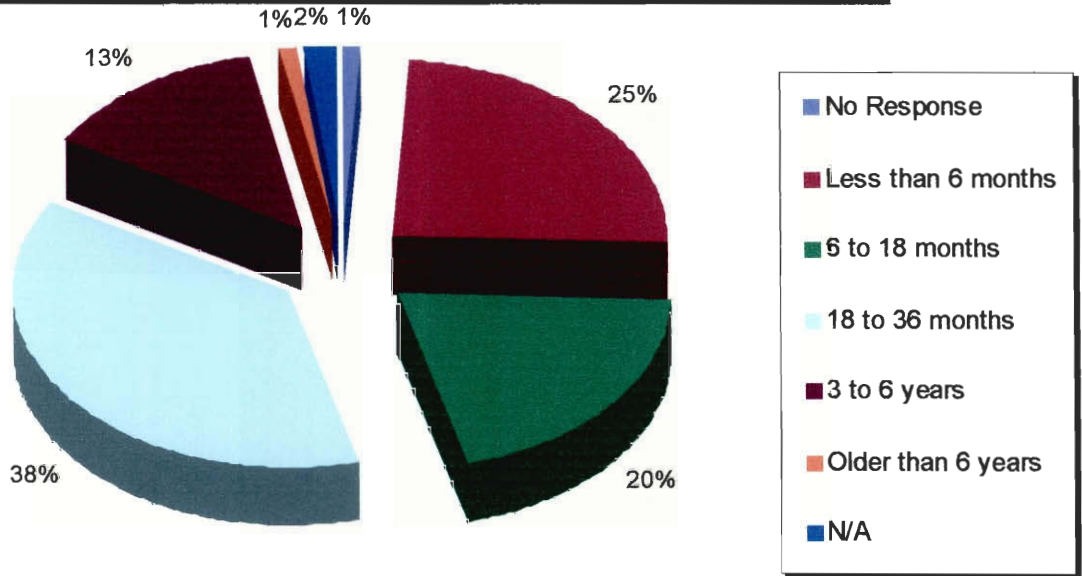
See Appendix A for details about each survey question.

Maximum Hours per Week of Computer Use

Another interesting observation from our data is that nearly 75% of the students surveyed use a computer for less than 2 hours per day, or 15 hours per week. Almost half use a computer for less than 45 minutes per day.

How many hours (maximum) per week do you spend using a computer?
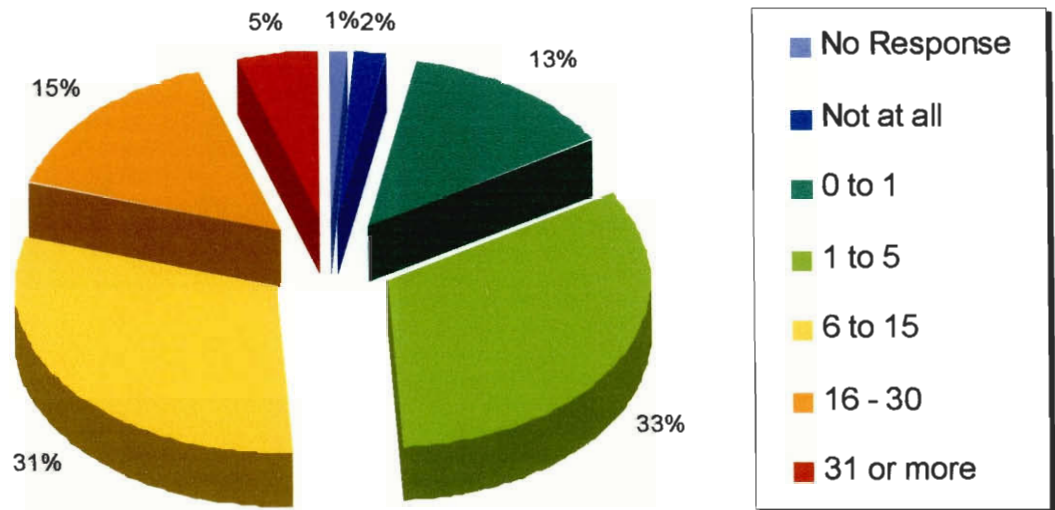
See Appendix A for details about each survey question.

Operating Systems

Microsoft Windows is by far the most common operating system on the market, but what is especially interesting to note is that Mac OS X and Mac OS 9 make up such a small portion of the responses, that they almost are non-existent amongst young students. Another interesting note that we should mention is that all of the people who answered "other" for this question, named particular pieces of software such as AOL or Instant Messenger, or manufacturers of PC's such as Dell, as their operating system.

What operating system do you use on the computer that you use the most?

Legend:
- No Response
- Windows NT / 2000 / XP
- Mac OS X
- UNIX
- Windows 95 / 98 / ME
- Mac OS 9 or Earlier
- Other

3% 4% 5% 38% 48% 0% 2%
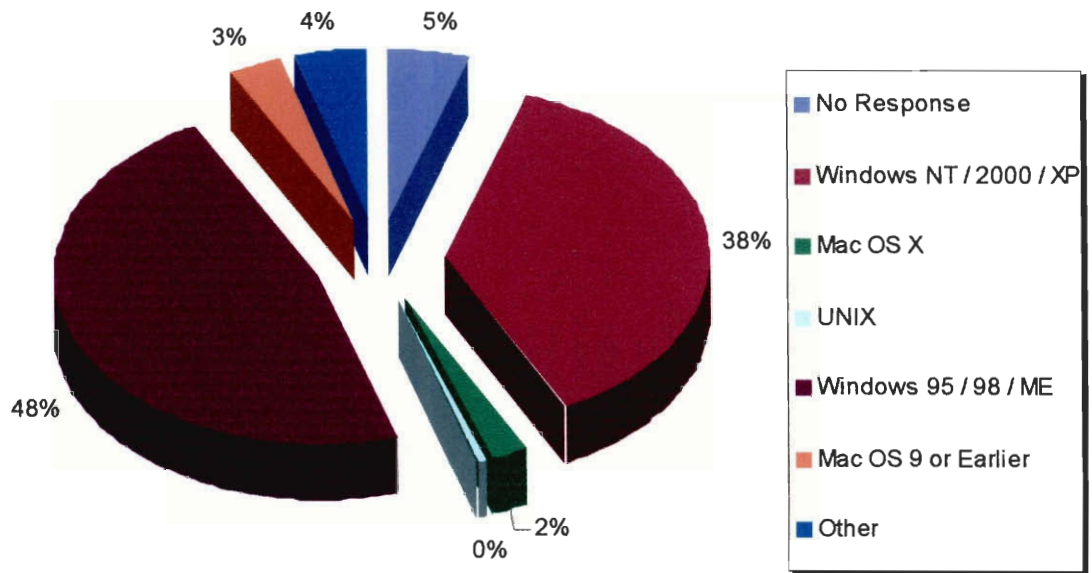
See Appendix A for details about each survey question.

Type of Computer

This next chart illustrates the percentages of people who own brand name computers, compared to those who have custom computer systems. It seems that nearly 85% of people have brand name machines, while only 13% built their own.

Is the computer you use the most a...

13%   1%1%   11%

No Response

Laptop

Brand-Name
Desktop

Home-made
Desktop

Other

74%
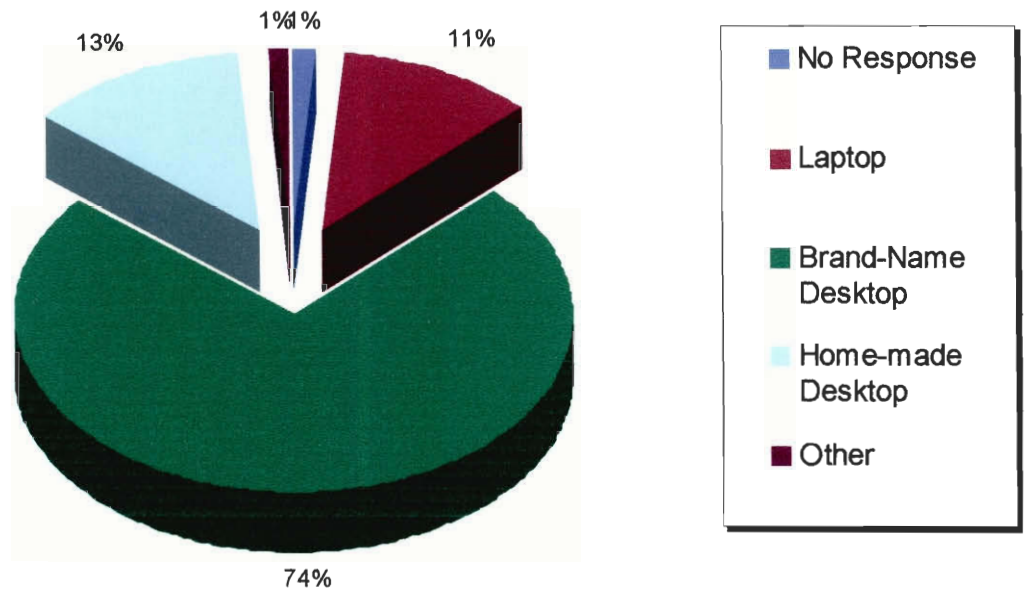
See Appendix A for details about each survey question.

Percentage of Students with E-Mail Accounts

Do you have your own e-mail account?

1%

21%

78%

No Response

Yes

No
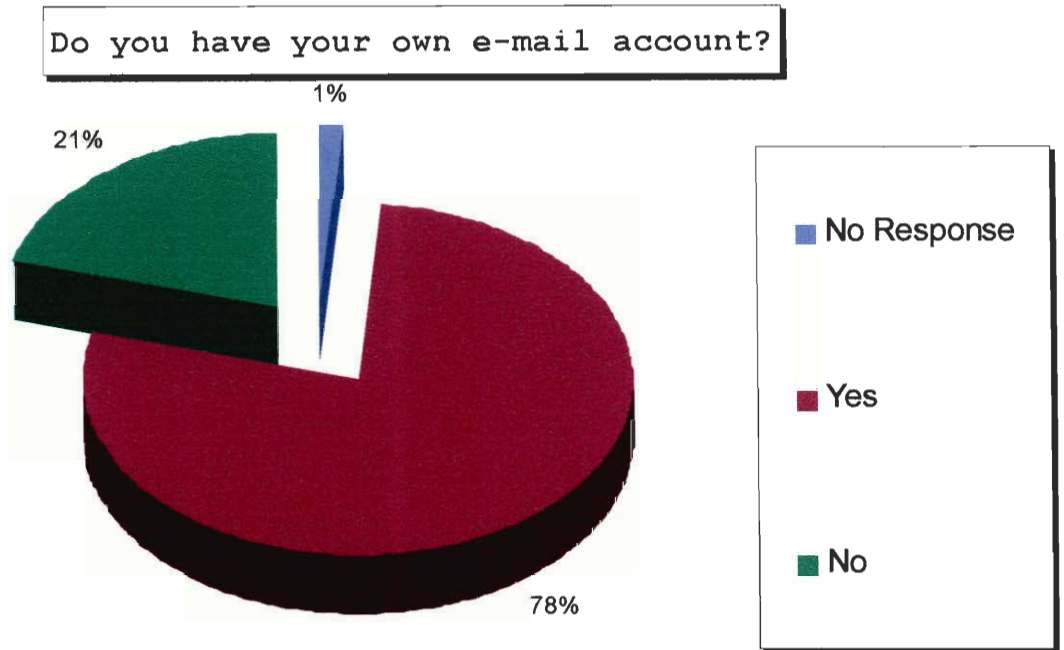
See Appendix A for details about each survey question.

Primary Computer Used to Access the Internet

This next chart shows the breakdown of what computer students primarily use to access the Internet. An interesting observation that can be made is that only 3% of students did not have any access to the Internet.
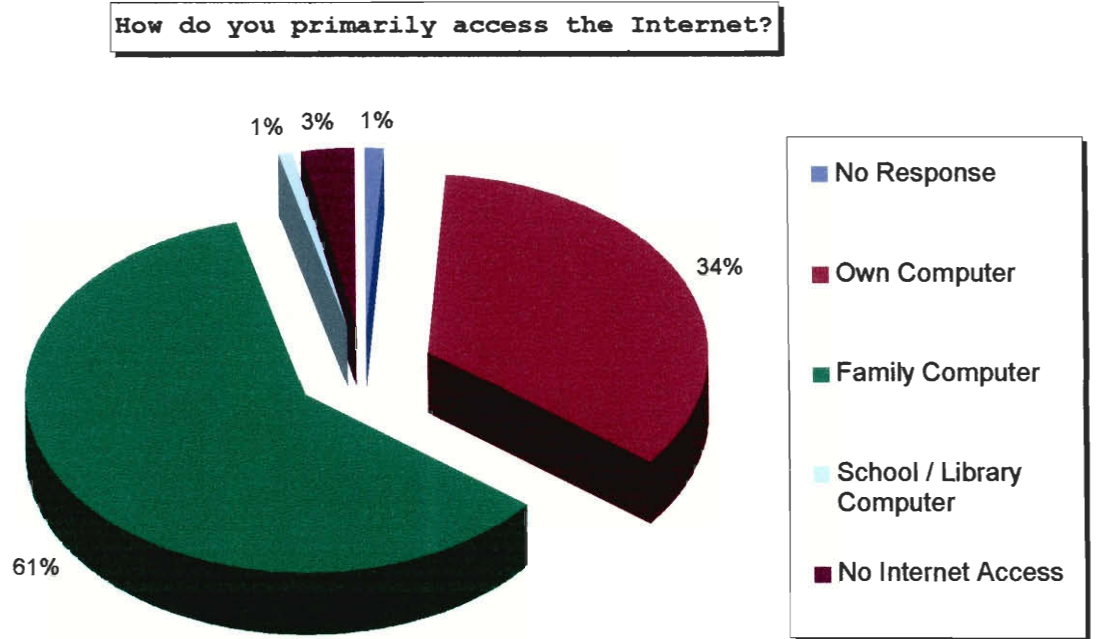
See Appendix A for details about each survey question.

Internet Connection Type

The following chart illustrates the different types of Internet connections by which students use to get online. "Broadband" refers to Cable Modem or Digital Subscriber Line (DSL), whereas "High Speed Network" refers to a LAN connected to the Internet with a T1 or T3. Because our survey population did not include any college students, the resulting percentage of students on a high speed network is much smaller than we would have expected.

What kind of Internet connection do you use?

- 2%
- 27%
- 33%
- 1%
- 3%
- 34%

Legend:
- No Response
- Dial-Up
- High Speed Network
- Broadband
- Other
- Don't Know
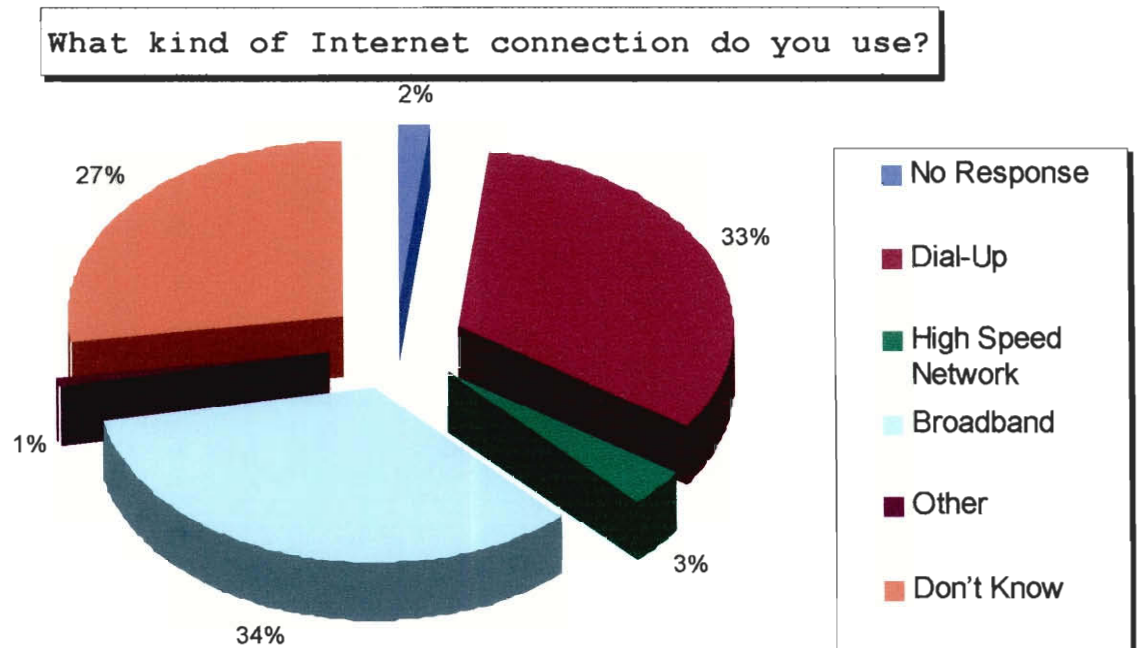
See Appendix A for details about each survey question.

Computer Related Activities

For most of the activities listed below, the majority of the responses indicated 1 – 5 hours per week, while surprisingly, the overwhelming majority of responses for watching videos using a computer were "none."

How many hours (maximum) per week do you spend doing the following using a computer?
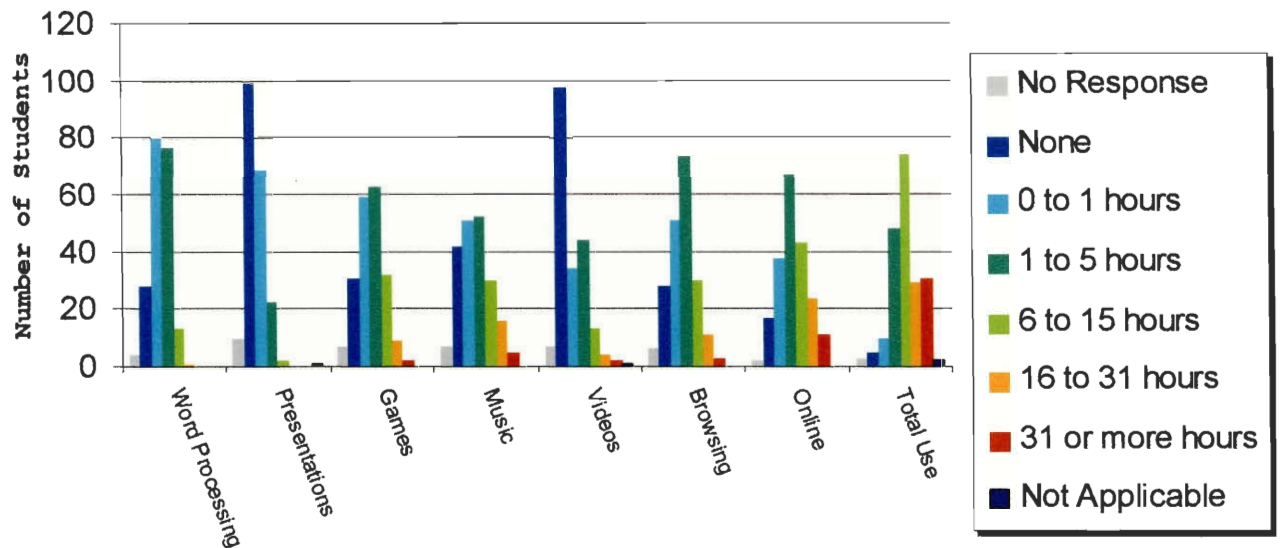
See Appendix A for details about each survey question.

Non-Computer Related Activities

The following graph illustrates the number of hours per week that students did each of the options listed below without the use of a computer.

How many hours per week (maximum) do you spend doing the following without the use of a computer or the Internet?
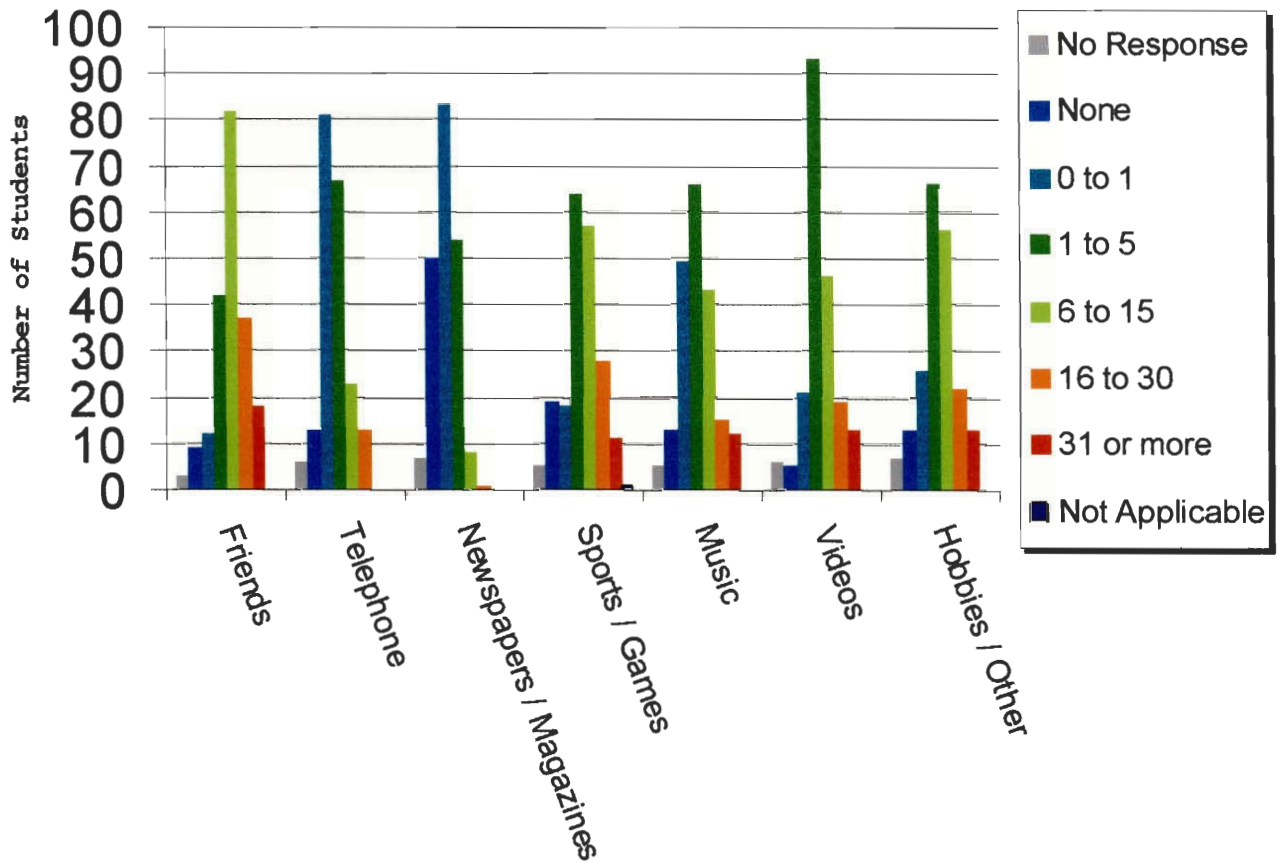
See Appendix A for details about each survey question.

Use of E-Mail

Not surprisingly our data suggests that for the majority of students who use e-mail, they primarily have it for personal use. This question was "Check all that apply," so responses were not mutually exclusive.

Do you use e-mail for (check all that apply) :

Number of Students	
180	
160	
140	
120	
100	
80	
60	
40	
20	
0	

Personal   Education   Business   Other
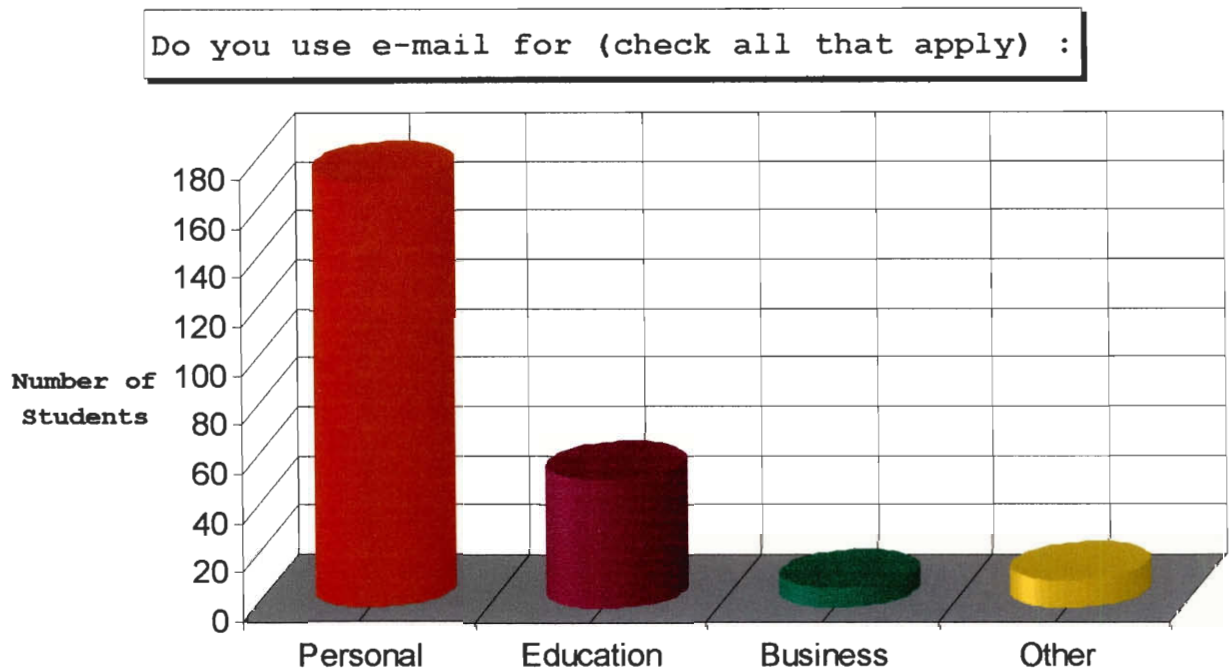
See Appendix A for details about each survey question.

Number of Unsolicited E-Mails per Week

An interesting observation that can be made from our data on this question was that nearly half of the students whom we surveyed receive less than 10 pieces of unsolicited e-mail per week.

How many pieces of unwanted / unsolicited / junk e-mail do you get per week in your primary account?

- No Response
- None
- 1 to 10
- 11 to 40
- 40 to 100
- 100 to 500
- More than 500
- Does Not Apply

3%, 5%, 0%, 7%, 20%, 18%, 28%, 19%
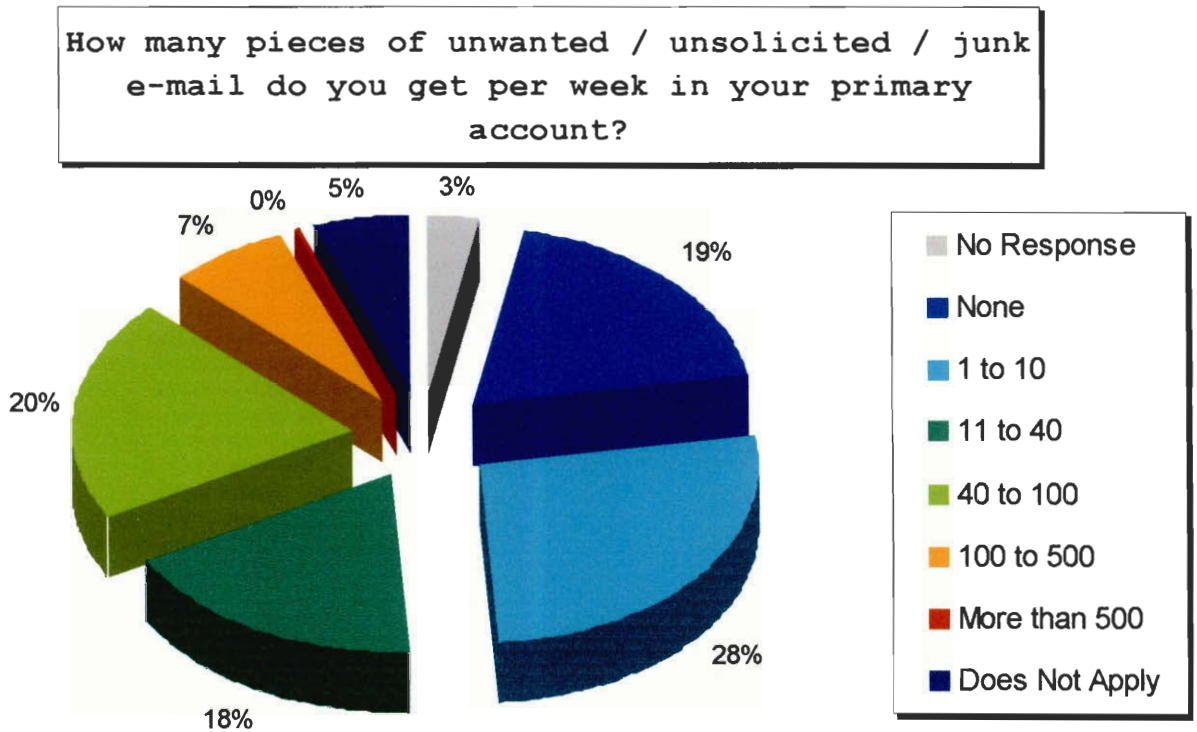
See Appendix A for details about each survey question.

Household Rules

The following chart shows the rules in our subjects' homes regarding computer and Internet use. Surprisingly almost half of the households did not have rules in place. The most common of the rules was what sites a person can visit, followed closely by how much time he or she can spend online. Very few parents have rules about what games their children can play online, and only a few more have rules about who their children can and cannot talk to. The types of rules were not mutually exclusive, but the option for "No Rules" was mutually exclusive from the rest.

Are there rules in your household regarding your internet use?

Number of Responses
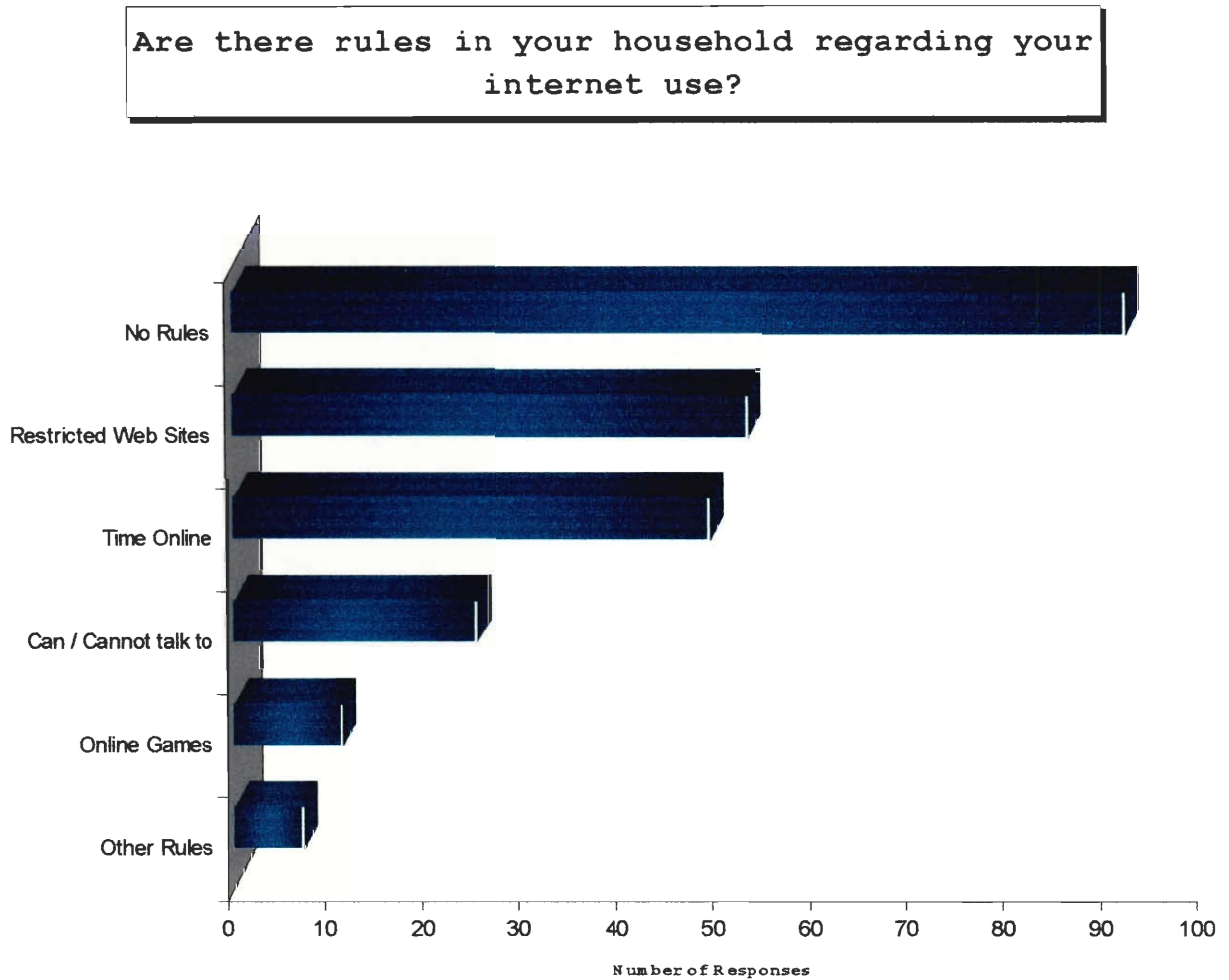
See Appendix A for details about each survey question.

Discussion of Acceptable Internet Use

We have found that 58% of parents talked to their children about acceptable Internet use.

Has a parent or guardian ever discussed acceptable Internet use with you?
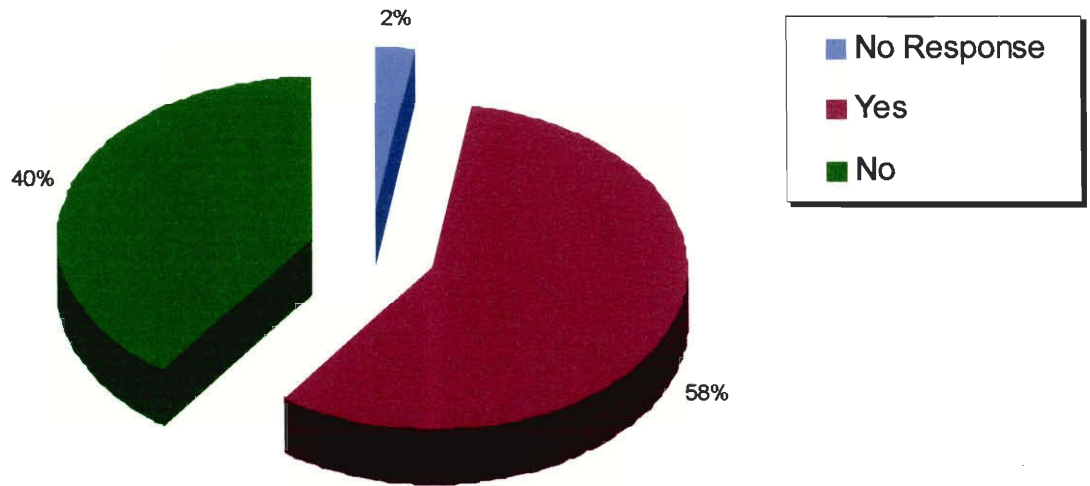
See Appendix A for details about each survey question.

Parents' Computer Usage

This chart illustrates the amount of time that parents use a computer. It shows that the largest distinct group of parents use a computer for one to five hours per week. An interesting observation from the data shows that 4% of parents do not use computers.

> Decide which of your parents or guardians uses a computer the most. Hom many hours per week does he / she use a computer (or best estimate)?

See Appendix A for details about each survey question.

43
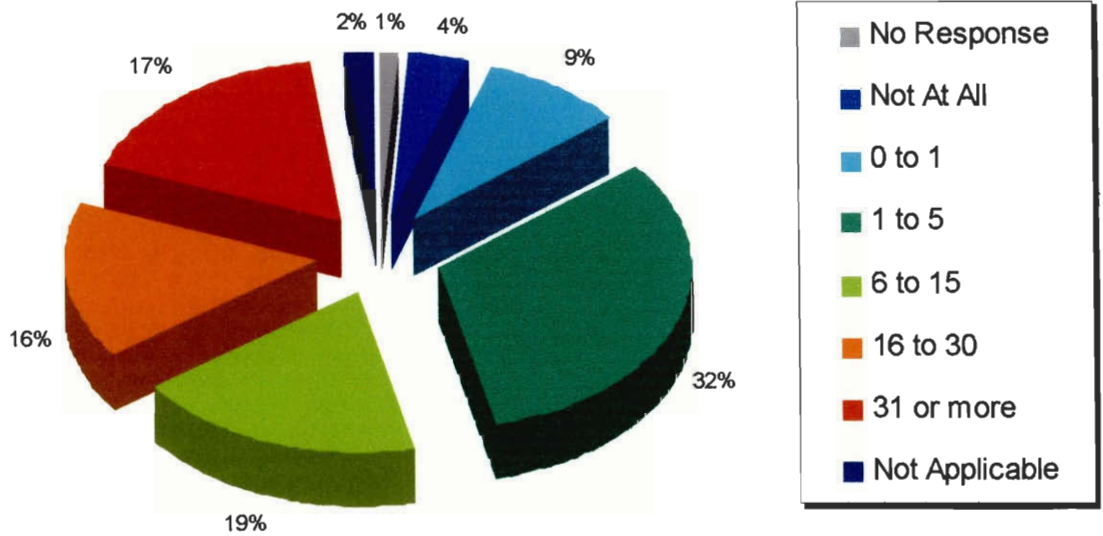
Computer Knowledge and Influences

Who knows the most about computers in your household?
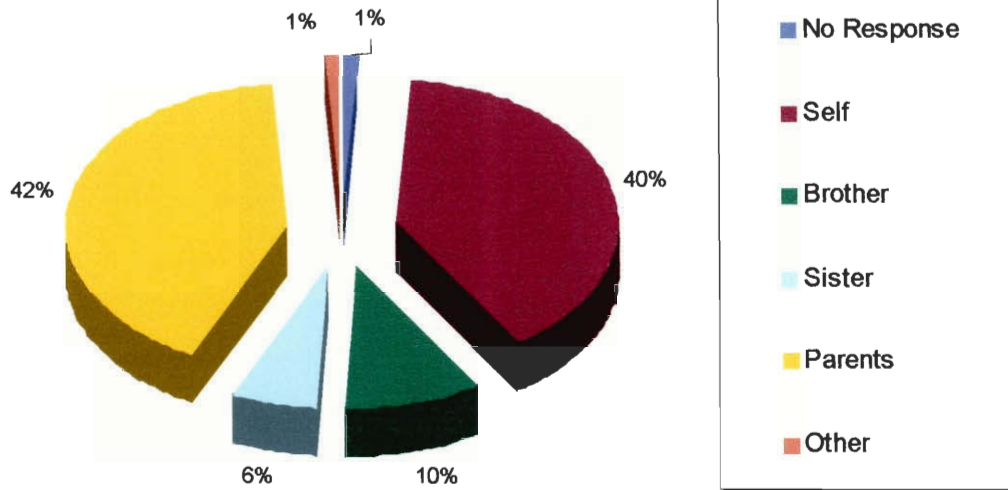
Legend: No Response, Self, Brother, Sister, Parents, Other

1% 1% 42% 40% 6% 10%

Who has helped you the most with computers?

Legend: No Response, Teachers, Parents, Friends, Siblings, Internet Resources / Books, Other

4% 2% 7% 8% 14% 42% 23%

See Appendix A for details about each survey question.

44
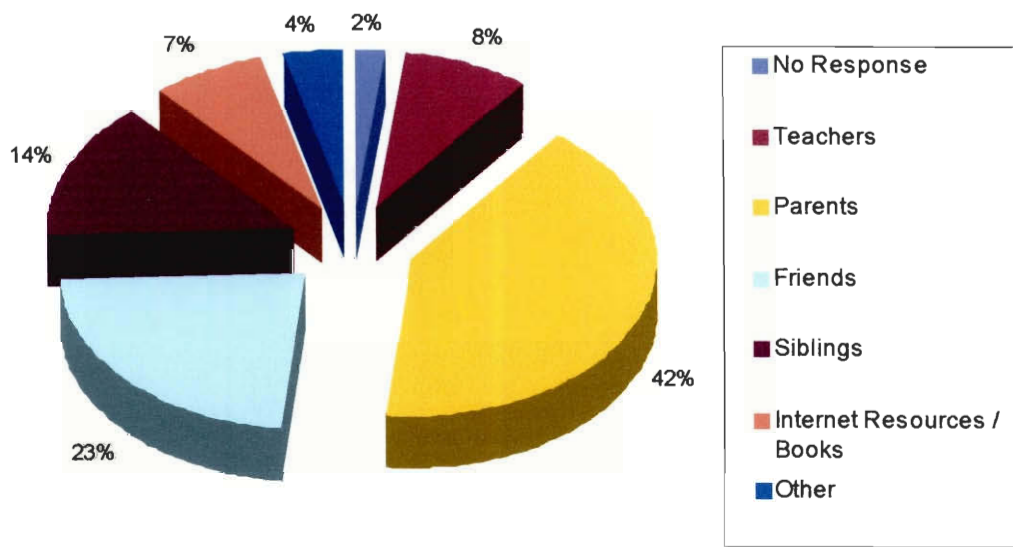
Results from Hypotheses

The right tools for the job

There were mainly two different approaches made to extract correlations from our data set. The first being a collection of custom built Perl scripts which we call miner and the second being a UNIX based statistical analysis package called SAS/Insight.

Originally our intent was to rely solely on the SAS package, but after many attempts at using the software we were unable to learn the interface well enough to have it find correlations in the data for us. We then decided to devote two of our team members and about one week of our time to write some Perl scripts that would search for correlations and remove any unusable or insignificant results for us. This set of scripting tools, referred to as miner, would output raw correlations, but did not give us a clean visual, so we decided to incorporate SAS back into the picture to run alongside miner.

miner proved to be useful in that it would give us ideas which we hadn't thought of and point us in the right direction, while SAS allowed us to explore further into all correlations which we believed were significant.

I'm not getting any younger – Grade vs. Total Time Online

Another hypothesis that we were able to find supporting evidence for was that children in younger grades tend to spend less time using the Internet then do children in older grades. This seems like a logical assumption because as children advance in grade school, their school related work gradually starts to require more and more research, and the World Wide Web has proven to be a useful source of information.

Through our data mining and statistical analysis software, we were able to determine that though the modes were very similar, in the older grades, there were many more extreme cases of people spending 31 or more hours online. The following chart illustrates this point.
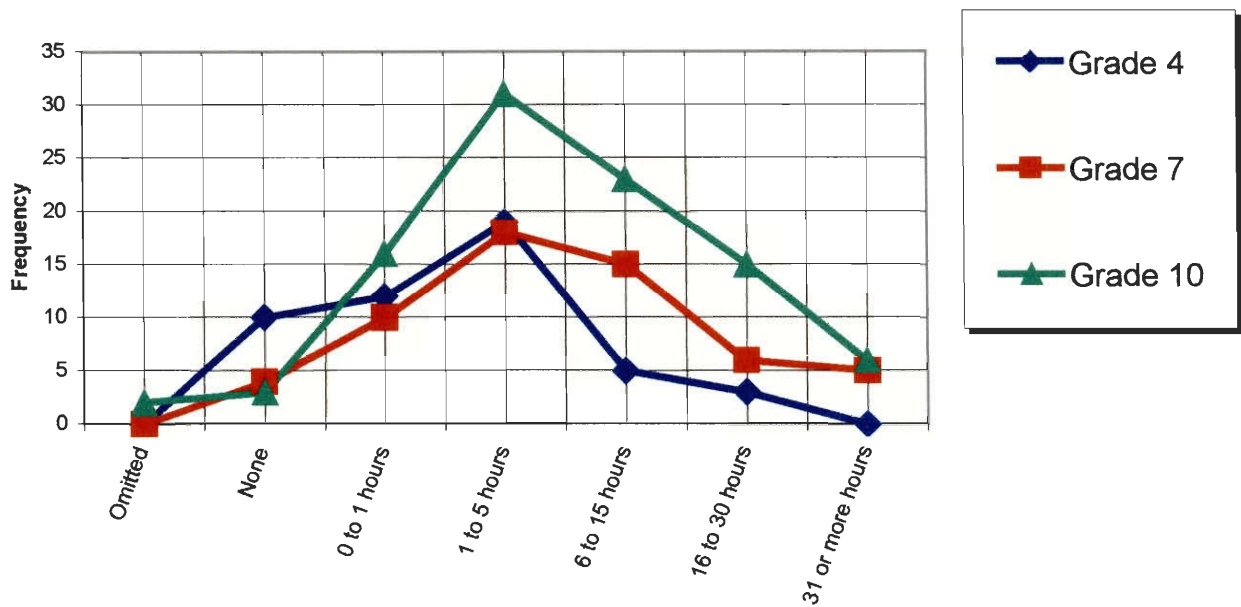
The battle of the sexes – Gender vs. Household Rules

Throughout history, the female sex has been portrayed as weaker. Any truths behind this theory are beyond the scope of this project, but because of this common perception, we have reason to suspect that young girls are disproportionately sheltered by their parents. Since our culture expects boys to be mischievous, it is likely that boys are more restricted by their parents in which websites they are allowed to visit, whereas girls are more restricted in whom they are allowed to talk to on the Internet.

Traditional expectations about gender roles seem to have had the expected effect on parental regulation of children's Internet use. The tendency is for girls to face more rules about whom they are allowed to talk to, whereas boys face more rules about the content they access, the time they spend on the computer and games that they play.

The following chart represents the actual data that we collected. Many more females fall into the category of having rules about whom they can talk to than males do, and there are many males who have rules about what websites they are allowed to visit than there are females. The most surprising comparison we found was that many more females had rules about how much time they can spend online than did males, but all in all, the resulting data does a thorough job of illustrating our hypothesis.

Rules vs. Gender

Other Rules
Who they can / cannot talk to
Online Games
Restricted Websites
Time Online

Female
Male

0   5   10   15   20   25   30

Number of Students
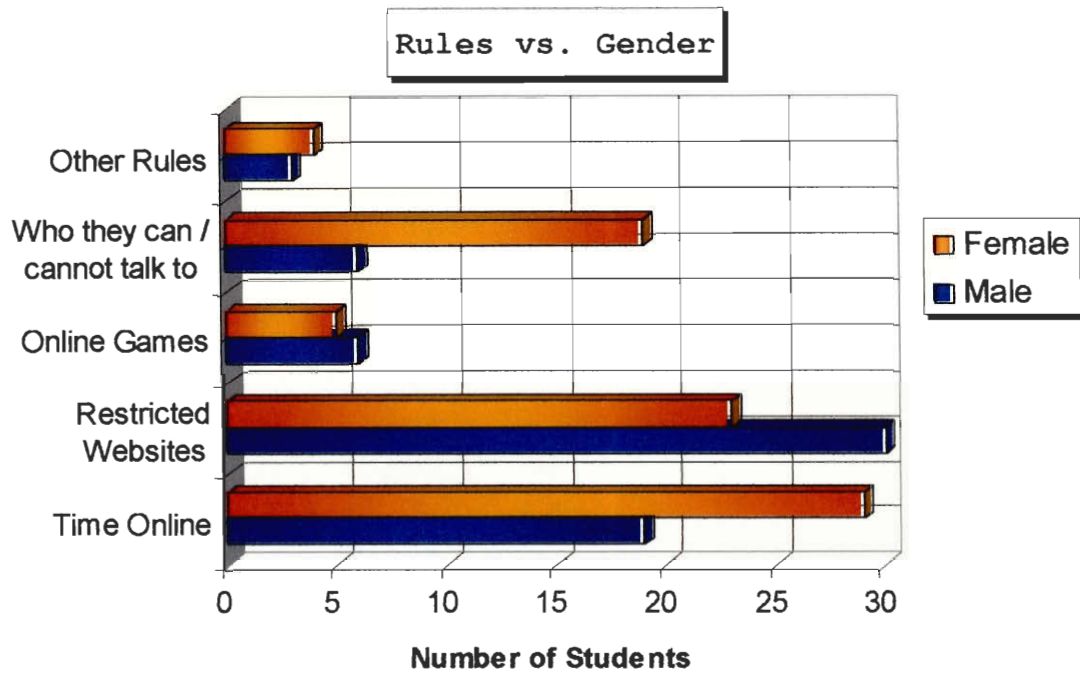
Another interesting correlation that we would like to illustrate is the difference in the genders with respect to parents or guardians discussing acceptable Internet use with them. Our data shows that males were more likely to have had a parent or guardian talk with them about proper use of computers and the Internet.

Has a parent or guardian ever discussed acceptable Internet use with you?

Female

Male

Yes
No
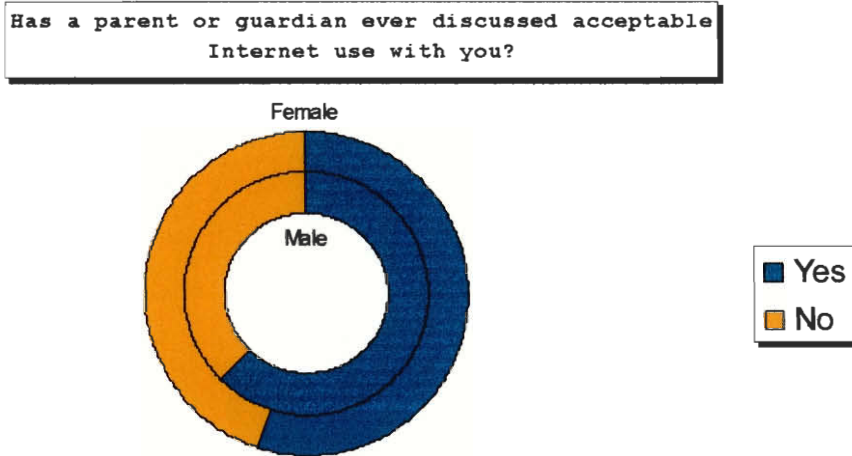
In this chart the inner circle represents the males and the outer circle represents the females. The blue is the percentage that responded with yes, and the orange is the percentage that replied with no. It shows the difference in the percentages between the two genders.

Can I play daddy? Age vs. Household Rules

One of our hypotheses was that age is inversely proportional to the extent of rules in the home. This should be expected because as children are younger it would be logical to assume that their parents would try harder to shelter them from inappropriate material and attempt to control who they can and cannot talk to on the Internet. As a child grows older and more mature it would be expected that parents grant a little space to their children by taking away some of the rules and hoping that their child will make the right decisions on their own.

In actuality, what we were able to pull out of the data was that the age group that had the most rules was the seventh grade. After putting a little more thought into why the results differed from our hypothesis we were able to reason that it was possible that by fourth grade, the students had not had as much interest in the dark side of the Internet. By the time students were in the seventh grade, and reaching the age of puberty, there was more of a need to restrict what sites they could visit. By 10th grade, students have mostly matured and parents eased some of the rules.

Another explanation we were able to form for our results is that fourth graders probably are unaware of what their parents have set for rules, because most of the things which rules are meant to prevent are not as interesting to a 9 or 10 year old.

Age vs. Household Rules
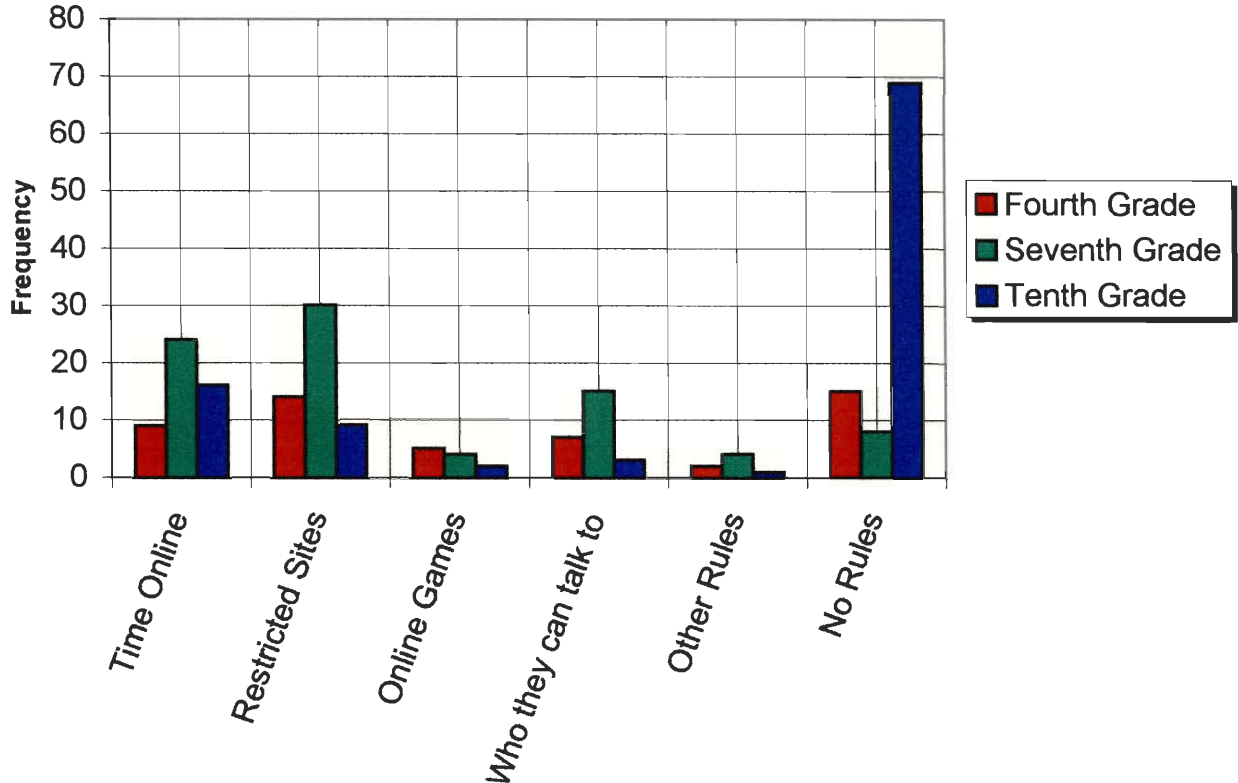
Start them young – Grade vs. Age of First Computer Use.

Another hypothesis that we felt we should take a closer look at was that children are beginning to use computers at a younger and younger age. We feel that one the main reasons for this is that computers are becoming more and more available.

Through our usage of miner and SAS we were able to extrapolate from our data a definitive conclusion that proves our hypothesis. The following graph shows the modes of the ages of first computer separated by grade. It is clear that the younger grades on average tend to start using computers at an earlier age then the older grades.

Grade vs. Age of First Computer Use
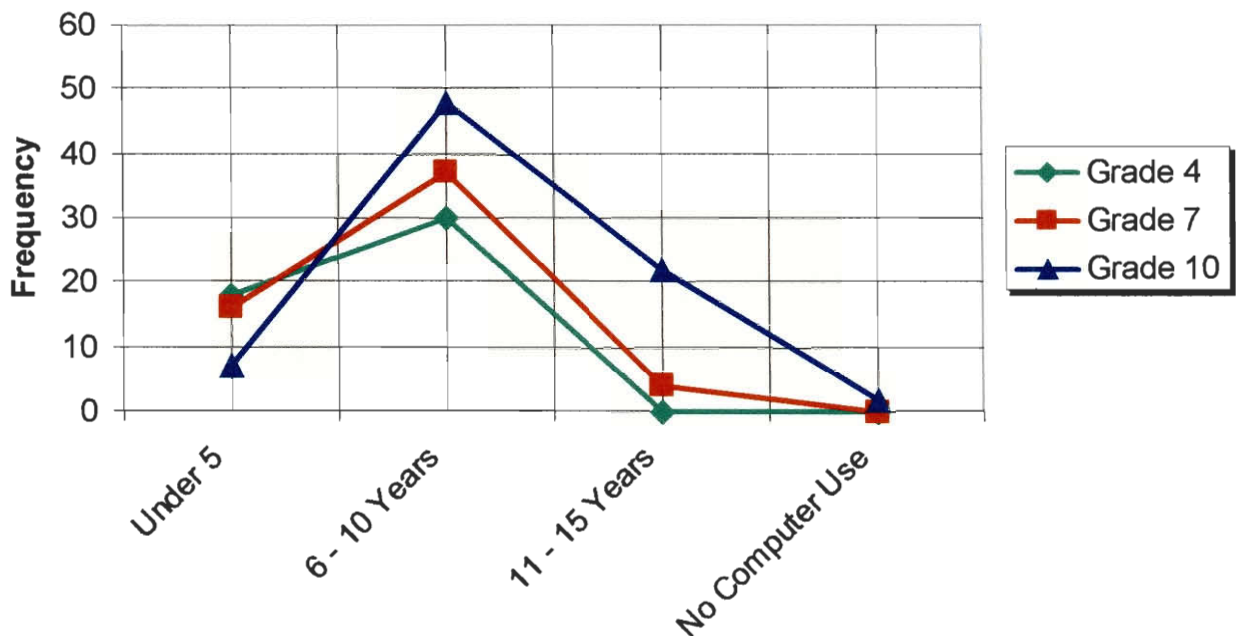
This seemed like a logical correlation to make because people who are older, probably did not have the same level of access to a computer, as did the younger generation. Some of the older subjects that we surveyed grew up during the late 1980s, a time when computers were as common as televisions during the 1950s. Most of the seventh grade students were not even alive until 1990 or so, and they were brought up in a time where most homes had at least one computer and many of them were connected to the Internet. The fourth graders spent most of their lives during the boom and bust of the dot-com era.

The Changing of the Guard – Computer Related Activities vs. Non-Digital Equivalents

One of the hypotheses that we as a group felt held considerable importance in our project was how computer related activities related to their non-digital equivalents. We believe that this is of significant value to how society is beginning to adopt the widespread use of computers. The original hypothesis that we decided upon was that computer related activities were replacing their non-digital counterparts.

In actuality, computers are not replacing the non-digital equivalents but instead the effects of the computer are transparent to the activities. The following set of graphs shows that students who spent a lot of time participating in the non-computer related activities also spend large portions of time using the computer for similar tasks. For example, students who devote a lot of time during the week to playing sports and games also spend a lot of time playing computer games.

Sports vs. Computer Games

Music vs. Computer Music

Legend:
- Listen to Music off computer
- Listen to Music on Computer

Categories: No Response, None, 0 to 1 hours, 1 to 5 hours, 6 to 15 hours, 16 to 31 hours, 31 or more hours, Not Applicable
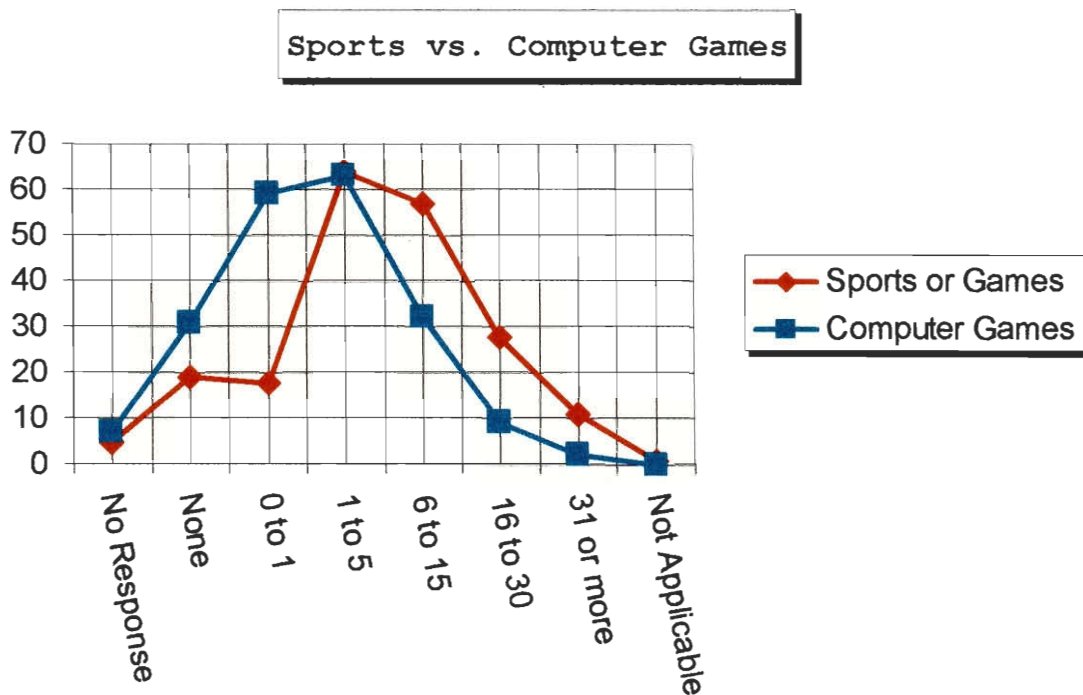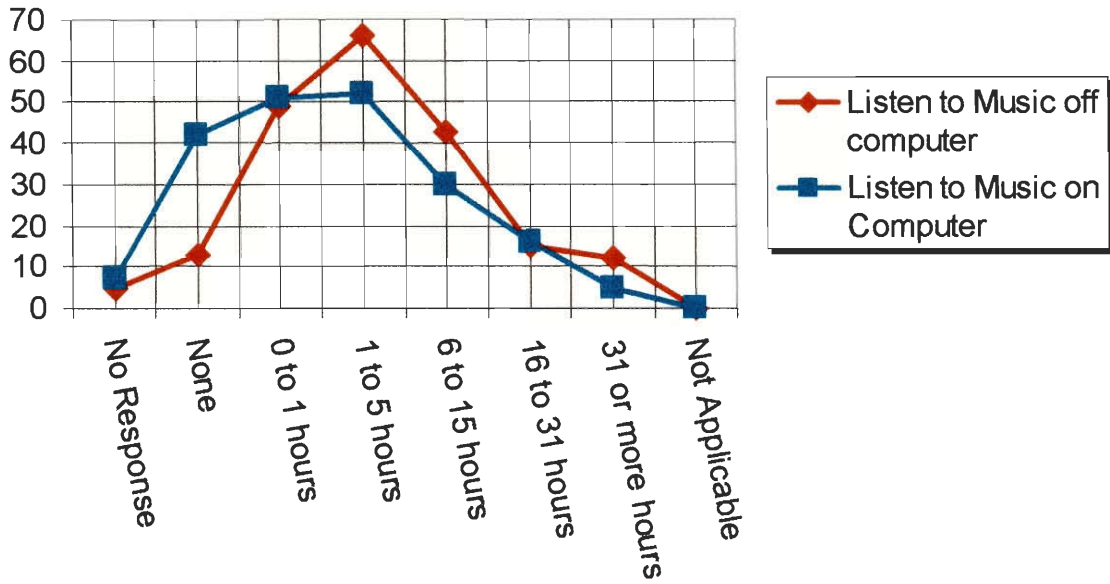
Videos vs. Computer Videos

Legend:
- Watching Videos off Computer
- Watching Videos on Computer

Categories: No Response, None, 0 to 1 hours, 1 to 5 hours, 6 to 15 hours, 16 to 31 hours, 31 or more hours, Not Applicable

51
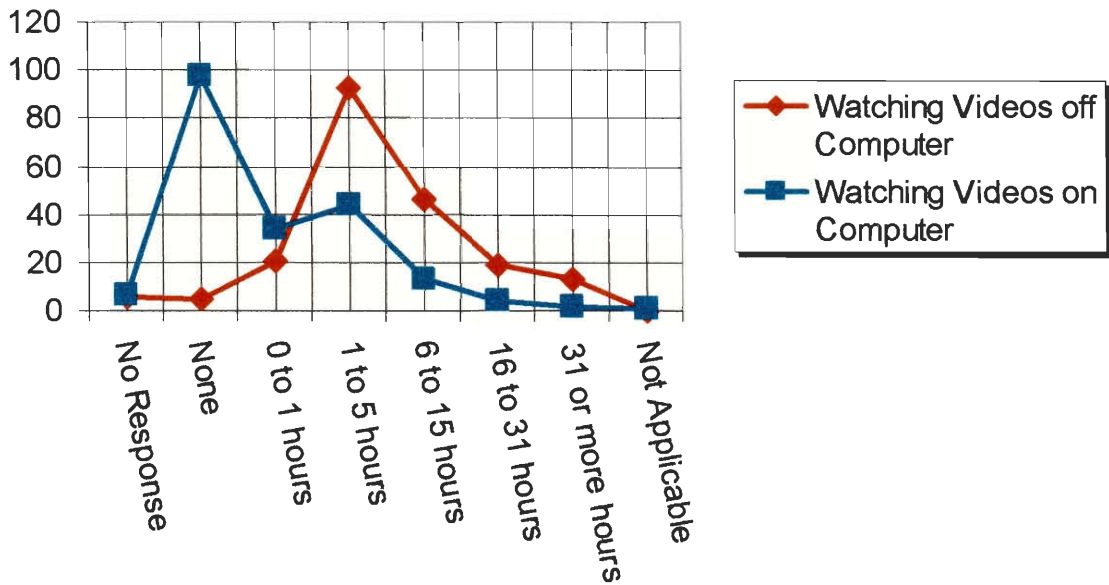
The need for speed – Connection Speed vs. Time Online

Another hypothesis that we were able to find supporting data for was that people with higher connection speeds would spend more time online. Unfortunately as a direct result of such a small data set, the transition is not as apparent as we would hope. The following graph illustrates this point more clearly.

Connection Speed vs. Total Time Online
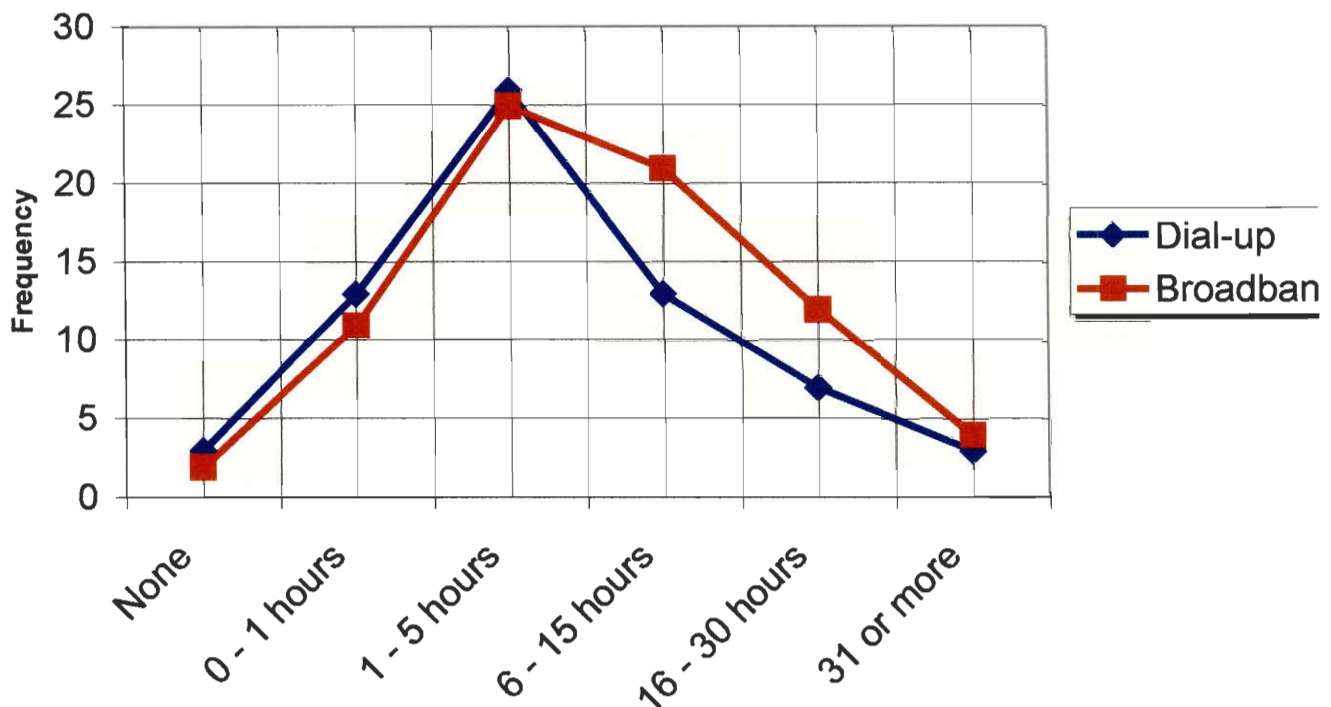
Although it is not very apparent in the graph, the transition is much better portrayed in the raw data and correlations discovered by the miner program.

Our reasoning behind the correlation is this: Dial-up Internet is slow. People tend to dislike waiting for pages to load and true streaming media is not conceivable on even the fastest modems. Most people who are limited to the frustration imposed by Dial-up connections do not want to spend any more time than they have to being online. As the graph indicates, this may not be an issue for limited use, but for many, spending more than 6 hours per week online using a Dial-up connection is too much to bear. Broadband connections are much faster, and as a benefit, browsing the web is a much less stressful activity. Therefore Broadband Internet users tend to have a more enjoyable time online and will tend to use the Internet more often.

Another reason why people on Dial-up connections tend to spend less time online is that it ties up a phone line. Although some Dial-up users will have a second phone line for their Internet connection, many still are limited to using the same line for voice. With Broadband, there is no need to tie up a phone line, and therefore users can remain online for longer periods of time.

Data Mining Results

Rules vs. Primary Access to Computers

It is clear that rules are more commonly enforced on family computers than on computers with a single user. However, there was an even distribution of parents talking to their children about acceptable Internet use across both of these categories. This would seem to indicate that parents do not change their attitudes about what their children should be doing online based on whether their children have their own computers; they simply make more rules when they are in direct control of the computer in question.

Rules vs Primary Access to Computer
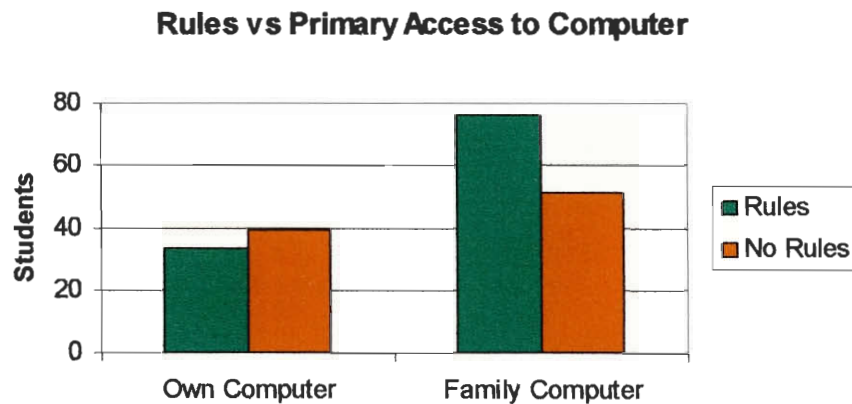
Type of computer vs. Time spent (offline) with friends.

Clearly, those who build their own computers spend less time with their friends than those who do not. This implies that the hobbyist instinct, manifested by building computers, often takes precedence over socialization. Of course, the fact that a student uses a hom-made computer does not necessarily imply that the student was the one that built it.

Type of computer vs. Time spent with friends (offline)
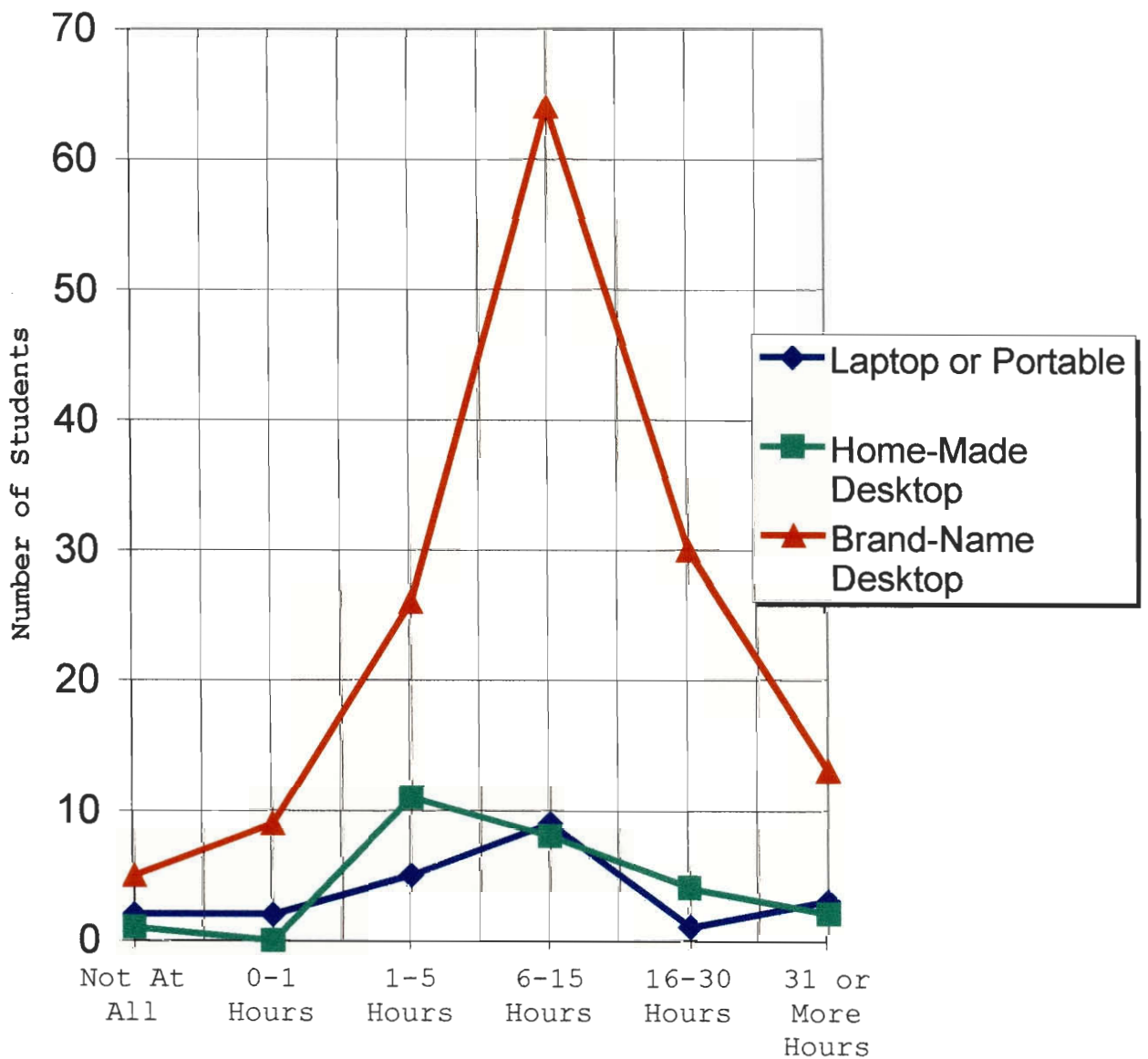
Knowledge vs. Helpfulness

This data is strong evidence of the supporting role of parents while children are growing up. Parents are accessible; they end up helping whenever they can. It stands to reason that someone who knows a lot about computers would be more varied in their quest for learning because no single source is adequate.

Knowledge vs. Who Helped Most
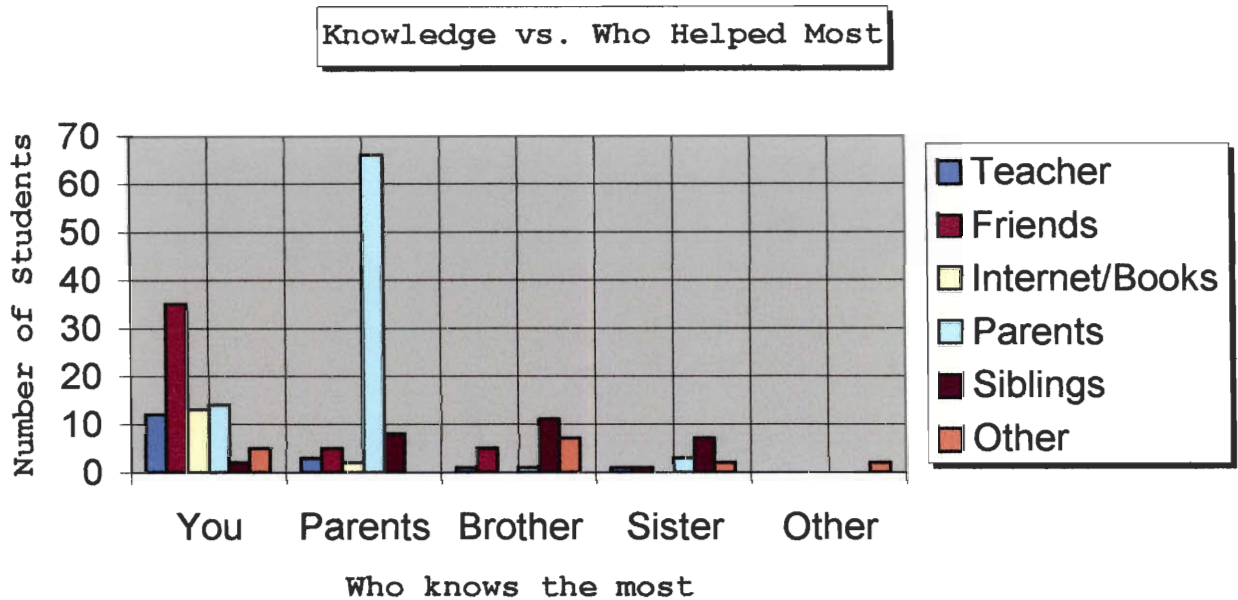
Concluding Remarks on Data Mining

Data mining is a potentially involved process. These conclusions are only a sampling of what we found interesting. Browsing the nearly limitless possibilities that miner provides is a complex task, with much left up to the individual. The data which we collected undoubtedly allows for many more unforeseen conclusions. The tools we used (and wrote) were helpful to us for locating significant conclusions as they should be for others who attempt the same task.

The views depicted in our interpretations of the data represent what we believe to be important correlations.

_6_

Conclusion

Conclusion
Suggestions for the future

Conclusion

Project 404 was very successful. We were able to achieve our initial goal, which was to determine the effects of the spread of information technology on youth. By administering our survey on several different grade levels in different towns across the state of Massachusetts, we were able to obtain conclusive data to support or disprove many of our initial hypotheses. Initially our plan was to survey four different grade levels, but due to our time constraints we could only gather information from three.

It is overwhelmingly clear that computers are here to stay. 99% percent of the students we surveyed use computers, 21% of them having started before the age of five. Most of the students we surveyed have more than one computer in their home. This means that any conclusion we draw must pertain to the effects of the widespread use of this technology.

Computers and the Internet are, fundamentally, ways of processing information. They are tools with many uses. We can only hope that they are used as such. Our data showed that tendencies to participate in an activity were paralleled by a tendency to participate in its digital equivalent. This means that most people are using computers transparently, not as a complete replacement of, or distraction from, conventional tools. Most people care much more about what they are doing than how they are doing it. Thus, we have found that computers are being used as the tools that they are. Fortunately, one of the most misunderstood tools of our time has been widely useful.

Another very positive thing we discovered was that parents have a tendency to talk about acceptable Internet use with their children more readily than making rules. This fosters understanding, not disobedience. We can only infer that fear of the Internet is outweighed by healthier, more balanced attitudes. Parents who know about computers show a propensity for helping their children with computers, in much the same way they help with so many other things.

All of this can only mean that the information age is reaching a healthy maturity. Computers do not require their users to have a hacker instinct, although those that are interested in the technology can enjoy it independently of the majority. The Internet is serving its purpose, even for those who do not undrstand it. Despite the fact that the Internet is filled with spam and illicit material, people find great utility and accessibility in it. Typically, parents are protecting their children from unhealthy aspects of technology in ways that encourage ethical thought instead of blind obedience. This maturing of the information age will allow Computers and the Internet to function in ways that are beneficial to society, and the individuals within it.

Suggestions for the future

One suggestion that we would have about the future is that another group of students perform an extension to our work and compare how the values change over time. The disadvantage to being the first group to gather this type of information is that we can only make broad generalizations about the current situation. If another IQP group were to pick up where we left off, they would be able to see differences, which we were initially aiming to capture. A good time to perform such a survey would be every three years, because by that time, the students that we surveyed will have moved up to the next survey grade level, so the fourth graders would become the seventh graders, and the seventh graders would be in tenth grade.

7

Bibliography

Bibliography

A Brief History of the Internet. Leiner, Barry M. et. al. 4 Aug. 2000. The Internet Society. 24 Apr 2003.
   <http://www.isoc.org/internet/history/brief.shtml>

About the WPI Projects Program Worcester Polytechnic Institute. 28 Apr 1999
   <http://www.wpi.edu/Academics/Projects/intro.html>

Analysis of Online Societies. Sheldon, Aathan A. and Peter L. Wright. 2002. Worcester Polytechnic Institute.
   <http://www.wpi.edu/Academics/Depts/IGSD/Interactions/viewproject.php?exec=0&r=02d123i>

Civin, Michael A. Male, Female, E-Mail. New York: Other Press. 2000.

Frates, Jeffrey. Computers, And Life. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. 1983.

Grossman, Wendy M. Net.wars. New York: New York University Press. 1997.

Introducing Young Children to Basic Computer Skills. Curry, Melissa J. et. al. 2000. Worcester Polytechnic Institute.
   <http://www.wpi.edu/Academics/Depts/IGSD/Interactions/viewproject.php?exec=0&r=00d017i>

KMUTT Internet Survey. Bresson, Jeffrey D. et. al. 2000. Worcester Polytechnic Institute.
   <http://www.wpi.edu/Academics/Depts/IGSD/Interactions/viewproject.php?exec=0&r=00c017i>

Napster: Impact on Intellectual Property Law. Morse, Carrie Ellen. 2001. Worcester Polytechnic Institute.
   <http://www.wpi.edu/Academics/Depts/IGSD/Interactions/viewproject.php?exec=0&r=01b022i>

Oteri, Anthony. Personal Communication. 21 Mar. 2002

Porter, David. Internet Culture. London: Routledge. 1997.

Project Xanadu History. 05 Oct 2001. The Xanadu Project. 24 Apr 2003.
   <http://www.xanadu.net/xuhistory.html>

Robertson, Douglas S. The New Renaissance. New York: Oxford University Press. 1998.

Rochlin, Gene I. Trapped in the Net. Princeton, New Jersey: Princeton University Press. 1997.

Shapiro, Andrew L. The Control Revolution: How the Internet is Putting Individuals in Charge and Changing the World We Know. New York: Century Foundation. 1999.

Shields, Rob. Ed. Cultures of Internet. London: SAGE Publications Ltd. 1996.

Sikes, Alfred C. Fast Forward. New York: William Morrow. 2000.

Silver, Gerald A. The Social Impact of Computers. New York: Harcourt Brace Jovanovich. 1979.

Talbott, Stephen L. The Future Does Not Compute: Transcending the Machines in Our Midst. Sebastopol, California: O'Reilly & Associates. 1995.

Tapscott, Don. Growing Up Digital. New York: McGraw-Hill. 1998.

Appendix A

The Survey

Survey as Given
Questions with Additional Information
Data Entry Web Site

Survey as Given

Project 404: REIEFS - Results of Electronic Information Exchange in a Free Society

----

Preliminary Survey:

To be read aloud to the survey group:
This survey is designed to collect information about your computer and internet use. We are also trying to find out a little about who you are and how you spend your time as far as it can be correlated to your computer and internet use. Your name and survey location will be recorded for statistical purposes, but not in conjuction with the survey you filled out. Please read and understand all questions carefully. Answer the questions to the best of your ability, but if you do not understand a question you may skip it. If you feel it necessary, you may write comments in the margin to help answer a question. Please be completely honest, as this survey is anonymous. Your name and any other personal information you submit will be kept strictly confidential.

Name _____

Date _____

School Name _____

Grade Level
( ) 4th
( ) 7th
( ) 10th
( ) College Freshman
( ) Other_____

Project 404: REIEFS

Personal Information

How old are you? _____

Are you ( ) Male or ( ) Female?

Are you..
- ( ) Christian?
- ( ) Jewish?
- ( ) Muslim?
- ( ) Other? (please specify) _____

- ( ) Hindu?
- ( ) Agnostic? (not sure about God/Religion)
- ( ) Atheist? (no belief in God)

What is your political preference?
- ( ) Democrat
- ( ) Republican
- ( ) Independent
- ( ) Other (please specify) _____

- ( ) Green Party
- ( ) Libertarian
- ( ) No Preference

What is your household income? (or best guess)
- ( ) Less than $30,000
- ( ) $30,000-$50,000
- ( ) $50,000-$70,000

- ( ) $70,000-$90,000
- ( ) $90,000-$120,000
- ( ) More than $120,000

One of your parents or guardians involved in... (check all that apply) (as a job or otherwise)
- ( ) Engineering/Science
- ( ) Education
- ( ) Social occupations
       (lawyer, social worker, sales, etc.)
- ( ) Arts/entertainment
- ( ) Government/military/police
- ( ) Other_____

- ( ) Management/Administration
- ( ) A professional trade (carpentry, plumbing, etc.)
- ( ) Homemaking
- ( ) Medical
- ( ) Physical labor (factory worker, construction, etc)
- ( ) Data entry/Processing
- ( ) Business/Finance

What is your favorite academic subject in school?
- ( ) English
- ( ) Social Studies (History, geography, etc)
- ( ) Foreign Language

- ( ) Mathematics
- ( ) Science
- ( ) Arts (visual or performing)

What is your least favorite academic subject in school?
- ( ) English
- ( ) Social Studies (History, geography, etc)
- ( ) Foreign Language

- ( ) Mathematics
- ( ) Science
- ( ) Arts (visual or performing)

The area you live in is:
- ( ) Urban (city)
- ( ) Suburban (small city/large town)

- ( ) Rural (small town)

How long have you lived in your current home?
- ( ) Less than a year
- ( ) 1-3 years

- ( ) More than 3 years, but not all your life
- ( ) All your life

Computer Usage

Check the statement that you agree with **most**:
- ( ) Technology is a positive influence on society.
- ( ) Technology does not greatly influence society.
- ( ) Technology is a negative influence on society.

Check the statement that you agree with **most**:
- ( ) Advanced technology is necessary for progress as a society.
- ( ) Technology is necessary, but should be limited.
- ( ) Advanced technology is unnecessary for progress as a society.

How old were you when you first started using computers?
- ( ) Under 5
- ( ) 6-10
- ( ) I don't use computers.
- ( ) 11-15
- ( ) Over 15

How do you primarily access a computer?
- ( ) I have my own
- ( ) My family has one
- ( ) No computer access
- ( ) At school/Library
- ( ) Other_____

How many computers are in your home?
- ( ) 0
- ( ) 2
- ( ) 4 or more
- ( ) 1
- ( ) 3

How old is the newest computer in your home?
- ( ) Less than 6 months old
- ( ) 18 months - 3 years old
- ( ) More than 6 years old
- ( ) 6 - 18 months old
- ( ) 3 - 6 years old
- ( ) N/A

How many hours (maximum) per week do you spend using a computer.
- ( ) Not at all
- ( ) 1-5
- ( ) 16-30
- ( ) 0-1
- ( ) 6-15
- ( ) 31 or more

What operating system do you use on the computer you use the most?
- ( ) Microsoft Windows (XP, 2000, or NT)
- ( ) Mac OS X
- ( ) Unix/Linux
- ( ) Microsoft Windows (ME, 98, 95, or 3.1)
- ( ) Mac OS (version 9 or earlier)
- ( ) Other _____

Is the computer you use the most a:
- ( ) Laptop/portable
- ( ) Home-made Desktop
- ( ) Brand-Name Desktop (Dell, Gateway, Compaq, etc.)
- ( ) Other_____

Do you have your own e-mail account?
- ( ) Yes
- ( ) No

How do you primarily access the Internet?
- ( ) My own computer
- ( ) My family computer
- ( ) No Internet access
- ( ) At school/library
- ( ) Other_____

What kind of internet connection do you use?
- ( ) Dial-up
- ( ) High speed network (T1, T3, etc.)
- ( ) I don't know
- ( ) Broadband (cable modem or dsl)
- ( ) Other_____

How many hours (maximum) per week do you spend doing the following, using a computer?

Word Processing	( )	( )	( )	( )	( )	( )	( )
Preparing Presentations	( )	( )	( )	( )	( )	( )	( )
Playing Games	( )	( )	( )	( )	( )	( )	( )
Listening to Music	( )	( )	( )	( )	( )	( )	( )
Watching videos (Movies or TV)	( )	( )	( )	( )	( )	( )	( )
Browsing the Web	( )	( )	( )	( )	( )	( )	( )
Online	( )	( )	( )	( )	( )	( )	( )
Total Computer Use	( )	( )	( )	( )	( )	( )	( )

Do you use e-mail for (check all that apply) :
( ) Personal   ( ) Business
( ) School   ( ) Other_____

How many pieces of unwanted/unsolicited/junk e-mail do you get per week in your primary account?:
( ) None   ( ) 1-10
( ) 11-40   ( ) 40-100
( ) 100-500   ( ) More than 500
( ) N/A

Are there rules in your household regarding your internet use?:
( ) Yes, How much time you can spend online   ( ) Yes, What sites you can visit
( ) Yes, What games you can play online   ( ) Yes, Who you can/cannot talk to
( ) Yes, other (Please Specify) _____   ( ) No

Has a parent or guardian ever discussed acceptable internet use with you?
( ) Yes   ( ) No

Decide which of your parents or guardians uses a computer the most. How many hours per week does he/she use a computer (or best estimate) ?
( ) Not at all   ( ) 0-1
( ) 1-5   ( ) 6-15
( ) 16-30   ( ) 31 or more
( ) N/A?

Who knows the most about computers in your household:
( ) You   ( ) Parents
( ) Brother   ( ) Sister
( ) Other_____

Who has helped you the most with computers:
( ) Teacher   ( ) Parents
( ) Friends   ( ) Siblings
( ) Internet Resources/books   ( ) Other_____

How many hours per week (maximum) do you spend doing the following *without* the use of a computer or the Internet?

Spending free time with friends?	( )	( )	( )	( )	( )	( )	( )
Using the telephone?	( )	( )	( )	( )	( )	( )	( )
Reading Newspapers or Magazines?	( )	( )	( )	( )	( )	( )	( )
Sports/Games?	( )	( )	( )	( )	( )	( )	( )
Listening to music?	( )	( )	( )	( )	( )	( )	( )
Watching videos (Movies/TV/Other) ?	( )	( )	( )	( )	( )	( )	( )
Hobbies/Other?	( )	( )	( )	( )	( )	( )	( )

Questions with Additional Information

The numbers in each option serve as an index for the values used for that answer in our digital version of our data. For the questions that were of the type "select all that apply" each one of the options was itself a question as far as the miner.pl program is concerned. Since more than one answer could be chosen, it had to be in a different format. The information that appears in parentheses is the question number and question key. The numbers next to each possible response were used in the comma-separated-value file that stored our data (See Appendix D for details). The first 3 (Questions start at 0) questions 0-2 were survey number, survey school, and survey grade.

How old are you? _____ **(Question 3 - age)**

Are you (1) Male or (2) Female? **(Question 4 - sex)**

The integer the survey taker entered in the age question was used as the value in the raw data. The gender and age of the survey taker were especially important to many correlations that will be drawn.

(Question 5 – religion)

Are you..
- (1) Christian?
- (3) Jewish?
- (3) Muslim?
- (7) Other? (please specify) _____

- (4) Hindu?
- (5) Agnostic? (not sure about God/Religion)
- (6) Atheist? (no belief in God)

This question asks the user for their religion. This was intended to compare Internet uses with people of different faiths to look for correlations.

This was one of the three questions that a lot of the school systems did not allow us to ask.

(Question 6 – politics)

What is your political preference?
- (1) Democrat
- (2) Republican
- (3) Independent
- (7) Other (please specify) _____

- (4) Green Party
- (5) Libertarian
- (6) No Preference

This question tries to probe the survey taker to find out their political preference. This would also make for interesting correlations, if there were any.

This was one of the three questions that a lot of the school systems did not allow us to ask.

(Question 7 – income)

What is your household income? (or best guess)	
(1) Less than $30,000	(4) $70,000-$90,000
(2) $30,000-$50,000	(5) $90,000-$120,000
(3) $50,000-$70,000	(6) More than $120,000

This question was added to compare the upper and lower classes, and see if it correlates to proficiency with computers.

This was one of the three questions that a lot of the school systems did not allow us to ask.

(Question 8 – parents_engineer)
(Question 9 - parents_education)
(Question 10 - parents_social)
(Question 11 – parents_arts)
(Question 12 - parents_gov)
(Question 13 - parents_management)
(Question 14 - parents_trade)
(Question 15 - parents_homemaking)
(Question 16 - parents_med)
(Question 17 - parents_labor)
(Question 18 - parents_dataent)
(Question 19 - parents_business)
(Question 20 - parents_other)

One of your parents or guardians involved in... (check all that apply) (as a job or otherwise)			
(1) Engineering/Science **(13)**		**(8)**	(1) Management/Administration
(1) Education **(14)**		**(9)**	(1) A professional trade (carpentry, plumbing, etc.)
(1) Social occupations **(15)**		**(10)**	(1) Homemaking
(lawyer, social worker, sales, etc.) **(16)**			(1) Medical
(1) Arts/entertainment etc) **(17)**		**(11)**	(1) Physical labor (factory worker, construction,
(1) Government/military/police **(18)**		**(12)**	(1) Data entry/Processing
(1) Other_____ **(19)**		**(20)**	(1) Business/Finance

This question was intended to obtain the trade their parents are in. Correlations between what the parents do for a living, and the proficiency with computers with their children were sought.

(Question 21 – favsub)

What is your favorite academic subject in school?	
(1) English	(4) Mathematics
(2) Social Studies (History, geography, etc)	(5) Science
(3) Foreign Language	(6) Arts (visual or performing)

Favorite subjects can give clues as to what kind of person someone is. It can also be correlated to gender. With a large number of surveys, we can see what people's favorite subjects were, and what people's least favorite subjects were, and draw correlations between the two.

(Question 22 – lfavsub)

What is your least favorite academic subject in school?	
(1) English	(4) Mathematics
(2) Social Studies (History, geography, etc)	(5) Science
(3) Foreign Language	(6) Arts (visual or performing)

A person's ability or lack thereof with a certain subject may have an influence on their perception of computers. Computer technology is based off of Science and Mathematics, so perhaps people that have Mathematics or Science, as their least favorite subject will also dislike computers.

(Question 23 – location)

The area you live in is:	
(1) Urban (city)	(3) Rural (small town)
(2) Suburban (small city/large town)	

We had originally intended to have a school system (4th, 7th, and 10th grades) from each section. Comparison between each of those schools could yield information such as what type of living environment is the most likely to have children that are proficient with computers grow up out of them.

(Question 24 - length-residence)

How long have you lived in your current home?	
(1) Less than a year	(3) More than 3 years, but not all your life
(2) 1-3 years	(4) All your life

This question was an addition to the previous question that would validate the data. If someone chose urban setting, but lived in a rural one most of his life, then the data isn't really accurate. This question allows us to probe that, and draw more accurate conclusions.

(Question 25 – influence)

Check the statement that you agree with **most**:
- (1) Technology is a positive influence on society.
- (2) Technology does not greatly influence society.
- (3) Technology is a negative influence on society.

This question probes the survey taker for their perception of the importance of technology. People who think it is of great influence, may take it more seriously and use it more.

This question had version control problems. In the incorrect version, option 2 said "Technology is necessary, but should be limited," which is incorrect. Thus, we had to throw out all the data from this question.

(Question 26 – necessity)

Check the statement that you agree with **most**:
- (1) Advanced technology is necessary for progress as a society.
- (2) Technology is necessary, but should be limited.
- (3) Advanced technology is unnecessary for progress as a society.

This is a similar question to the previous one, but it's more serious. It doesn't just ask if it's an influence, it asks if it's necessary. If society couldn't function without it. It gives us a better representation of their attitude than the previous question by themselves, and forces them to either back down a bit from their statement, or strongly reinforce it. By examining the two questions, we can determine how strong their opinion on technology is, and what that opinion is.

This question also had version control problems. In the incorrect version, option 2 did not exist, and there was only option 1 and option 3. Since these questions were not consistent, we had to throw out all the data from this question.

(Question 27 – firstcompuse_age)

How old were you when you first started using computers?
- (1) Under 5
- (2) 6-10
- (3) 11-15
- (4) Over 15
- (5) I don't use computers.

This question asks the survey taker when they first started using computers. This question allows us to see if people are starting to use computers at a younger age. We can compare the results from different grade levels to see if there is a change. There most likely will be a change, because for the 10th graders, computers were not nearly as commercially available when they were extremely young as they are now.

(Question 28 – comp_acess)

How do you primarily access a computer?	
(1) I have my own	(3) At school/Library
(2) My family has one	(5) Other_____
(4) No computer access	

This question helps us understand how exposed to a computer a survey taker is. If they have their own computer, then they, or their parents, bought them one for a reason. It likely gets used often. If they use the family computer, they probably still use it a lot, but probably not as much as if they had their own computer. The question that asks them how much time they spend using a computer, allows us to compare those results with these and determine the appropriate correlations.

(Question 29 – num_systems)

How many computers are in your home?	
(1) 0	(2) 1
(3) 2	(4) 3
(5) 4 or more	

We asked this questions because we wanted to know if there would be an increase over previous generation. When we were in public school as youths, one computer was considered plenty. Now, families might have several computers in their household. This question examines that trend.

(Question 30 – oldest_comp)

How old is the newest computer in your home?	
(1) Less than 6 months old	(2) 6 - 18 months old
(3) 18 months - 3 years old	(4) 3 - 6 years old
(5) More than 6 years old	(6) N/A

This question will tell us how up-to-date the technology a survey taker is using.

(Question 31 – hours_max)

How many hours (maximum) per week do you spend using a computer:	
(1) Not at all	(2) 0-1
(3) 1-5	(4) 6-15
(5) 16-30	(6) 31 or more

This question directly asks the user how much time they spend on a computer.

(Question 32 – os)

What operating system do you use on the computer you use the most?

(1) Microsoft Windows (XP, 2000, or NT)   (4) Microsoft Windows (ME, 98, 95, or 3.1)
    (2) Mac OS X   (5) Mac OS (version 9 or earlier)
    (3) Unix/Linux   (6) Other _____

Sometimes the operating system has an impact on a person's perception of technology. A poor interface or unstable system can give a person a negative outlook on technology. Each of these operating systems is different, and could lead to different survey results. Correlations may be drawn.

(Question 33 – computer_type)

Is the computer you use the most a:
(1) Laptop/portable   (3) Brand-Name Desktop (Dell, Gateway, Compaq, etc.)   (2) Home-made Desktop   (4) Other_____

This question examines what kind of computer hardware they are using. Most people will have a Brand-name Desktop, but there will be some that have a Home-made Desktop or a Laptop.

(Question 34 – email_account)

Do you have your own e-mail account?
(1) Yes   (2) No

This question indicates control over technology. If a survey taker has their own email account it shows an element of control over technology, and not just using what their parents set up for them. We were thinking ahead and we figured we could get some correlations by asking this question and comparing them to other questions.

(Question 35 – internet_access)

How do you primarily access the Internet?

(1) My own computer   (4) At school/library
    (2) My family computer   (5) Other_____
    (3) No Internet access

This question not only asks what computer they use to connect to the Internet, but if they even have access to it at all. Consistency is also an issue here, to see if the answer here, is in accordance to the question that asked how they primarily accessed a computer, to see if it is the same.

(Question 36 – connection)

What kind of Internet connection do you use?
- (1) Dial-up
- (2) High speed network (T1, T3, etc.)
- (3) I don't know
- (4) Broadband (cable modem or dsl)
- (7) Other_____

The type of Internet connection they have determine the speed. This is important to know because people might become less frustrated with the Internet if it is fast. Also, if they are not using dial-up, they have a constant connection to the Internet, without having to dial up. This fact may skew their perception of the Internet, which we can test for in other questions.

(Question 37 – hours_wordprocessing)
(Question 38 – hours_presentations)
(Question 39 – hours_games)
(Question 40 – hours_music)
(Question 41 – hours_movies)
(Question 42 – hours_browsing)
(Question 43 – hours_online)
(Question 44 – hours_total)

How many hours (maximum) per week do you spend doing the following, using a computer?							
	Not at all	0 - 1	1 - 5	6 - 15	16 - 30	31 or more	N A
Word Processing **(37)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Preparing Presentations **(38)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Playing Games **(39)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Listening to Music **(40)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Watching videos (Movies or TV) **(41)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Browsing the Web **(42)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Online **(43)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total Computer Use **(44)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)

This question does a total breakdown of how many hours someone spends on a computer, and what they do while on the computer. Many of these activities overlap, but we can get a general idea of what people are doing the most while using a computer. Averaged out with all the survey takers, we can see what kids use computers for, and what they don't or at least very minimally use a computer for.

Do you use e-mail for (check all that apply) :			
(1) Personal	**(Question 45 – personal)**	(1) Business	**(Question 47 – business)**
(1) School	**(Question 46 – education)**	(1) Other_____	**(Question 48 – other)**

We asked this question because we need to know if people are using the Internet for social reasons, business reasons, academic reasons, or any combination of the three. It will give us insight on the outlook people have for electronic communication, as well as dependence from businesses, teachers, friends, etc. on electronic communication.

(Question 49 – spam)

How many pieces of unwanted/unsolicited/junk e-mail do you get per week in your primary account?:	
(1) None	(2) 1-10
(3) 11-40	(4) 40-100
(5) 100-500	(6) More than 500
(7) N/A	

The level of spam someone gets is a direct result from their behavior online. People don't like spam, and if someone gets a minimal amount of spam, than they know how to use the Internet anonymously. On the other hand, someone who gets a minimal amount of spam, but doesn't spend any time online, isn't getting spam for the same reasons. A lot of spam comes from giving sites your email address voluntarily, and they in turn sell the information to spammers. People that give Internet sites junk email addresses instead of their primary ones will get less spam.

(Question 50 – yes_timeonline)
(Question 51 – yes_sites)
(Question 52 – yes_games)
(Question 53 – yes_talk)
(Question 54 – yes_other)
(Question 55 – no_rules)

Are there rules in your household regarding your Internet use?:	
(51) (1) Yes, How much time you can spend online	(50)(1) Yes, What sites you can visit
(53) (1) Yes, What games you can play online	(52)(1) Yes, Who you can/cannot talk to
(55) (1) Yes, other (Please Specify) _____	(54)(1) No

A huge aspect of the Internet with regards to social values are rules. There is a lot of uncensored individuals, information, and obscene material on the Internet, and several correlations are going to be drawn with respect to rules.

(Question 56 – acceptable_use)

Has a parent or guardian ever discussed acceptable Internet use with you?
(1) Yes   (2) No

This question brings the parent's behavior into question, not the survey taker. Parents that know about the content available on the Internet will likely discuss acceptable use with their children. On the other hand, parents who are uninformed will not, and this question allows us to examine how informed the survey taker's parents are.

(Question 57 – guardian_hours)

Decide which of your parents or guardians uses a computer the most. How many hours per week does he/she use a computer (or best estimate) ?
(1) Not at all   (2) 0-1
(3) 1-5   (4) 6-15
(5) 16-30   (6) 31 or more
(7) N/A?

Lifestyles of a parent often rub off onto a child. The amount of time their parents use a computer could have an affect on how much the child uses computers.

(Question 58 – household_knowledge)

Who knows the most about computers in your household:
(1) You   (2) Parents
(3) Brother   (4) Sister
(7) Other_____

We asked this question because we wanted to know the source of information. The sources of information have an affect on information itself. This is an indirect source of information because it does not necessarily mean that they are shown how to use a computer well by this person.

(Question 59 – helped_you)

Who has helped you the most with computers:	
(1) Teacher	(4) Parents
(2) Friends	(5) Siblings
(3) Internet Resources/books	(7) Other_____

This is the direct source of information for our survey takers. Who has helped them the most is a very important question in a social sense. If it was their brother or sister, they might have been taught some really interesting social stuff to do online, such as finding chat rooms, using file sharing programs to get music, etc. Learning from a teacher would be quite different, in that they would likely be taught how to use search engines to find information and do research. Being taught by parents would be more influenced by their rules, while Internet resources/books allows the survey taker to develop their own vision of the Internet, develop their own preferences, and learn what they want to learn.

(Question 60 – nocomp_friends)
(Question 61 – nocomp_telephone)
(Question 62 - nocomp_newspapers)
(Question 63 - nocomp_sports)
(Question 64 - nocomp_music)
(Question 65 - nocomp_videos)
(Question 66 - nocomp_hobbies)

How many hours per week (maximum) do you spend doing the following *without* the use of a computer or the Internet?

	Not at all	0 - 1	1 - 5	6 - 15	16 - 30	31 or more	NA
Spending free time with friends? **(60)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Using the telephone? **(61)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Reading Newspapers or Magazines? **(62)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sports/Games? **(63)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Listening to music? **(64)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Watching videos (Movies/TV/Other) ? **(65)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hobbies/Other? **(66)**	(1)	(2)	(3)	(4)	(5)	(6)	(7)

This question, like the computer question similar to it earlier, is asking about how much time they spend doing certain activities. A lot of these questions are directly related to the computer questions, only without computers. This question, combined with the earlier similar computer question, compares and contrasts the two, and we wanted to be able to determine if computers detract and take time away from non-computer related activities.

Data Entry Web Site

Appendix B

Public Relations Package

Cover Letter
Brochure
Logo
Elements of website

Cover Letter to Survey Candidate Schools

After contacting candidate survey schools on the phone, we often sent out a personalized copy of this Cover Letter along with a Brochure, to help explain Project 404.

<School Name>
<Address>
<City, ST ZIP>

Dear <School Name>:

   We are a group of undergraduate students at Worcester Polytechnic Institute researching the effects of the spread of information technology on our society. The project is entitled "Project-404: **R**esults of **E**lectronic **I**nformation **E**xchange in a **F**ree **S**ociety". Our hope is to survey students from your school system. Specifically, we would like to survey a class of 20-30 students from the 4th, 7th, and 10th grades. The survey is short, easy, and a great catalyst for meaningful classroom discussion. We request the use of one of your classes (such as homeroom or an unleveled, required class) for 15-20 minutes. Wednesdays are the best time for us to survey your students. Enclosed is an informational brochure explaining our project in more detail, including the types of conclusions we wish to draw. The survey we have formulated is designed to draw parallels between activities that utilize Electronic mediums (like the Internet), and their non-Electronic counterparts. There are other areas of importance on the survey as well; please review the copy we have enclosed.

It should be stressed that our survey is **anonymous**. There is a cover page that requests each student's name and grade: it is only to ascertain the validity of each survey. This cover page will be removed and separated from each survey as they are collected, to retain anonymity.

We would appreciate your participation in our project. Our website, at http://www.project-404.net, will be updated with our progress as we go. At the completion of our project, our data will consist of results from test groups at over a dozen public Elementary, Middle/Junior, and High Schools, and public or private College Institutions. This project is classified as an IQP (Interactive Qualifying Project). For more info on the IQP see our brochure or http://www.wpi.edu/Pubs/Catalogs/Ugrad/Current/iqp.html .

Thank you,

The R.E.I.E.F.S. Team:
Kevin D'Aquila
Scott Longley
Craig O'Brien
Anthony Oteri
Project Advisor Prof. Michael Ciaraldi

info@project-404.net
http://www.project-404.net

Enclosures

Brochure

To assist with communication of our Project goal to faculty at prospective survey sites, we designed an informational brochure. This document was created in Microsoft Publisher. To produce a printed brochure, the first page is printed, and the sheet is turned over and fed back into the printer. The second page of the document is printed on the back side of the first print. Next, the margins are trimmed off, and the document is folded three ways. For the purpose of the Appendix, the Brochure follows on two separate pages.

Have you ever wondered:

Can I become addicted to the Internet?

Is "Big Brother" watching me?

Do computers improve education?

Is our society too dependent on technology?

Is my data secure?

Can I safely shop on the Internet?

Is the Internet a haven for anarchy?

Are children misusing the Internet?

Is the Internet actually anonymous?
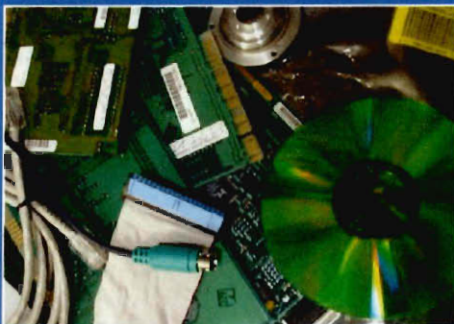
Our Team:

Kevin J. D'Aquila
Scott E. Longley
Craig M. O'Brien
Anthony P. Oteri

Project Advisor
Prof. Michael Ciaraldi

WPI
http://www.wpi.edu

For more information,
visit our Website at:
http://www.project-404.net
or E-Mail
info@project-404.net

REIEFS

Results of
Electronic
Information
Exchange in a
Free
Society

WORCESTER POLYTECHNIC INSTITUTE · 1865 · **WPI**

http://www.project-404.net

The Idea?

R.E.I.E.F.S is a study of the effects of technology upon society. Our goal is to gather conclusive data about the social effects of increased usage and availability of information technology. Our topics of interest include: primary uses of the Internet, technology addictions, benefits and drawbacks of early exposure to computers, online shopping and gambling, effects of widely distributed adult content, etc. The Internet has major societal effects that influence our daily lives. As popularity of the Internet increases, the way it interacts with society deserves serious examination.

Project Overview:

We have prepared a short survey designed to gather conclusive data regarding social effects of electronic media.

All data collected is completely *anonymous.*

Our survey is composed of approximately thirty questions, and requires no more than twenty minutes to take.

Questions will gather information about time spent on various activities, both electronic and otherwise. We will also attempt to ascertain basic social tendencies.

You can follow the project at: *http://www.project-404.net* .

The Interactive Qualifying Project (IQP):

The IQP challenges students to identify, investigate, and report on a self-selected topic examining how science or technology interacts with societal structures and values. The objective of the IQP is to enable WPI graduates to understand, as citizens and as professionals, how their careers will affect the larger society of which they are part.

The interactive project is a broad and integrative educational experience, employed to relate the scientific or technological component to a social issue.

WPI

For more information, visit our Website at:
http://www.project-404.net
or E-Mail
info@project-404.net

Project 404 Logo

Our Project needed an identity, so we created a graphical logo. The phrase "Project 404" was formed by darkening sections of the background in the shapes of these characters. "Results of Electronic Information Exchange in a Free Society" is denoted by the acronym REIEFS in the foreground of the logo. This logo was used in our website (see below) and informational brochure.

Project 404 Website: http://www.project-404.net

Early in the development of our IQP, we designed and implemented a project website. Especially due to the nature of Project 404, it served an important purpose. This site contained links to Progress Reports, Schedules for our group meetings, links to download materials related to our project, and general Project 404 information. Upon conducting surveys, we would leave the teacher and class with the URL to this site so they could learn more about the project and see our conclusions.

Appendix C

Source Code

miner
nix
kr
NIX
num2pr
genex

The following is the source code for miner and the associated utilities. These were used for data mining and refining results into more useable forms. This code is responsible for the creation of several hypotheses. These programs appear in the order that they were usually used on a set of data.

miner

```perl
#!/usr/bin/perl -Tw

 #      #       #     #      #    ######   #####
 ##   ##       #     ##     #    #         #     #
 # ## #        #     # #    #    #####     #     #
 #     #       #     #  # #      #          #####
 #     #       #     #   ##      #         #     #
 #     #       #     #     #    ######    #      #

# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
# Datamining Program
#
# miner was written by
# Scott Longley and Kevin D'Aquila
# # # # # # # # # # # # # # # # # # # # # #
# R.E.I.E.F.S. IQP Project
# http://www.project-404.net
# info@project-404.net
#
# This program is designed to do some lightweight datamining on a set of
# data stored in a comma-seperated-value list. @rawData is an array that
# holds the entire list of data, one line of comma-seperated-values per
# array index.
#
# There are a few subroutines with their own descriptions. Once data is
# run through sortData, a two-dimensional array is returned. This array
# can be used to obtain data from any position in the .CSV:
#     Ex: To get the fifth value from the fifth line of a file:
#         $sortedData[4]->[4];
#     Note that index values start at 0 in Perl.
# # # # # # # # # # # # # # # # # # # # # #
# USAGE: ./miner.pl [filename]
#
# If a filename is given to the program, it will attempt to parse the file. If
# multiple filenames are given, all but the first are ignored. If no filename
# is given,
#
# INPUT FILE FORMAT:
# The file is a ".csv" (comma separated value file). Each line of the file,
# except the first line, represents a survey. The first value of each line is
# only an identifier for the survey (line). This means that the first value in
# each line is not data, and is not parsed as such. All other values in each
# line are integers that denote the answer to the corresponding question. The
# questions are defined by the first line of the file. Each value in this line
# is a string that summarizes the meaning of the question.
#
# EXAMPLE INPUT FILE:
#
# line_identifier,question1,another_question,What's the answer?
```

```perl
#  1,1,4,42
#  2,2,1,0
#  3,1,1,42
#  ???,2,123,27
#  anything_at_all,2,123,21
#
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #

use strict;

my @splitData;                     # data is put here later

# no question has more than MAX_ANSWERS answers
use constant MAX_ANSWERS => 20;

# Check for command-line arguments
my $temp = shift @ARGV;
open (MYFILE, "<$temp") or die "Couldn't open $temp for reading.";

# Read data from file
my @tempFileData  = <MYFILE>;
close MYFILE;

# Main code section - call each necessary subroutine
my @tempList = sortData(\@tempFileData);
@tempList = analyzeData(\@tempList);
@tempList = formatResults(\@tempList);
@tempList = sortResults(\@tempList);
printResults(\@tempList);

# sortData (@array): sorts data into a two-dimensional array
#    arguments: @array of raw data to sort. Each index of the array should
#               contain a string of comma-seperated-values.
#    NOTE: The @array argument MUST be passed by reference, not value!
#          Ex: sortData (\@myArray);
#    Return value: sortData will return a two-dimensional array
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
sub sortData {
      # If the @array argument is omitted, display an error.
      if ($#_ < 0) { die "ERROR: sortData needs to be passed an array.\n"; }

      my @rawData = @{$_[0]};         # Take in the @array reference
      my @returnData;                 # Used to return the array

      # Organize data
      for (my $i = 0; $i < $#rawData + 1; $i++) {
          my @tempData = split /,/, $rawData[$i];
          $returnData[$i] = \@tempData;
      }

      return @returnData;             # Return our two-dimensional array
}

# analyzeData (@array): finds recurring patterns in data
#    arguments: @array should be a reference to a two-dimensional array
#               of data to analyze (one that is extracted from a .csv file by
#               sortData)
```

```perl
#    NOTE: The @array argument MUST be passed by reference, not value!
#         Ex: analyzeData (\@myArray);
#    Return value: analyzeData will return a four-dimensional array containing
#                  counts of how many times a pattern was found. To reference
#                  elements in the array:
#               $results[$question1][$question2][value1][value2]
#                  ...where valueX is an answer for questionX.
#                  A pattern consists of 2 values for two different questions.€
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
sub analyzeData {
    my @splitData = @{$_[0]};          # Take in the @array reference
    my @matches;
#increment $corrolator through the array index for each entry on the first line
of the file (except survey_number)
    for (my $corrolator = 1;$corrolator < $#{$splitData[0]};$corrolator++) {
        for (my $corrolatee = $corrolator + 1;$corrolatee <=
$#{$splitData[0]};$corrolatee++) {          #increment $corrolatee through all
possibilities after $corrolatee
            for (my $ref_row_num = 1;$ref_row_num <= $#splitData;$ref_row_num++)
{      #go through each row (survey), to use as a reference for comparison
                my $coror_val = $splitData[$ref_row_num][$corrolator];
        #get reference values
                my $coree_val = $splitData[$ref_row_num][$corrolatee];
                if (!defined
$matches[$corrolator][$corrolatee][$coror_val][$coree_val]) {
    #if we haven't counted this pattern yet
                    $matches[$corrolator][$corrolatee][$coror_val][$coree_val] =
0;    #initialize match count
                    for (my $row_num = 1;$row_num <= $#splitData;$row_num++) {
                        if ($coror_val ==
$splitData[$row_num][$corrolator]and$coree_val ==
$splitData[$row_num][$corrolatee]) {        #if it matches ...

$matches[$corrolator][$corrolatee][$coror_val][$coree_val]++;
                #...count it!
                        }
                    }
                }
            }
        }
    }
    return @matches;                      #YAY! results...
}


# formatResults (@array): formats results from a four-dimensional array for
#                          use in a text file
#   arguments: @array should be a reference to a four-dimensional array
#              of results to print (one that is created by analyzeData)
#   NOTE: The @array argument MUST be passed by reference, not value!
#         Ex: printData (\@myArray);
#   Return value: formatResults will return an array of scalars (strings),
#                 each representing a line of the final output of the program.
#                 each of these strings is in this format:
#
#            (number of matches),$question1,$question2,value1,value2
#
#                 ...where valueX is an answer for questionX
```

```perl
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
sub formatResults {
    my @results = @{$_[0]};            # Take in the @array reference
    my @formattedResults;

    # A 4-level loop to navigate through a 4-dimensional array
    foreach my $i (1..scalar(@results) -2) {
        foreach my $j ($i + 1..scalar(@results)-1) {
            foreach my $k (0..MAX_ANSWERS) {
                foreach my $l (0..MAX_ANSWERS) {
                    if (defined $results[$i][$j][$k][$l]) {
                        push @formattedResults,
"$results[$i][$j][$k][$l],$i,$j,$k,$l";
                    }
                }
            }
        }
    }

    return @formattedResults;
}

# sortResults (@array): sorts and returns formatted results
#    arguments: @array should be a reference to an array of formatted results,
#               as created by &formatData.
#    NOTE: The @array argument MUST be passed by reference, not value!
#         Ex: printData (\@myArray);
#    Return value: formatResults will return an array in the same format as
#                  @array, but with elements sorted by the first number that
#                  appears in them.
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
sub sortResults {
    my @formattedData = @{$_[0]};    # Take in the @array reference
    my @res = sort (sortCriteria @formattedData);
    return @res;

}

# sortCriteria : decides which of two lines of formatted results should come
#                   first
#    NOTE: $a and $b are both taken care of by &sortResults
#    Return value: returns a -1 if $a should come first, a 1 if $b should
#                  come first, or a 0 if they have equal ordering precedence.
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
sub sortCriteria {
        (my $aVal) = ($a =~ /(\d+),/);
        (my $bVal) = ($b =~ /(\d+),/);
        return $bVal <=> $aVal;
}

# printResults (@array): prints formatted results to STDOUT.
#    arguments: @array should be a reference to an array of formatted results,
#               as created by &formatData and &sortData.
#    NOTE: The @array argument MUST be passed by reference, not value!
#         Ex: printData (\@myArray);
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
sub printResults {
```

```perl
    my @formattedData = @{$_[0]};    # Take in the @array reference
    foreach my $lineNumber (@formattedData) {
        print "$lineNumber\n";
    }
}


# printData (@array): prints data in a two-dimensional array
#    arguments: @array should be a reference to a two-dimensional array
#                 of data to print (one that is created by sortData)
#    NOTE: The @array argument MUST be passed by reference, not value!
#          Ex: printData (\@myArray);
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
sub printData {
    # If the @array argument is omitted, display an error.
    if ($#_ < 0) { die "ERROR: printData needs to be passed an array.\n"; }

    my @pData = @{$_[0]};            # Take in the @array reference
    # Get it all out
    foreach my $tempData (@pData) {
        foreach my $data (@{$tempData}) {
            print "$data:";
        }
    }
}
```

nix

```perl
#!/usr/bin/perl -Tw

 #     #     #     #     #
 ##    #     #      #   #
 # #   #     #       ##
 #  #  #     #       ##
 #   ##      #      #   #
 #    #      #     #     #

# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
# A Program for eliminating non-results
#
# nix was written by
# Scott Longley and Kevin D'Aquila
# # # # # # # # # # # # # # # # # # # # # #
# R.E.I.E.F.S. IQP Project
# http://www.project-404.net
# info@project-404.net
#
# This program is designed to remove correlations from the results of miner
# that correlate non-data to something. A "0" in field four or five means that
# no answer was given for that question. If that question represents only one
# element in a check-all-that-apply question, there is no problem, but
# otherwise, that zero represents non-data. Correlations involving non-data
# are not important, so this program outputs every line from the input to
# STDOUT, except those that contain non-data.
# # # # # # # # # # # # # # # # # # # # # # #
# USAGE: ./nix [filename]
#
# If a filename is given to the program, it will attempt to parse the file. If
# multiple filenames are given, all but the first are ignored. If no filename
# is given, STDIN is read in place of the file.
#
# INPUT FILE FORMAT:
# The file is a ".csv" (comma separated value file) of the form that miner
# outputs. Each line of the file represents a correlation that may lead to a
# conclusion. The first value of each line is the number of times (or a
# percentage of the time) that the correlation occured. The second and third
# values in the file are integers or strings that represent questions. The
# fourth and fifth fields are integers that represent the answers to these
# questions.
#
# EXAMPLE INPUT FILE:
#
# 188,47,54,0,0
# 156,7,47,0,0
# 1,63,65,5,0
#
# In this case, only the second line would be removed, because only it contains
# non-data. Here's the output:
#
# EXAMPLE OUTPUT FILE:
#
```

```perl
# 188,47,54,0,0
# 1,63,65,5,0
#
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #

use strict;

my @data;    # Used for data input and manipulation

if (@ARGV) {        # Open the file
    my $fileName = shift(@ARGV);
    open (MYFILE, "<$fileName") or die "Couldn't open $fileName.\n";
    @data = <MYFILE>;
    close MYFILE;
}
else {                          # Use STDIN if there is no file specified
    @data = <STDIN>;
}

foreach my $line (@data) {
    $line =~ /\d*?,(\d*?),(\d*?),(\d*?),(\d*)/;
    # Print line unless "0" is answer for a non check-all-that-apply question
    unless (!(8<$1 and $1<21 or 45<$1 and $1<50 or 50<$1 and $1<57) and $3 == 0
        or !(8<$2 and $2<21 or 45<$2 and $2<50 or 50<$2 and $2<57) and $4 == 0)
{
    print "$line";
    }
}
```

kr

```perl
#!/usr/bin/perl -Tw

 #     #  #####
 #  #  #      #
 ####     #      #
 #  #     #####
 #  #  #  #      #
 #     #  #      #

# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
# Key replacement program
#
# kr was written by
# Kevin D'Aquila
# # # # # # # # # # # # # # # # # # # # # #
# R.E.I.E.F.S. IQP Project
# http://www.project-404.net
# info@project-404.net
#
# This program is designed to replace the integers in fields 2 and 3 of
# every line of the input file with a corresponding key from the keyfile.
# The result is printed to STDOUT.
# # # # # # # # # # # # # # # # # # # # # # # #
# USAGE: ./kr keyfile [filename]
#
# keyfile is a one-line comma-separated-values file. If a filename is
# given to the program, it will attempt to parse the file. If multiple
# filenames are given, all but the first are ignored. If no filename is
# given, STDIN is read in place of the file.
#
# INPUT FILE FORMAT:
# The file is a ".csv" (comma separated value file) of the form that miner
# outputs. Each line of the file represents a correlation that may lead to a
# conclusion. The first value of each line is the number of times (or a
# percentage of the time) that the correlation occured. The second and third
# values in the file are integers or strings that represent questions. The
# fourth and fifth fields are integers that represent the answers to these
# questions.
#
# EXAMPLE INPUT FILE:
#
# 188,47,54,0,0
# 156,7,47,0,0
# 1,63,65,5,0
#
# EXAMPLE OUTPUT FILE:
#
# 188,email_business,yes_other,0,0
# 156,income,email_business,0,0
# 1,nocomp_sports,nocomp_videos,5,0
#
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
```

```perl
use strict;

$/ = undef; # Use 'slurp' mode

my $keyFile;
my @keys;
my $resultsFile;

if (@ARGV == 0) { die "Usage: kr keyfile [resultsfile]\n"; }

# If there is a filename specified, attempt to open it for reading
open (MYFILE, "<$ARGV[0]") or die "Couldn't open $ARGV[0] for reading.";
$keyFile = <MYFILE>;
close MYFILE;

# Convert from comma-delimited to a Perl array
@keys = split ",", $keyFile;

# Check the number of arguments specified
if (@ARGV > 1) {      # Read results from a file if specified
     open (RESFILE, "<$ARGV[1]") or die "Couldn't open $ARGV[1].";
     $resultsFile = <RESFILE>;
     close RESFILE;
} else {              # If results not specified, read from STDIN
     $resultsFile = <STDIN>;
}

# Replace the question numbers with the corresponding keys
$resultsFile =~ s/([^,]*?,)(\d+?),(\d+?)(,.*)/$1$keys[$2],$keys[$3]$4/g;

# Output final results
print $resultsFile;
```

NIX

```
#!/sbin/sh

#       #    ###   #       #
##      #     #     #     #
# #     #     #      # #
#  #    #     #       #
#   #   #     #      # #
#    ## #     #     #   #
#       #    ###   #       #

# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
# Eliminates results that we probably don't care about
#
# NIX was written by
# Scott Longley
# # # # # # # # # # # # # # # # # # # # #
# R.E.I.E.F.S. IQP Project
# http://www.project-404.net
# info@project-404.net
#
# This program gets rid of uninteresting lines.
# Since this script would be 2 lines if not for the comments: 'NUFF SAID!
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #

grep -v "0,0" $1
```

num2pr

```perl
#!/usr/bin/perl -Tw

                        #####
  #    #  #    #  #    # #    # #####  #####
  ##   #  #    #  ##  ## #    # #    # #    #
  # #  #  #    #  # ## # #####  #    # #    #
  #  # #  #    #  #    #    #    #####  #####
  #   ##  #    #  #    #    #    #      #   #
  #    #  ####   #    # #######  #      #    #


# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
# A Program for converting datamining results to express their count as a
#
# num2pr was written by
# Scott Longley and Kevin D'Aquila
# # # # # # # # # # # # # # # # # # # # # # # #
# R.E.I.E.F.S. IQP Project
# http://www.project-404.net
# info@project-404.net
#
# This program is designed to replace an integer that counts number of
# correlations with a percentage of a specified total number of occurences.
# # # # # # # # # # # # # # # # # # # # # # #
# USAGE: ./num2pr total [filename]
#
# total is the total out of which the percentages will be calculated. If
# the data being analyzed originated from 203 surveys, specifying 203 will
# show a the percentage of the total surveys that displayed that
# correlation. If a filename is given to the program, it will attempt to
# parse the file. If multiple filenames are given, all but the first are
# ignored. If no filename is given, STDIN is read in place of the file.
#
# INPUT FILE FORMAT:
# The file is a ".csv" (comma separated value file) of the form that miner
# outputs. Each line of the file represents a correlation that may lead to a
# conclusion. The first value of each line is the number of times (or a
# percentage of the time) that the correlation occured. The second and third
# values in the file are integers or strings that represent questions. The
# fourth and fifth fields are integers that represent the answers to these
# questions.
#
# EXAMPLE INPUT FILE:
#
# 188,47,54,0,0
# 156,7,47,0,0
# 1,63,65,5,0
#
# In this case, only the second line would be removed, because only it contains
# non-data. Here's the output:
#
# EXAMPLE OUTPUT FILE:
#
# 188,47,54,0,0
```

```
#  1,63,65,5,0
#
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #

use strict;

my @data;                       # Used for data input and manipulation
my $total = shift(@ARGV);       # User specified

if (@ARGV) {         # Open the file
    my $fileName = shift(@ARGV);
    open (MYFILE, "<$fileName") or die "Couldn't open $fileName.\n";
    @data = <MYFILE>;
    close MYFILE;
}
else {                          # Use STDIN if there is no file specified
    @data = <STDIN>;
}

# Convert the cardinality (number of correlations) to a percentage
foreach my $line (@data) {
    $line =~ /(\d*?)(,.*)/;
    my $p = ($1*100)/$total;
    my $theRest = $2;
    $p =~ s/([^\.]\.[\d]{0,2}).*/$1/;
    print "$p\%$theRest\n";
}
```

genex

```perl
#!/usr/bin/perl -Tw

   ####    ######  #     #  ######  #     #
  #    #   #       ##    #  #        #   #
  #        #####   # #   #  #####     ##
  #  ###   #       #  #  #  #         ##
  #    #   #       #   ## #  #        #   #
   ####    ######  #     #  ######  #     #

# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
# Generic Extractor program
#
# genex was written by
# Kevin D'Aquila
# # # # # # # # # # # # # # # # # # # # #
# R.E.I.E.F.S. IQP Project
# http://www.project-404.net
# info@project-404.net
#
# The Generic Extractor program may be used to search the output of
# the miner program. Input to this program, via a file or STDIN, must
# be in the format output from miner (5 comma-delimited values per line).
#
# Switches 1 through 5 may be used to specify a regular expression to use
# for searching the results. Only matching results will be returned.
#
#     ex) ./genex -1 15 -2 "[^\d]*" -5 "3" resultsfile.txt
#                  -or-
#          ./miner mydata.csv | kr keys.csv | genex -1 20
# # # # # # # # # # # # # # # # # # # # # # # #
# USAGE: ./genex [-1 regex] ... [-5 regex] [filename]
#
# If a filename is given to the program, it will attempt to parse the file. If
# multiple filenames are given, all but the first are ignored. If no filename
# is given, STDIN is read in place of the file.
#
# INPUT FILE FORMAT:
# The file is a ".csv" (comma separated value file) of the form that miner
# outputs. There should be 5 values in each line.
#
# EXAMPLE INPUT FILE:
#
# 188,47,54,0,0
# 156,7,47,0,3
# 1,63,65,5,0
# 3,24,31,1,2
#
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #

use strict;

my $errorMsg = "Usage: genex [-1 regex1] [-2 regex2] [-3 regex3] " .
               "[-4 regex4] [-5 regex5] [file]\n";
```

```perl
my @patterns;       # Stores regex's passed via switches 1-5
my $fileName;       # Stores the file to read data from, if any
my @data;    # Used for data input and manipulation

if (@ARGV < 1) { die $errorMsg; }    # Make sure there are arguments

# Parse the command arguments
for (my $i = 0; $i < @ARGV; $i = $i + 2) {
     if ($ARGV[$i] =~ /^-(\d+)/) {
           # Make sure the syntax is correct
           unless ($ARGV[$i + 1]) { die $errorMsg; }

           # In case a switch is specified more than once
           if ($patterns[$1 - 1]) {
                die "Specify each argument only once.\n"; }

           # Switches must be between 1 and 5
           if ($1 < 1 or $1 > 5) {
                die "You may only specify arguments 1 through 5.\n"; }

           # If it's valid, save the switch argument
           $patterns[$1 - 1] = $ARGV[$i + 1];
     }
     # This block recognizes a filename, if one has been specified
     elsif ($ARGV[$i] =~ /^([^-].*)/) {
           $fileName = $1;
           $i--;
     }
     # If we get here, there's an error
     else { die $errorMsg; }
}

# Fill in non-specified switches with a regex to capture anything
for (my $i = 0; $i < 5; $i++) {
     unless (defined $patterns[$i]) { $patterns[$i] = ".*?"; }
}

if (defined $fileName) {        # Open the file
     open (MYFILE, "<$fileName") or die "Couldn't open $fileName.\n";
     @data = <MYFILE>;
     close MYFILE;
}
else {                          # Use STDIN if there is no file specified
     @data = <STDIN>;
}

# Parse each line of data
foreach my $line (@data) {
     if ($line =~ /($patterns[0],  # Try to match
            $patterns[1],          # each pattern
            $patterns[2],          # specified by
            $patterns[3],          # the user's switches
            $patterns[4]$)/x)
     {
                print "$1\n";
     }
}
```

11

Appendix D

Data

Data from Surveys
Data Output from miner
Data Output from nix
Data Output from kr
Data Output from NIX
Data Output from num2pr

Data from Surveys

Survey Data: Key File

Data from each survey are stored in simple comma-separated value files (.csv). Keys that respectively correspond to data are as follows:

```
survey_number,survey_school,survey_grade,age,sex,religion,politics,income,parent
s_engineer,parents_education,parents_social,parents_arts,parents_gov,parents_man
agement,parents_trade,parents_homemaking,parents_med,parents_labor,parents_datae
nt,parents_business,parents_other,favsub,lfavsub,location,length_residence,influ
ence,necessity,firstcompuse_age,comp_access,num_systems,oldest_comp,hours_max,os
,computer_type,email_account,internet_access,connection,hours_wordprocessing,hou
rs_presentations,hours_games,hours_music,hours_movies,hours_browsing,hours_onlin
e,hours_total,email_personal,email_education,email_business,email_other,spam,yes
_timeonline,yes_sites,yes_games,yes_talk,yes_other,no_rules,acceptable_use,guard
ian_hours,household_knowledge,helped_you,nocomp_friends,nocomp_telephone,nocomp_
newspapers,nocomp_sports,nocomp_music,nocomp_videos,nocomp_hobbies
```

If these keys are prepended to a data file, they will serve as labels for each line of data (note that for the numeric data below, each two lines denotes a single survey):

```
survey_number,survey_school,survey_grade,age,sex,religion,politics,income,parent
s_engineer,parents_education,parents_social,parents_arts,parents_gov,parents_man
agement,parents_trade,parents_homemaking,parents_med,parents_labor,parents_datae
nt,parents_business,parents_other,favsub,lfavsub,location,length_residence,influ
ence,necessity,firstcompuse_age,comp_access,num_systems,oldest_comp,hours_max,os
,computer_type,email_account,internet_access,connection,hours_wordprocessing,hou
rs_presentations,hours_games,hours_music,hours_movies,hours_browsing,hours_onlin
e,hours_total,email_personal,email_education,email_business,email_other,spam,yes
_timeonline,yes_sites,yes_games,yes_talk,yes_other,no_rules,acceptable_use,guard
ian_hours,household_knowledge,helped_you,nocomp_friends,nocomp_telephone,nocomp_
newspapers,nocomp_sports,nocomp_music,nocomp_videos,nocomp_hobbies
1,1,4,8,2,4,4,0,1,0,0,0,0,0,0,0,0,0,0,0,0,4,1,2,4,1,1,2,1,3,1,3,1,1,1,1,2,3,0,0,
0,0,1,3,0,1,1,1,1,6,0,0,0,0,0,1,1,3,3,1,0,0,0,0,0,0,0
2,1,4,9,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,5,4,2,3,1,1,2,2,4,2,3,1,3,1,2,1,1,1,2,
2,2,2,3,4,0,0,0,1,5,0,1,0,1,0,0,1,2,2,4,5,2,2,4,2,3,3
3,1,4,9,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,2,4,2,3,2,1,2,2,2,2,3,4,2,2,2,1,3,2,3,
1,2,1,1,3,0,0,0,0,0,0,0,0,0,1,2,4,2,1,4,2,2,4,5,3,0
4,1,4,10,2,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,5,3,2,3,2,1,1,2,4,3,5,4,3,1,2,3,3,1,5
,4,3,4,5,5,1,0,0,0,2,0,1,0,0,0,0,2,1,2,4,4,3,2,4,3,5,4
5,1,4,12,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,4,6,2,2,1,1,2,3,3,4,3,5,3,2,4,3,3,1,1
,1,1,2,2,4,0,0,0,1,7,0,0,0,0,1,0,2,3,0,5,4,1,2,3,3,4,3
```

. . .
```

Data from Survey Site 1: A single survey consists of data as seen below, every two lines.

1,1,4,8,2,4,4,0,1,0,0,0,0,0,0,0,0,0,0,0,0,4,1,2,4,1,1,2,1,3,1,3,1,1,1,1,2,3,0,0,
0,0,1,3,0,1,1,1,1,6,0,0,0,0,0,1,1,3,3,1,0,0,0,0,0,0,0,0
2,1,4,9,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,5,4,2,3,1,1,2,2,4,2,3,1,3,1,2,1,1,1,2,
2,2,2,3,4,0,0,0,1,5,0,1,0,1,0,0,1,2,2,4,5,2,2,4,2,3,3
3,1,4,9,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,2,4,2,3,2,1,2,2,2,2,3,4,2,2,2,1,3,2,3,
1,2,1,1,3,0,0,0,0,0,0,0,0,0,0,1,2,4,2,1,4,2,2,4,5,3,0
4,1,4,10,2,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,5,3,2,3,2,1,1,2,4,3,5,4,3,1,2,3,3,1,5
,4,3,4,5,5,1,0,0,0,2,0,1,0,0,0,0,2,1,2,4,4,3,2,4,3,5,4
5,1,4,12,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,4,6,2,2,1,1,2,3,3,4,3,5,3,2,4,3,3,1,1
,1,1,2,2,4,0,0,0,1,7,0,0,0,0,1,0,2,3,0,5,4,1,2,3,3,4,3
6,1,4,9,1,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,4,3,2,3,1,1,2,2,5,1,6,1,3,1,2,2,2,1,4,
3,4,1,4,6,1,0,0,0,1,0,0,0,0,0,1,1,6,2,4,3,2,1,4,1,3,2
7,1,4,9,2,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,4,2,2,3,2,3,2,2,5,3,3,1,3,2,2,3,2,2,3,
2,1,2,3,4,1,0,0,0,3,0,0,0,1,0,0,1,5,2,4,4,2,2,3,2,4,4
8,1,4,10,2,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,4,5,2,3,2,1,2,2,3,3,4,1,3,2,4,3,2,3,3
,2,1,2,3,4,1,0,0,0,2,1,0,0,1,0,0,1,5,2,4,4,2,3,4,2,3,3
9,1,4,9,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,6,1,2,3,1,3,1,2,2,4,4,4,3,2,2,3,2,2,4,
2,1,2,2,4,1,1,0,0,2,0,1,0,0,0,0,1,5,2,4,4,3,2,5,3,6,3
10,1,4,10,2,0,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,0,5,2,3,2,1,1,2,4,3,3,0,3,2,2,3,2,2,
4,2,3,3,3,5,0,0,0,0,0,0,0,0,0,1,2,0,4,4,4,3,2,5,4,5,3
11,1,4,9,2,0,0,0,1,1,1,0,0,0,0,0,0,0,1,0,0,6,3,2,4,1,0,2,1,5,3,3,1,1,1,2,3,1,1,3
,1,3,3,2,4,1,0,0,0,2,0,0,1,0,0,0,2,4,2,4,2,3,2,3,3,3,3
12,1,4,10,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,4,3,2,3,1,1,2,2,4,3,2,1,3,2,2,3,2,1,
2,1,1,1,2,2,0,0,0,0,0,0,1,0,0,0,0,1,3,2,4,2,2,1,3,2,2,2
13,1,4,9,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,5,2,4,2,1,2,2,5,1,3,4,3,1,2,1,2,2,3
,2,1,2,2,4,1,0,0,0,3,0,1,0,0,0,0,1,6,2,4,4,2,1,3,2,4,2
14,1,4,10,2,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,5,3,2,3,2,1,1,2,4,3,3,0,1,2,0,0,2,2,
3,1,1,2,3,3,0,0,0,0,0,0,0,0,0,1,2,3,2,0,4,3,1,4,2,3,3
15,1,4,9,2,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,5,2,3,1,1,1,2,3,3,2,5,3,1,2,3,2,3,2
,2,3,2,3,4,1,0,0,0,1,0,1,0,0,0,0,1,3,2,5,3,2,2,3,2,2,4
16,1,4,10,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,2,4,2,3,1,1,1,2,5,2,0,4,3,1,2,4,2,3,
5,4,4,5,4,6,0,0,0,0,2,0,0,0,0,1,0,1,6,1,3,4,3,2,3,2,4,6
17,1,4,9,2,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1,4,1,2,1,1,1,1,4,1,3,4,3,1,1,3,3,3,4
,5,5,1,5,6,1,0,0,0,2,0,0,0,1,0,0,1,6,2,4,6,4,3,3,5,5,3
18,1,4,10,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,4,3,2,3,2,1,1,2,5,2,2,4,3,2,2,3,2,2,
3,1,3,3,4,5,1,1,0,0,1,0,0,0,1,0,0,1,5,2,5,3,2,3,3,4,3,4
19,1,4,10,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,5,3,2,1,2,1,1,1,2,1,2,1,1,1,2,3,1,1,
3,1,2,1,2,2,1,0,0,0,2,0,0,0,0,0,1,2,4,1,4,6,2,3,5,2,3,1
20,1,4,10,2,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,5,3,2,3,2,1,2,3,3,2,4,1,3,2,1,3,1,1,
4,3,1,3,4,5,1,0,0,0,2,0,0,0,1,0,0,1,3,1,2,2,2,1,4,3,6,2
21,1,4,10,2,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,3,2,2,2,1,1,1,4,1,3,4,3,0,2,0,2,2,
3,2,1,3,3,2,1,0,0,0,7,0,0,0,0,0,1,1,3,2,4,4,3,2,4,3,5,4
22,1,4,9,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,4,2,2,3,1,1,2,2,4,4,3,4,3,2,2,1,2,2,3
,1,1,3,3,4,1,0,0,0,5,0,0,0,0,0,1,1,6,2,4,4,1,2,4,2,3,4

Data from Survey Site 2: A single survey consists of data as seen below, every two lines.

```
1,2,10,16,2,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,6,1,2,1,1,1,2,2,4,2,5,1,3,1,2,4,3,3,
2,3,3,3,4,4,1,1,0,0,3,1,0,0,0,0,0,1,5,3,5,4,3,3,3,3,3,3
2,2,10,16,1,1,6,3,0,0,0,0,0,0,0,0,0,0,1,0,0,2,1,2,3,1,1,3,2,4,3,3,4,2,1,1,2,2,1,
2,4,2,4,5,4,1,1,0,0,2,0,0,0,0,0,1,1,5,2,4,5,3,2,5,4,4,3
3,2,10,16,1,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,5,3,2,3,1,1,1,1,5,2,4,3,1,1,1,4,3,2,
2,3,3,4,3,4,1,1,1,0,4,1,0,0,0,0,0,1,6,1,4,3,2,3,3,2,3,3
4,2,10,16,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,4,2,4,2,1,1,2,5,3,4,4,3,1,2,4,3,1,
5,4,3,3,5,5,1,1,0,0,1,0,0,0,0,0,1,1,5,1,2,5,2,3,5,4,6,4
5,2,10,15,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,3,6,2,2,1,1,1,1,5,2,6,1,1,1,1,4,4,2,
4,5,4,5,5,6,1,1,0,0,3,0,0,0,0,0,1,2,3,1,2,4,3,3,3,4,3,4
6,2,10,14,1,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,1,5,2,2,1,1,2,1,4,1,4,4,2,1,1,4,3,1,
1,4,2,3,4,4,1,1,0,0,5,0,0,0,0,0,1,2,6,2,4,5,2,3,3,4,3,4
7,2,10,15,2,0,0,0,1,0,0,0,0,0,0,1,1,0,0,0,0,2,3,2,3,2,1,2,2,5,1,4,1,3,1,1,3,3,1,
2,4,3,3,4,4,1,0,0,0,2,0,0,0,0,0,1,2,5,2,5,5,4,3,3,5,4,3
8,2,10,16,2,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,2,4,2,3,1,1,2,1,4,3,3,1,1,1,1,2,3,2,
3,3,3,3,3,3,1,0,0,0,4,0,1,0,0,0,2,3,1,5,4,3,2,3,3,3,4
9,2,10,15,2,2,6,0,0,0,0,1,0,0,0,0,1,0,0,0,0,6,4,2,3,1,1,2,2,2,3,5,2,3,1,2,1,3,1,
2,1,2,4,5,5,1,0,0,0,5,1,0,0,0,0,2,5,1,2,5,2,3,2,6,1,4
10,2,10,15,1,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,1,4,2,3,1,1,2,2,3,3,4,4,3,1,2,4,3,2
,1,4,3,4,5,6,1,0,0,0,5,0,0,0,0,0,1,2,6,1,3,5,3,4,5,4,4,6
11,2,10,16,1,6,1,0,0,0,1,0,0,0,0,0,1,0,0,0,0,3,4,2,3,2,3,2,1,5,3,3,5,3,1,1,1,3,1
,1,3,1,3,3,3,1,1,0,0,3,0,0,0,0,0,1,1,3,3,5,3,3,3,4,3,2,3
12,2,10,15,1,1,2,3,0,0,1,0,0,0,0,0,0,0,0,0,0,2,4,2,3,2,1,2,2,3,1,2,1,3,1,2,1,1,1
,1,1,1,1,2,2,1,0,0,0,4,0,0,0,0,0,1,1,4,2,4,2,1,1,1,6,3
13,2,10,16,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,6,4,2,4,2,1,2,2,3,1,5,4,3,1,2,1,3,2
,2,4,1,4,5,6,1,1,0,0,4,0,0,0,0,0,1,1,5,2,4,5,4,3,4,4,5,5
14,2,10,16,1,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,2,1,2,3,2,3,2,1,5,3,4,6,3,1,1,4,2,2
,3,3,3,4,2,4,1,1,0,0,2,0,1,0,0,0,1,5,1,4,3,2,2,3,2,3,3
15,2,10,16,1,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0,6,5,2,3,2,1,2,1,5,2,3,1,3,1,1,4,2,1
,3,1,1,0,3,3,1,1,0,0,3,0,0,0,0,0,1,1,6,2,4,4,3,3,5,3,4,3
16,2,10,16,2,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,6,1,2,1,1,1,2,2,2,2,5,4,3,1,2,4,3,1
,2,5,5,2,5,5,1,0,0,0,4,1,0,0,0,0,1,3,1,7,2,2,1,1,4,3,3
17,2,10,15,1,1,1,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,6,4,3,2,1,1,1,4,3,4,4,3,1,1,1,3,2
,4,4,3,5,5,6,1,0,0,0,2,1,0,0,0,0,1,2,1,7,4,3,2,4,4,4,4
18,2,10,15,2,0,0,0,0,0,0,1,1,0,0,0,1,0,0,0,0,3,4,2,1,1,3,2,2,3,1,4,2,3,1,1,4,2,3
,2,3,1,3,4,4,1,1,1,0,1,1,0,0,0,0,2,5,1,3,5,3,3,4,4,4,4
19,2,10,15,1,0,0,0,1,0,0,0,1,0,1,0,0,0,0,0,0,2,4,3,2,2,3,2,2,4,2,2,4,3,1,2,1,2,1
,2,1,1,2,2,3,1,0,0,0,4,0,0,0,0,0,1,1,3,1,5,5,4,4,4,5,4,4
20,2,10,15,1,7,1,2,1,0,0,1,0,0,0,0,1,0,0,0,0,6,5,2,3,1,1,2,1,5,1,4,1,3,1,1,4,3,2
,3,3,4,3,5,5,1,1,0,0,4,1,0,0,0,0,1,5,1,3,4,2,3,4,3,3,4
21,2,10,15,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,4,3,2,3,2,1,2,1,3,2,4,4,1,1,2,4,4,3
,4,2,1,4,4,4,1,1,0,0,2,0,0,1,0,0,1,5,2,4,3,3,2,1,1,4,4
22,2,10,15,2,2,1,5,1,0,0,1,0,0,0,0,0,0,0,0,0,5,4,2,4,1,1,1,1,5,1,6,4,3,1,1,1,5,1
,1,5,1,4,6,6,1,1,0,0,4,0,0,0,0,0,1,1,5,1,2,4,5,2,1,5,4,4
23,2,10,15,1,0,0,0,0,0,1,1,0,0,0,1,0,0,0,1,0,5,2,2,4,1,1,2,1,5,3,4,1,1,1,1,1,3,2
,2,3,2,2,3,4,1,1,0,0,2,0,0,0,0,0,1,1,6,2,4,4,2,3,3,5,3,3
24,2,10,15,1,0,6,4,0,0,0,1,0,0,0,0,0,0,1,0,0,6,4,2,3,0,3,2,1,5,3,3,1,1,1,1,4,3,2
,2,3,2,3,3,3,1,1,0,0,2,1,0,0,0,0,1,6,2,4,3,2,2,2,3,3,4
```

Data from Survey Site 3: A single survey consists of data as seen below, every two lines.

```
1,3,7,12,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,5,2,3,3,1,1,2,1,3,1,5,1,3,1,1,1,3,2,3
,4,3,4,5,6,1,0,0,0,3,1,0,0,0,0,0,1,2,1,2,3,2,3,4,4,3,5
2,3,7,12,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,5,3,3,3,1,1,2,2,3,1,4,5,3,1,2,4,3,2,3
,4,1,3,4,4,1,0,1,0,2,1,0,0,0,0,0,1,5,2,3,3,2,2,3,3,4,4
3,3,7,12,1,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,5,1,2,3,1,1,2,1,3,2,4,1,3,1,1,4,2,1,4
,3,2,3,4,4,1,0,0,0,2,0,1,0,0,0,0,2,4,2,4,4,2,2,3,3,3,2
4,3,7,12,1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,5,3,2,4,2,1,2,2,3,3,4,1,3,1,2,4,2,1,2
,2,2,2,2,4,1,0,0,0,2,1,0,0,0,0,0,2,5,2,2,3,1,2,5,3,5,2
5,3,7,13,1,0,0,0,0,0,0,0,0,1,1,0,0,0,0,1,0,4,3,3,3,1,1,1,2,2,2,5,1,3,1,2,4,3,3,3
,4,3,4,5,6,1,0,0,0,4,1,1,0,0,0,0,2,4,1,3,5,4,3,4,3,6,3
6,3,7,13,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,5,6,2,4,1,1,2,2,3,2,4,1,3,1,2,4,2,1,4
,4,2,2,4,4,1,0,0,0,1,0,1,0,0,0,0,1,3,3,5,4,2,3,4,3,4,1
7,3,7,12,2,0,0,0,0,1,1,0,0,1,0,0,0,0,0,1,0,3,2,2,4,2,3,2,2,4,2,3,4,1,2,2,3,3,2,2
,1,2,1,2,4,0,1,0,0,7,1,0,0,0,0,0,1,4,4,5,3,2,2,3,3,2,2
8,3,7,12,2,0,0,0,1,1,0,0,0,0,0,1,0,0,0,1,0,5,0,2,2,1,1,2,2,3,3,3,4,3,1,2,4,2,3,2
,1,1,1,2,3,1,0,0,0,1,1,1,1,1,1,0,1,5,2,4,3,3,3,5,3,3,2
9,3,7,12,1,0,0,0,0,0,0,0,1,1,0,1,0,0,0,0,0,3,2,4,1,3,1,2,4,2,4,4,1,1,2,4,2,3,3
,2,1,3,3,4,1,0,0,0,1,0,1,0,0,0,0,2,6,2,4,5,2,1,4,2,3,5
10,3,7,12,1,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,3,2,3,1,1,2,1,5,1,5,1,1,1,1,4,3,2,
3,3,1,3,4,5,1,1,0,1,1,0,1,0,0,0,0,2,6,2,5,4,3,1,4,2,3,3
11,3,7,12,1,0,0,0,1,0,0,1,1,1,1,0,0,1,1,1,0,5,1,2,3,1,3,1,1,3,2,4,2,3,1,1,1,3,2,
4,2,1,3,3,4,1,0,0,0,3,1,0,0,0,0,0,2,6,2,4,3,2,3,4,3,3,4
12,3,7,13,2,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,6,3,2,3,1,1,2,2,3,3,4,5,2,1,2,4,2,1,
3,1,1,3,4,4,1,0,0,0,3,1,0,0,0,0,0,1,4,2,4,3,3,3,4,2,3,3
13,3,7,13,2,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,5,0,2,2,1,1,2,1,3,1,4,1,1,1,1,1,4,2,
2,2,3,3,2,4,1,0,0,0,3,0,1,0,0,0,0,1,3,1,3,1,1,2,3,2,3,3
14,3,7,13,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,6,2,2,2,2,1,2,1,5,2,4,1,1,1,1,4,2,1,
4,2,3,2,1,4,1,0,0,0,2,0,0,0,0,0,1,1,6,2,5,3,2,2,3,2,3,4
15,3,7,13,2,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0,6,4,2,4,1,1,2,2,4,3,5,4,3,1,2,4,2,1,
1,2,1,5,5,5,1,0,0,0,2,1,0,0,0,0,0,2,6,2,4,3,2,1,4,3,3,2
16,3,7,13,2,0,0,0,1,0,0,1,0,0,0,0,1,0,0,1,0,5,1,2,4,1,1,1,2,2,3,5,4,3,1,2,4,4,2,
3,3,1,3,3,5,1,0,0,0,2,0,0,0,0,0,1,2,6,1,3,4,4,3,3,4,2,4
17,3,7,12,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,5,6,2,3,1,1,2,2,3,1,4,1,3,1,2,4,3,2,
3,3,1,2,2,4,1,0,0,0,2,1,1,0,1,0,0,1,4,2,4,3,1,2,4,2,3,3
18,3,7,12,2,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,2,0,2,4,2,3,1,2,5,3,3,4,2,1,1,4,3,2,
2,1,1,2,2,3,1,1,0,0,2,1,0,0,0,0,0,1,6,2,4,3,3,3,3,2,3,4
19,3,7,12,1,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,5,3,2,2,2,1,2,2,3,6,5,2,2,2,2,3,3,4,
5,1,1,5,5,5,1,0,0,0,7,0,1,0,0,0,0,1,6,2,4,4,3,2,6,2,5,5
20,3,7,12,1,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0,4,3,2,3,1,1,2,2,2,2,4,1,3,1,2,3,3,2,
3,1,1,4,4,4,0,0,0,0,1,1,1,1,1,0,0,2,5,2,4,3,2,3,5,2,3,3
21,3,7,13,1,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,5,3,2,3,2,1,1,2,3,3,4,1,2,1,2,4,1,2,
4,1,1,2,4,4,0,0,0,0,1,0,1,0,0,0,0,1,6,2,4,3,3,3,1,2,3,3
```

Data from Survey Site 4: A single survey consists of data as seen below, every two lines.

```
1,4,4,10,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,2,4,2,3,1,1,2,2,3,2,3,4,3,2,3,3,1,1,3
,1,1,1,1,3,1,0,0,0,2,1,0,0,0,0,0,1,4,2,4,5,2,2,2,3,3,4
2,4,4,10,1,2,6,1,0,0,0,0,0,0,0,0,0,0,0,1,0,2,2,3,1,1,2,2,4,1,6,1,6,4,2,3,7,1,1,1
,1,1,1,1,1,0,0,0,1,7,0,0,0,0,0,1,2,7,7,7,1,1,1,1,1,1,1
3,4,4,10,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,4,1,3,3,1,3,2,2,3,1,1,1,3,2,2,3,1,1,2
,2,1,1,1,3,0,0,0,1,1,0,1,0,0,0,0,2,3,4,5,4,3,2,4,2,4,3
4,4,4,10,2,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,2,1,2,2,1,1,2,2,2,4,2,4,3,2,2,3,1,1,2
,1,1,1,2,2,0,0,0,1,7,1,0,0,0,0,0,1,2,2,4,3,2,1,1,2,3,3
5,4,4,10,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,2,1,2,2,1,1,2,1,3,1,2,4,2,1,1,3,2,1,3
,1,1,1,2,4,1,0,0,0,2,0,0,0,0,0,1,2,3,2,4,1,2,1,1,1,2,1
6,4,4,10,2,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,4,1,2,3,1,2,1,2,3,3,3,1,3,0,2,3,3,1,2
,3,1,3,3,3,1,0,0,0,4,1,1,0,0,0,0,2,4,4,5,4,2,2,3,2,3,4
7,4,4,9,2,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,4,0,2,4,1,0,2,2,2,4,3,0,3,2,2,3,3,1,3,
2,1,2,1,4,0,0,0,1,1,0,1,0,0,0,0,2,3,2,5,4,2,1,2,1,3,5
8,4,4,10,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,5,4,2,4,1,2,1,2,3,4,4,2,2,2,3,2,1,4
,3,3,1,1,6,1,0,0,0,2,0,0,1,0,0,0,1,0,0,0,3,2,1,3,1,3,1
9,4,4,10,1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,5,4,2,2,1,1,2,2,2,3,2,1,2,2,2,1,1,1,1
,1,1,1,2,3,1,0,0,0,1,0,1,0,0,0,0,1,3,2,4,4,2,3,6,1,6,6
10,4,4,10,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,5,3,2,3,1,0,2,1,5,1,3,4,1,1,2,1,3,1,
5,3,4,3,3,4,1,0,0,0,1,0,0,0,0,0,1,1,6,2,4,4,3,2,5,3,4,2
11,4,4,9,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,4,1,1,2,0,0,2,2,2,3,3,1,3,1,2,7,3,0,3
,2,1,2,3,3,1,0,0,0,1,0,0,0,0,0,1,1,3,2,4,2,2,1,3,2,2,3
12,4,4,9,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,2,4,3,3,1,3,1,2,2,3,1,4,3,2,3,3,1,1,2
,1,2,1,1,1,0,0,0,1,1,0,0,1,0,0,0,2,6,2,4,4,2,1,4,2,5,3
13,4,4,10,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,5,2,2,4,1,2,2,2,3,1,2,4,3,2,2,1,2,1,
2,1,1,1,1,2,0,0,0,1,1,1,0,0,0,0,1,3,2,4,4,2,2,3,2,3,3
14,4,4,9,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,5,0,2,0,0,0,0,2,2,5,0,0,0,2,3,3,1,1,2
,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,2,1,4,2,0,1,1,2,2,1
15,4,4,10,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,4,2,2,4,1,1,1,2,3,2,4,4,3,1,2,4,1,1,
3,2,4,3,3,4,1,0,0,0,2,0,1,0,0,0,0,1,2,3,5,3,2,1,4,2,4,2
16,4,4,9,1,0,0,4,0,1,0,0,0,0,0,0,0,0,0,0,0,0,4,2,0,3,1,0,1,2,3,3,2,0,3,2,2,3,1,2,4
,2,4,2,1,3,0,0,0,0,1,0,0,1,0,0,0,1,3,2,4,2,2,1,7,2,4,6
17,4,4,9,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,4,2,2,4,0,0,2,2,4,3,3,0,1,2,2,0,1,0,2
,2,2,2,3,4,0,0,0,0,1,0,0,0,0,0,1,5,2,4,4,3,2,3,4,2,0
18,4,4,10,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,5,4,2,4,0,0,1,1,5,1,4,0,3,2,2,4,0,0,
3,0,0,0,3,4,0,0,0,0,1,0,0,0,0,0,1,0,2,2,4,4,4,2,4,5,3,3
19,4,4,10,2,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,5,4,3,4,0,0,2,1,5,3,2,4,3,1,1,3,2,1,
3,2,4,2,2,5,1,0,0,0,1,0,0,0,0,0,1,1,3,2,4,3,2,2,1,3,3,3
20,4,4,9,2,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,5,4,2,2,0,0,2,2,2,4,3,0,3,1,2,3,3,2,3
,2,2,2,2,2,1,0,0,0,1,1,1,0,0,0,0,0,2,1,4,2,2,2,3,2,2,3
21,4,4,10,2,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,5,4,2,3,1,1,2,2,2,4,4,3,1,2,1,2,3,
4,1,2,3,3,4,1,0,0,0,1,0,1,0,0,0,0,1,4,2,4,3,3,4,3,4,4,3
22,4,4,10,2,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,5,4,2,4,1,0,1,2,3,2,4,4,3,2,2,4,2,1,
4,5,3,4,3,6,1,0,0,0,1,1,0,0,0,0,0,2,4,4,5,4,5,3,6,5,2,6
23,4,4,10,2,1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,6,4,2,3,0,0,2,2,2,5,2,4,2,2,2,3,1,1,
3,2,2,3,3,4,0,1,0,0,1,0,1,0,0,0,0,2,3,2,4,3,2,1,2,1,3,2
24,4,4,10,2,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,1,1,2,4,1,2,2,2,2,1,4,4,3,1,2,3,3,2,
5,4,3,6,5,6,1,0,0,0,3,0,0,0,1,0,0,1,4,1,4,1,1,1,1,1,1,1
25,4,4,9,1,1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,5,2,2,4,3,3,1,2,3,3,2,4,3,2,2,4,3,1,5
,3,6,3,4,6,1,0,0,0,2,0,0,0,0,0,1,2,3,4,4,4,5,1,6,3,5,2
26,4,4,10,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,5,2,2,3,2,1,2,2,3,3,2,4,3,1,2,3,1,1,
3,1,3,2,2,3,1,0,0,0,2,0,0,1,0,0,0,2,5,2,4,1,1,1,4,1,4,3
27,4,4,9,2,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,4,2,2,3,2,1,2,2,4,1,2,1,3,1,2,3,1,1,2
,2,1,2,3,3,1,1,0,0,2,1,0,0,0,0,1,2,4,5,3,2,1,3,3,3,3
```

Data from Survey Site 5: A single survey consists of data as seen below, every two lines.

1,5,10,16,1,1,2,4,1,0,0,0,0,0,0,0,0,0,0,0,0,2,4,3,4,1,1,2,2,2,3,4,1,3,1,2,1,3,1,
4,1,1,3,3,4,1,0,0,0,2,0,1,0,0,0,0,1,3,2,4,4,2,0,3,3,3,3
2,5,10,17,2,1,6,6,1,0,1,0,0,0,1,0,0,0,0,1,0,1,4,3,3,1,1,3,2,2,2,5,4,3,1,2,4,3,2,
1,5,2,4,5,5,1,0,0,0,2,0,0,0,0,0,1,2,3,1,2,4,2,2,4,3,3,2
3,5,10,16,1,1,2,6,0,0,0,0,0,1,0,0,1,0,0,0,0,2,5,3,3,1,1,2,2,3,2,3,4,3,1,2,4,3,1,
2,2,1,3,0,4,1,0,0,0,4,0,0,0,1,0,0,1,3,3,2,4,2,4,5,3,3,4
4,5,10,15,1,5,2,2,0,0,0,0,0,0,1,0,0,0,0,0,0,6,3,3,2,1,2,2,2,3,3,3,4,3,1,2,1,2,1,
1,1,1,3,3,3,1,0,0,0,2,0,0,0,0,0,1,1,4,1,2,4,2,2,4,3,3,2
5,5,10,17,1,1,2,4,0,0,0,0,0,0,1,0,1,0,0,0,0,2,5,3,4,1,2,2,1,3,4,3,4,3,1,1,1,3,0,
0,0,0,0,3,3,1,0,0,0,2,1,0,0,0,0,0,1,4,1,2,4,2,3,3,2,3,2
6,5,10,17,2,7,1,3,0,0,1,0,0,0,0,0,0,0,0,0,1,0,4,2,3,1,1,3,2,1,5,4,6,4,3,1,2,3,4,1,
4,5,1,6,6,6,1,0,0,0,2,0,0,0,0,0,1,1,5,4,5,4,2,2,2,3,4,3
7,5,10,17,2,1,6,4,0,0,0,0,0,0,0,0,0,0,0,1,0,2,5,3,3,1,1,2,2,3,3,4,4,2,1,2,4,4,2,
1,4,1,4,4,5,1,1,0,0,3,0,0,0,0,0,1,2,4,1,2,5,3,2,5,3,2,4
8,5,10,16,2,1,6,5,0,0,1,0,0,0,0,0,0,1,0,0,0,1,2,3,3,1,2,3,2,3,2,3,1,3,1,2,4,2,2,
1,2,1,2,3,3,1,1,0,0,2,0,0,0,0,0,1,2,3,2,2,5,3,2,4,2,3,3
9,5,10,16,2,1,6,0,1,0,0,0,0,0,0,0,1,0,0,0,0,6,4,3,3,1,1,3,2,3,3,4,4,2,1,1,1,3,2,
1,3,1,3,4,4,1,0,0,0,5,1,0,1,0,1,0,2,3,2,4,5,4,3,4,4,2,4
10,5,10,17,2,0,6,1,0,0,0,0,0,0,0,0,0,1,1,0,0,5,1,3,4,1,1,2,1,5,4,6,5,3,1,1,4,3,1
,0,5,2,3,6,6,1,0,0,0,2,0,0,0,0,0,1,1,6,2,3,5,3,2,3,3,4,3
11,5,10,16,2,1,1,3,0,1,0,0,0,0,1,0,0,0,0,1,0,6,4,3,3,1,1,2,1,4,2,4,4,3,1,1,4,3,2
,1,4,1,4,4,4,1,0,0,0,4,0,0,0,0,0,1,2,6,2,4,3,3,2,3,4,3,2
12,5,10,15,2,5,3,3,0,0,1,0,0,1,0,0,0,0,0,1,0,1,4,3,4,1,2,1,2,4,4,4,4,2,1,2,4,3,2
,1,3,1,3,3,4,1,1,0,0,3,0,0,0,0,0,1,1,4,2,5,4,2,3,1,2,3,3
13,5,10,16,1,6,4,3,1,1,0,0,0,0,0,0,0,0,0,0,2,4,2,3,3,3,2,2,3,1,3,1,3,1,2,1,2,1
,1,2,1,2,2,3,1,0,0,0,3,0,0,0,0,0,1,1,2,1,2,5,2,2,5,4,3,3
14,5,10,17,1,1,6,2,0,0,0,0,0,0,0,0,1,0,0,1,0,6,4,3,4,1,2,2,1,3,3,4,1,2,1,1,4,2,1
,2,3,3,4,4,4,1,0,0,0,3,0,0,0,0,0,1,2,7,1,3,3,2,2,3,2,2,4
15,5,10,17,1,1,1,4,0,0,0,0,0,1,0,0,0,1,0,1,0,2,3,3,4,1,1,2,2,4,3,4,4,3,1,2,4,2,2
,3,4,3,3,3,5,1,1,0,0,3,0,0,0,0,0,1,1,3,1,1,4,3,2,3,5,3,3
16,5,10,17,1,2,2,5,0,1,1,0,0,1,0,0,0,0,0,1,0,2,6,2,4,0,2,3,1,4,1,5,6,3,1,1,3,3,1
,1,2,2,4,4,4,1,1,1,0,7,0,0,0,0,0,1,1,7,3,5,4,2,1,1,1,4,5
17,5,10,17,2,1,5,4,0,0,0,0,0,1,0,0,0,0,0,0,0,6,4,3,2,1,2,3,2,2,3,2,4,3,2,1,1,2,0
,0,2,0,0,2,2,1,0,0,0,2,0,0,0,0,0,1,1,3,2,4,5,3,2,2,3,3,4

Data from Survey Site 6: A single survey consists of data as seen below, every two lines.

```
1,6,11,17,1,2,2,6,0,0,0,0,0,1,1,0,1,0,0,1,0,2,5,3,3,1,1,2,2,3,2,3,4,3,1,2,4,3,2,
1,3,1,3,3,3,1,0,0,0,5,0,0,0,0,0,1,1,4,2,4,4,3,3,5,3,3,4
2,6,11,17,1,1,6,0,0,1,0,0,0,0,0,0,0,0,1,0,2,4,3,2,1,2,2,2,3,3,4,4,3,1,2,4,3,1,
2,5,2,3,4,6,1,0,0,0,3,0,0,0,0,0,1,1,4,3,5,5,4,3,3,6,4,5
3,6,11,18,2,1,6,4,1,0,0,0,0,1,1,0,0,1,1,1,0,5,2,3,4,1,2,3,2,2,3,3,4,3,1,2,4,3,1,
3,4,3,3,3,5,1,1,0,0,5,0,0,0,0,0,1,2,6,1,2,4,2,2,4,4,3,4
4,6,11,17,2,1,6,4,1,0,0,0,0,1,0,0,0,0,0,0,0,5,3,3,3,1,1,2,2,2,3,3,4,3,1,2,1,3,1,
1,1,1,1,2,3,1,0,0,0,3,0,0,0,0,0,1,2,1,4,2,4,2,2,4,2,3,3
5,6,11,17,2,1,6,2,0,0,0,0,0,0,0,0,0,0,0,1,0,4,2,3,4,1,1,3,1,2,2,3,4,3,1,1,4,3,1,
1,3,1,2,3,3,1,0,0,0,4,0,0,0,0,0,1,2,1,1,2,4,3,2,2,2,3,4
6,6,11,17,2,1,6,0,0,0,0,0,0,0,0,0,0,0,0,0,1,5,3,3,4,1,2,2,2,3,2,5,1,3,1,2,1,3,2,
2,6,2,5,5,6,1,1,0,0,5,0,0,0,0,0,1,1,4,1,2,4,4,2,5,6,3,3
7,6,11,18,1,1,4,4,1,0,0,0,1,0,0,0,0,0,0,0,0,6,4,3,3,1,1,2,1,4,1,3,1,3,1,1,4,2,2,
2,3,2,2,3,4,1,0,0,0,3,0,0,0,0,0,1,2,4,1,2,6,4,2,4,4,3,4
8,6,11,17,1,1,1,5,0,0,0,0,1,0,0,0,0,0,0,0,0,5,4,3,3,1,1,2,1,4,1,4,4,2,1,1,4,2,1,
4,3,3,2,3,4,1,1,0,0,2,0,0,0,0,0,1,1,3,1,2,3,4,3,4,3,3,4
9,6,11,17,1,1,1,5,1,0,0,0,0,1,0,0,0,0,0,0,0,1,3,3,3,2,2,2,1,3,2,5,1,3,1,1,4,4,3,
2,3,3,3,5,5,1,1,0,0,4,0,0,0,0,0,1,1,5,1,2,5,3,0,6,3,4,3
10,6,11,16,1,5,6,5,1,0,0,0,0,0,0,0,0,0,0,0,0,4,1,3,4,1,1,2,1,2,3,5,4,3,1,1,4,3,1
,3,5,3,4,1,6,1,0,0,0,2,0,0,0,0,0,1,2,2,1,2,6,2,3,6,5,5,4
11,6,11,17,2,1,1,1,0,0,0,0,0,0,0,0,0,1,1,0,0,0,6,5,3,3,1,2,2,1,3,1,4,4,3,1,1,4,2,2
,4,0,1,3,4,5,1,1,0,0,4,0,0,0,0,0,1,2,4,1,2,4,3,2,5,4,4,3
12,6,11,17,2,1,6,3,0,0,0,0,0,0,0,0,0,1,0,0,0,1,4,3,3,1,1,3,1,4,1,3,2,1,1,1,1,2,1
,2,5,1,3,0,4,1,0,0,0,4,0,0,0,0,0,1,2,1,3,7,4,4,3,2,5,3,4
13,6,11,17,2,6,6,5,1,0,0,0,0,1,0,1,0,0,0,0,0,6,2,3,3,1,2,1,1,4,3,3,4,3,1,1,1,3,2
,3,0,2,3,3,4,1,1,0,0,4,1,0,0,0,0,0,1,4,1,3,4,3,1,2,2,3,4
14,6,11,16,2,1,3,4,0,0,1,0,1,0,0,0,0,0,0,0,0,1,5,3,2,1,2,3,2,3,1,3,1,3,1,2,4,2,2
,1,1,1,2,3,3,1,0,0,0,3,0,0,0,0,0,1,2,3,3,5,4,4,3,3,3,4,4
15,6,11,17,1,1,2,6,0,1,0,0,0,0,0,0,0,0,0,1,0,2,3,3,3,1,1,1,2,4,3,4,4,3,1,2,1,3,2
,4,3,2,3,4,4,1,0,0,0,3,1,0,0,0,0,1,4,1,2,4,3,3,4,4,4,3
16,6,11,16,2,1,2,4,1,0,0,0,0,0,0,1,0,0,0,0,4,5,3,3,1,1,2,2,3,3,5,4,2,1,2,4,4,2
,1,4,2,3,6,6,1,1,0,0,5,0,0,0,0,0,1,1,4,2,2,3,3,1,5,3,3,2
17,6,11,17,1,1,3,0,1,0,0,0,1,0,0,0,0,0,0,0,2,3,3,3,1,1,3,2,2,3,2,1,3,1,2,1,2,0
,0,0,0,0,2,0,1,0,0,0,2,0,0,0,0,0,1,1,6,2,4,5,5,3,5,2,2,2
```

Data from Survey Site 7: A single survey consists of data as seen below, every two lines.

```
1,7,10,16,2,0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,0,6,3,3,3,1,1,2,2,2,4,5,1,3,1,2,3,3,1,
3,3,1,2,4,4,1,0,0,0,2,0,0,0,0,0,1,2,3,1,2,4,3,2,4,2,3,4
2,7,10,16,2,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,6,4,2,4,1,1,2,2,5,1,3,1,3,1,2,1,2,1,
1,2,1,2,2,3,1,1,0,0,4,0,0,0,0,0,1,2,4,2,5,5,4,1,3,4,5,5
3,7,10,15,2,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,4,5,3,4,1,1,2,2,2,3,3,1,3,1,2,1,3,3,
2,2,1,2,3,3,1,0,0,0,4,0,0,0,0,0,1,1,3,4,1,6,3,2,6,6,6,3
4,7,10,16,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,2,3,2,4,1,1,2,2,2,3,3,4,3,1,2,1,2,1,
2,3,1,2,3,3,0,1,0,0,3,0,0,0,0,0,1,2,2,1,1,5,2,2,3,3,3,3
5,7,10,15,1,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,2,3,2,4,1,1,2,2,2,4,3,1,3,1,2,4,2,2,
2,2,1,3,2,3,1,0,0,0,3,0,0,0,0,0,1,1,4,1,1,5,2,2,5,3,5,4
6,7,10,15,2,0,0,0,1,0,0,0,0,0,0,0,0,1,0,1,0,3,4,3,4,1,1,3,2,2,3,4,4,3,1,2,1,3,1,
2,4,1,3,4,6,1,1,0,0,4,0,1,0,0,0,0,1,6,2,2,4,5,2,3,5,3,3
7,7,10,16,1,7,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,6,4,3,3,1,1,2,1,5,1,4,1,3,1,1,1,4,2,
2,5,4,4,4,6,1,0,0,0,2,0,0,0,0,0,1,1,5,1,4,4,3,4,3,4,4,4
8,7,10,16,2,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,0,5,2,3,3,1,1,2,2,2,3,5,1,3,1,2,1,3,2,
2,3,1,3,5,5,1,0,0,0,3,0,0,0,0,0,1,2,3,1,2,5,5,3,3,4,4,3
9,7,10,16,1,1,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,4,3,3,1,2,2,1,2,4,4,4,3,1,1,1,2,1,
3,2,1,4,4,4,1,0,0,0,2,0,0,0,0,0,1,1,2,1,3,3,1,2,2,2,3,4
10,7,10,15,1,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,2,3,2,3,1,1,3,2,2,3,3,4,3,1,2,4,2,1
,2,2,1,3,2,3,1,1,0,0,4,0,0,0,0,0,1,2,3,1,1,4,2,2,4,3,4,5
11,7,10,15,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,4,1,3,2,1,1,3,1,3,1,3,1,3,2,2,1,2,1
,3,2,1,3,3,4,1,0,0,0,7,1,1,0,0,0,0,2,3,2,1,4,5,2,4,3,4,3
12,7,10,16,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,4,3,2,3,1,1,2,1,2,2,6,1,3,1,1,1,3,1
,0,6,2,3,5,6,1,0,0,0,4,0,1,0,0,0,0,1,3,1,4,4,2,2,3,4,5,5
13,7,10,16,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,4,3,3,4,1,2,2,1,2,1,4,1,3,1,1,1,2,2
,3,2,1,3,3,4,1,0,0,0,4,0,0,0,0,0,1,1,2,1,1,5,3,1,6,4,4,3
14,7,10,16,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,4,2,3,3,1,1,2,2,3,2,4,4,3,1,2,1,3,2
,4,4,3,4,4,4,1,1,0,0,4,0,1,0,0,0,0,1,3,3,2,5,3,2,5,4,4,5
15,7,10,15,2,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0,6,4,2,3,1,2,3,2,2,3,3,4,3,1,2,1,2,7
,1,6,3,3,3,3,1,0,0,0,3,0,0,0,0,0,1,0,2,1,2,5,4,2,5,6,3,3
16,7,10,16,1,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,4,5,2,3,1,1,2,1,3,1,5,4,3,1,1,1,2,2
,3,4,5,5,6,6,1,0,0,0,3,0,0,0,0,0,1,2,3,1,2,6,3,3,3,5,5,4
17,7,10,16,1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,4,2,3,1,1,1,2,2,2,4,3,4,3,1,2,4,3,1
,3,3,1,3,3,4,1,0,0,0,3,0,0,0,0,0,1,1,3,1,1,3,2,3,4,3,3,3
18,7,10,15,2,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,6,2,3,3,1,1,3,1,3,3,4,1,3,1,1,1,3,2
,3,3,1,4,4,4,1,1,0,0,2,0,0,0,0,0,1,2,2,1,1,4,4,3,2,5,3,5
19,7,10,16,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,4,3,3,4,1,1,2,1,5,1,4,4,3,1,1,1,3,3
,4,2,2,3,4,4,1,1,0,0,5,0,0,0,0,0,1,2,3,1,7,4,2,2,3,3,3,3
```

Data from Survey Site 8: A single survey consists of data as seen below, every two lines.

1,8,10,16,1,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,1,2,2,3,1,1,2,2,3,3,4,4,1,1,2,3,2,3,
4,0,2,2,2,3,1,1,0,0,4,0,1,0,0,0,0,1,5,3,2,2,0,1,3,3,6,5
2,8,10,16,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,5,4,3,2,1,1,3,1,4,2,2,4,2,1,1,1,2,1,
1,2,2,1,1,2,1,0,0,0,2,0,0,0,0,0,1,2,2,2,4,3,1,2,3,3,3,1
3,8,10,16,2,0,0,0,0,1,1,0,1,0,1,0,0,0,0,0,0,6,4,2,3,1,2,3,2,2,2,4,4,2,1,2,1,2,1,
3,4,1,3,4,4,1,0,0,0,4,0,0,0,0,0,1,1,3,1,4,6,3,3,1,3,3,6
4,8,10,16,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,5,2,4,1,2,5,4,1,6,1,6,4,2,3,3,1,1,
1,1,1,1,1,1,0,0,0,1,1,0,0,0,0,0,1,2,1,7,7,5,1,1,1,3,3,5
5,8,10,16,1,0,0,0,1,0,0,0,0,1,1,0,0,1,0,1,0,2,4,3,3,1,1,2,2,3,1,3,1,3,1,2,1,1,1,
2,2,1,2,2,3,1,0,0,0,3,0,0,0,0,0,1,1,3,1,2,4,3,2,1,3,3,1
6,8,10,16,2,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,2,4,3,3,1,1,3,1,3,1,5,1,3,2,1,4,2,1,
3,4,1,4,5,7,1,0,0,0,3,0,0,0,1,0,0,1,3,3,2,4,5,1,2,4,4,2
7,8,10,15,1,7,7,6,1,0,0,0,0,0,1,0,0,1,0,0,0,2,0,3,4,1,1,2,1,3,3,3,1,3,1,1,1,2,1,
1,3,3,3,3,3,1,0,1,0,4,0,0,0,0,0,1,2,6,1,2,6,2,2,3,4,4,4
8,8,10,16,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,4,3,4,1,2,3,2,2,4,2,4,2,2,1,4,2,1,
2,3,1,2,2,3,1,0,0,0,7,0,0,0,0,0,1,2,3,1,1,4,2,2,3,3,4,4
9,8,10,15,1,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,2,3,2,3,1,2,3,1,2,1,4,1,3,1,1,4,2,1,
4,3,1,2,2,4,1,0,0,0,3,0,0,0,0,0,1,1,3,1,2,4,2,1,5,2,3,3
10,8,10,15,1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,2,4,2,4,1,2,3,2,2,4,3,4,3,1,1,1,2,1
,3,2,1,3,3,4,1,0,0,0,4,0,0,0,0,0,1,2,4,1,2,4,2,1,3,2,6,6
11,8,10,16,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,6,4,1,3,1,1,5,1,2,2,2,6,3,1,2,3,1,1
,1,1,1,1,2,2,0,0,0,0,7,0,0,0,0,0,1,2,3,4,4,6,3,2,3,3,2,4
12,8,10,16,2,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1,3,2,3,1,1,3,2,2,1,4,6,2,1,2,1,2,1
,4,6,3,2,3,5,1,0,0,0,5,0,0,0,0,0,1,2,5,1,1,6,4,1,1,6,3,5
13,8,10,16,1,1,2,0,0,0,0,0,1,0,0,0,0,0,0,0,4,1,2,3,1,1,1,1,5,4,3,1,1,1,1,1,3,1
,3,1,1,2,2,4,1,0,0,0,1,1,0,0,0,0,0,1,6,2,4,6,2,3,4,4,3,6
14,8,10,16,2,1,1,0,0,0,0,0,0,0,1,0,0,0,0,0,1,2,4,2,4,1,1,3,1,2,4,6,1,2,1,1,4,2,1
,3,4,1,5,6,6,1,0,0,0,2,1,0,0,0,0,0,2,1,1,7,4,3,1,1,6,3,2
15,8,10,16,2,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,3,5,2,3,1,2,2,2,3,1,5,4,3,1,2,4,2,1
,3,3,3,3,4,5,1,0,0,0,4,0,0,0,0,0,1,2,3,2,4,1,3,1,3,3,3,1
16,8,10,16,2,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0,1,2,2,2,2,1,3,1,3,3,4,4,3,1,1,4,3,3
,2,1,1,3,4,5,1,1,0,0,4,1,0,0,1,0,0,1,5,2,4,5,5,3,3,3,4,3
17,8,10,16,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,5,3,3,1,2,3,1,2,3,3,4,3,1,1,1,1,1
,3,3,3,1,3,3,0,0,0,0,2,0,0,0,0,0,1,2,1,1,2,5,2,1,2,2,1,2
18,8,10,17,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,2,4,3,4,1,2,3,2,2,2,2,1,3,1,2,1,2,1
,2,1,2,3,3,3,1,0,0,0,1,0,1,0,0,0,0,1,5,1,2,6,5,5,3,6,6,6
19,8,10,16,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,4,1,3,4,1,2,2,2,3,3,3,4,3,1,2,4,2,1
,2,3,3,3,3,5,1,0,0,0,3,0,0,0,0,0,1,2,3,3,5,6,2,1,5,4,2,6

Data from Survey Site 9: A single survey consists of data as seen below, every two lines.

```
1,9,7,12,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,2,1,2,3,2,1,2,2,2,3,3,1,3,1,2,1,3,2,3
,3,0,3,3,4,1,0,0,0,4,0,0,0,1,0,0,2,3,1,1,2,0,0,0,0,0,0
2,9,7,13,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,3,2,4,1,1,2,2,2,4,3,1,3,2,2,3,3,1,3
,2,1,3,3,4,1,1,0,0,3,1,0,0,0,0,0,1,5,3,0,1,3,3,4,4,3,5
19,9,7,13,2,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,5,4,2,3,1,2,2,2,2,4,5,4,3,1,2,3,1,1,
3,4,4,3,4,5,1,0,0,0,1,0,1,0,1,0,0,1,3,3,5,4,3,2,4,3,3,4
3,9,7,12,2,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,2,5,2,2,1,1,2,2,2,3,3,5,3,1,2,3,0,0,2
,3,3,3,4,4,1,0,0,0,4,1,0,0,1,0,0,2,4,1,2,6,5,3,4,4,4,4
4,9,7,13,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,2,1,2,3,1,1,1,2,2,3,3,6,3,1,2,4,1,2,2
,3,1,2,2,3,1,0,0,0,4,0,1,0,0,0,0,2,3,3,4,0,0,0,0,0,0,0
5,9,7,12,2,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,2,1,2,3,1,2,2,2,3,1,3,6,3,2,2,3,3,1,4
,1,3,2,1,1,1,0,0,0,1,0,1,0,0,0,0,2,3,3,2,4,3,2,5,4,3,5
20,9,7,15,2,0,6,1,0,0,1,0,0,0,0,0,0,0,0,0,1,0,2,5,1,4,1,1,1,2,2,6,4,4,3,2,2,1,3,1,
3,2,1,1,2,3,0,0,0,0,1,0,0,0,0,1,0,2,2,2,4,3,2,3,1,3,2,1
6,9,7,13,2,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,5,1,2,3,1,2,3,2,2,2,4,1,3,1,2,3,0,1,3
,1,1,1,3,4,1,0,0,0,2,0,1,0,0,0,0,1,5,2,4,3,3,1,3,1,3,1
7,9,7,13,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,6,1,2,3,1,2,2,1,4,1,5,1,3,1,1,4,2,2,3
,3,2,3,3,5,1,0,0,0,4,1,0,0,0,0,0,1,4,1,7,3,2,2,2,3,2,2
21,9,7,13,2,1,7,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1,3,3,1,2,1,2,2,4,6,1,3,1,2,3,2,3,
6,1,6,4,6,4,1,0,0,0,3,0,0,0,0,0,1,0,6,1,4,6,3,1,6,4,6,6
8,9,7,12,1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,4,3,2,4,1,1,1,2,4,4,5,4,3,1,0,0,3,1,4
,3,1,2,3,5,1,1,0,0,4,0,0,0,0,0,1,1,3,1,3,4,3,1,5,4,4,2
9,9,7,13,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,2,4,2,4,1,2,2,1,3,3,2,4,1,1,3,3,2,0,2
,2,1,1,1,3,0,0,0,0,4,0,1,0,0,0,0,1,2,2,4,4,2,1,5,4,0,5
22,9,7,12,1,1,2,0,0,0,0,0,1,0,0,0,0,0,0,0,0,6,1,3,1,1,2,1,1,4,4,5,4,3,1,1,1,1,2,
3,2,3,4,4,4,1,0,0,0,2,0,1,0,0,0,0,1,6,1,3,1,3,1,4,2,4,3
10,9,7,12,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,2,5,2,3,1,1,2,2,3,1,5,4,3,2,2,2,3,2,
2,4,7,2,5,6,1,0,0,0,1,0,1,0,0,0,0,2,3,3,5,3,4,2,3,4,4,2
11,9,7,12,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,4,3,2,4,1,2,2,1,3,3,2,1,3,2,2,4,2,3,
2,2,2,2,3,3,0,0,0,0,7,0,1,0,0,0,0,1,4,4,5,4,1,1,3,2,3,3
23,9,7,13,2,1,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,0,6,4,3,4,1,2,2,2,3,3,3,1,3,1,2,3,2,2,
3,2,0,4,4,4,1,0,0,0,1,0,1,0,0,0,0,2,5,2,4,4,2,3,4,2,5,5
12,9,7,12,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,6,3,2,3,1,1,2,2,3,3,2,4,3,2,2,3,2,0,
2,2,1,2,2,3,1,0,0,0,3,0,1,0,1,0,0,2,5,2,4,3,2,1,4,2,1,4
13,9,7,13,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,5,6,2,4,1,2,2,2,3,3,3,4,2,2,2,1,2,2,
1,1,1,2,3,3,1,0,0,0,5,0,0,0,1,0,0,1,3,2,4,3,2,2,4,3,3,4
24,9,7,13,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,6,5,2,4,1,1,2,1,3,1,3,0,3,1,1,2,2,2,
3,3,3,3,3,3,1,0,0,0,2,1,0,0,0,0,1,4,3,5,4,3,2,6,3,5,5
14,9,7,12,1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,2,1,2,2,2,1,2,2,2,3,3,1,0,1,2,3,1,1,
2,3,3,0,3,3,0,0,0,0,5,0,1,0,0,0,0,2,3,1,2,5,4,0,0,0,0,0
15,9,7,13,2,0,0,0,0,0,0,0,1,0,0,1,0,0,1,0,2,1,2,4,1,3,2,2,3,3,2,4,3,1,2,3,1,1,
2,3,1,1,1,3,1,0,0,0,1,1,0,0,0,0,0,2,3,2,4,4,2,1,2,4,3,2
25,9,7,13,2,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,6,4,2,3,1,2,1,2,2,3,5,4,3,0,2,1,0,2,
5,4,1,3,3,7,1,0,0,0,1,1,1,0,1,0,0,1,1,1,1,5,2,3,4,6,2,4
16,9,7,13,1,0,0,0,0,1,1,0,0,0,1,0,0,0,0,0,2,1,2,4,1,1,2,2,2,3,3,1,1,1,2,1,4,1,
6,4,1,5,4,6,1,0,0,0,4,0,0,0,0,0,1,2,6,1,3,1,3,4,4,3,4,4
17,9,7,13,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,5,1,2,4,1,2,3,2,2,0,4,1,3,1,2,3,3,3,
4,6,4,5,6,0,1,0,0,0,4,0,0,0,0,1,0,1,7,4,0,0,0,0,0,0,0,0
26,9,7,12,2,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,3,3,4,2,1,3,2,2,4,6,0,3,2,2,3,4,1,
2,5,1,2,6,6,1,0,0,0,1,1,1,1,1,0,0,1,3,1,4,5,3,2,3,4,2,3
18,9,7,13,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,5,1,2,4,1,1,1,1,3,2,4,1,2,1,1,1,3,3,
3,3,3,3,3,4,1,0,0,0,1,0,0,0,0,0,1,2,6,1,4,4,4,2,5,4,4,6
27,9,7,14,2,1,3,6,0,0,1,0,0,0,0,0,0,1,0,0,1,6,6,2,1,2,1,1,1,4,1,6,1,1,1,1,2,4,3,
4,5,4,6,6,6,1,1,0,0,4,0,0,0,0,0,1,2,6,1,2,6,5,2,4,6,6,1
```

28,9,7,13,1,1,3,0,0,0,0,1,0,0,0,0,0,0,0,0,0,6,4,2,2,2,2,2,2,3,4,3,1,3,1,2,1,2,2,
2,3,1,3,3,4,1,0,0,0,3,0,0,0,1,0,0,1,4,1,4,5,3,3,3,3,3,4
32,9,7,12,1,1,1,6,0,0,1,0,0,0,0,0,0,0,0,0,0,2,1,3,2,1,1,2,2,2,3,3,4,3,1,2,1,2,2,
3,3,3,3,4,5,1,0,0,0,3,0,0,0,0,0,1,1,3,1,2,4,3,2,3,4,5,3
29,9,7,13,2,1,6,0,1,0,0,0,0,0,0,0,0,0,0,0,1,2,1,2,3,1,2,1,1,2,4,4,4,3,1,1,1,2,2,
2,2,1,4,4,4,1,1,0,0,1,1,0,0,0,0,0,1,4,2,4,2,3,2,2,3,3,3
30,9,7,13,2,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,4,1,3,4,1,1,2,1,3,2,5,4,3,1,1,1,4,4,
5,5,5,4,6,6,1,1,1,0,2,0,1,0,0,0,0,2,4,1,1,6,4,3,6,5,5,6
33,9,7,12,2,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,6,4,3,3,2,2,0,2,3,1,3,4,3,1,2,3,2,3,
0,3,3,3,3,3,1,1,0,0,2,1,1,0,1,0,0,1,3,2,4,5,4,4,4,3,4,5
31,9,7,13,2,1,1,0,1,0,0,0,0,0,0,0,0,0,0,0,6,1,3,3,1,2,2,2,3,3,4,1,2,1,2,4,3,1,
2,3,1,3,3,4,1,0,0,0,1,0,0,0,1,1,0,1,4,2,4,4,3,4,4,3,3,4
34,9,7,13,2,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,5,4,3,3,1,2,3,1,2,0,2,6,0,1,1,3,2,1,
1,2,2,2,3,3,1,0,0,0,4,1,1,1,1,0,0,1,3,2,1,4,3,2,4,3,3,3
35,9,7,12,2,1,6,0,0,0,0,0,0,0,1,0,0,0,0,0,0,4,5,2,1,1,1,2,1,3,4,5,1,3,1,1,3,3,2,
3,4,3,4,5,5,1,1,0,0,2,0,1,0,0,0,0,1,4,2,2,4,5,2,2,6,6,3
37,9,7,13,2,1,6,0,1,0,0,0,0,0,0,1,0,0,0,0,0,6,4,2,3,1,1,2,1,2,3,4,4,3,1,1,4,2,1,
2,2,3,2,2,4,1,0,0,0,4,1,1,0,1,0,0,1,3,1,2,6,3,1,3,3,5,4
36,9,7,13,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,2,1,2,3,1,1,1,1,4,1,5,1,2,1,1,4,3,2,
4,5,4,5,4,6,1,0,1,0,3,1,1,0,0,0,0,1,6,1,4,3,2,1,4,6,3,5

Data output from miner

Each line represents a single correlation made by miner. The first field is the cardinality of the correlation (how many times it occurs through out our entire data set of 203 surveys). Second is the first question number (correlator). Third is the second question number (correlatee). These numbers reference our key listing. The fourth field is the answer to the correlator's question, and the fifth and final field is the answer to the correlatee's question. These results are sorted according to their cardinality, from highest to lowest. The full results file is 32637 lines long. A sample of the output is shown below.

```
188,47,54,0,0
187,52,54,0,0
186,48,54,0,0
185,18,47,0,0
185,18,54,0,0
185,47,48,0,0
184,47,52,0,0
183,11,47,0,0
183,11,54,0,0
183,15,54,0,0
183,18,52,0,0
183,20,54,0,0
182,18,48,0,0
182,20,47,0,0
182,48,52,0,0
181,11,18,0,0
181,12,47,0,0
181,12,54,0,0
181,15,47,0,0
181,20,48,0,0


...


1,63,64,1,4
1,63,64,1,5
1,63,64,6,1
1,63,64,6,2
1,63,64,6,6
1,63,64,7,2
1,63,65,1,6
1,63,65,2,2
1,63,65,2,6
1,63,65,4,1
1,63,65,5,0
1,63,65,6,2
1,63,65,7,4
1,64,65,1,2
1,64,65,2,6
1,64,65,4,0
1,64,65,5,2
1,64,65,6,1
1,64,65,6,2
1,64,65,6,4
(not a full data listing)
```

Data output from nix

nix is used to remove correlations made with NULL (0) values on non-multiple choice survey questions. It may be used on data set output from miner, kr, NIX, num2pr, or a combination thereof.

```
188,47,54,0,0
187,52,54,0,0
186,48,54,0,0
185,18,47,0,0
185,18,54,0,0
185,47,48,0,0
184,47,52,0,0
183,11,47,0,0
183,11,54,0,0
183,15,54,0,0
183,18,52,0,0
183,20,54,0,0
182,18,48,0,0
182,20,47,0,0
182,48,52,0,0
181,11,18,0,0
181,12,47,0,0
181,12,54,0,0
181,15,47,0,0
181,20,48,0,0
...

1,62,65,3,1
1,62,65,5,6
1,63,64,1,4
1,63,64,1,5
1,63,64,6,1
1,63,64,6,2
1,63,64,6,6
1,63,64,7,2
1,63,65,1,6
1,63,65,2,2
1,63,65,2,6
1,63,65,4,1
1,63,65,6,2
1,63,65,7,4
1,64,65,1,2
1,64,65,2,6
1,64,65,5,2
1,64,65,6,1
1,64,65,6,2
1,64,65,6,4
(not a full data listing)
```

Data Output from kr

kr is the "key replace" script. If run on a data set output from miner, nix, NIX, num2pr, or a combination thereof, it will replace the second and third fields with the textual keys that they reference. A sample of the output is shown below.

```
188,email_business,yes_other,0,0
187,yes_games,yes_other,0,0
186,email_other,yes_other,0,0
185,parents_dataent,email_business,0,0
185,parents_dataent,yes_other,0,0
185,email_business,email_other,0,0
184,email_business,yes_games,0,0
183,parents_arts,email_business,0,0
183,parents_arts,yes_other,0,0
183,parents_homemaking,yes_other,0,0
183,parents_dataent,yes_games,0,0
183,parents_other,yes_other,0,0
182,parents_dataent,email_other,0,0
182,parents_other,email_business,0,0
182,email_other,yes_games,0,0
181,parents_arts,parents_dataent,0,0
181,parents_gov,email_business,0,0
181,parents_gov,yes_other,0,0
181,parents_homemaking,email_business,0,0
181,parents_other,email_other,0,0

...

1,nocomp_sports,nocomp_music,1,4
1,nocomp_sports,nocomp_music,1,5
1,nocomp_sports,nocomp_music,6,1
1,nocomp_sports,nocomp_music,6,2
1,nocomp_sports,nocomp_music,6,6
1,nocomp_sports,nocomp_music,7,2
1,nocomp_sports,nocomp_videos,1,6
1,nocomp_sports,nocomp_videos,2,2
1,nocomp_sports,nocomp_videos,2,6
1,nocomp_sports,nocomp_videos,4,1
1,nocomp_sports,nocomp_videos,5,0
1,nocomp_sports,nocomp_videos,6,2
1,nocomp_sports,nocomp_videos,7,4
1,nocomp_music,nocomp_videos,1,2
1,nocomp_music,nocomp_videos,2,6
1,nocomp_music,nocomp_videos,4,0
1,nocomp_music,nocomp_videos,5,2
1,nocomp_music,nocomp_videos,6,1
1,nocomp_music,nocomp_videos,6,2
1,nocomp_music,nocomp_videos,6,4
(not a full data listing)
```

Data Output from NIX

NIX is a simple script to remove NULL to NULL correlations. These are numerous in multiple choice questions. NIX may be used on data set output from miner, nix, kr, num2pr, or a combination thereof.

```
172,email_personal,email_other,1,0
170,email_personal,yes_other,1,0
166,email_personal,email_business,1,0
166,email_personal,yes_games,1,0
165,parents_dataent,email_personal,0,1
163,parents_other,email_personal,0,1
161,parents_arts,email_personal,0,1
161,parents_homemaking,email_personal,0,1
159,parents_gov,email_personal,0,1
158,parents_labor,email_personal,0,1
157,parents_education,email_personal,0,1
155,email_account,email_other,1,0
153,email_account,yes_other,1,0
151,email_account,yes_games,1,0
151,email_personal,yes_talk,1,0
150,email_account,email_business,1,0
149,parents_dataent,email_account,0,1
149,influence,email_business,1,0
149,influence,yes_other,1,0
149,email_account,email_personal,1,1
...

1,nocomp_newspapers,nocomp_videos,3,1
1,nocomp_newspapers,nocomp_videos,5,6
1,nocomp_sports,nocomp_music,1,4
1,nocomp_sports,nocomp_music,1,5
1,nocomp_sports,nocomp_music,6,1
1,nocomp_sports,nocomp_music,6,2
1,nocomp_sports,nocomp_music,6,6
1,nocomp_sports,nocomp_music,7,2
1,nocomp_sports,nocomp_videos,1,6
1,nocomp_sports,nocomp_videos,2,2
1,nocomp_sports,nocomp_videos,2,6
1,nocomp_sports,nocomp_videos,4,1
1,nocomp_sports,nocomp_videos,6,2
1,nocomp_sports,nocomp_videos,7,4
1,nocomp_music,nocomp_videos,1,2
1,nocomp_music,nocomp_videos,2,6
1,nocomp_music,nocomp_videos,5,2
1,nocomp_music,nocomp_videos,6,1
1,nocomp_music,nocomp_videos,6,2
1,nocomp_music,nocomp_videos,6,4
(not a full data listing)
```

Data Output from num2pr

num2pr replaces the cardinality field (the first field) with a percentage. The percentage is computed according to a user-provided parameter specifying the total number of surveys. num2pr may be used on data set output from miner, nix, kr, NIX, or a combination thereof.

```
84.72%,email_personal,email_other,1,0
83.74%,email_personal,yes_other,1,0
81.77%,email_personal,email_business,1,0
81.77%,email_personal,yes_games,1,0
81.28%,parents_dataent,email_personal,0,1
80.29%,parents_other,email_personal,0,1
79.31%,parents_arts,email_personal,0,1
79.31%,parents_homemaking,email_personal,0,1
78.32%,parents_gov,email_personal,0,1
77.83%,parents_labor,email_personal,0,1
77.33%,parents_education,email_personal,0,1
76.35%,email_account,email_other,1,0
75.36%,email_account,yes_other,1,0
74.38%,email_account,yes_games,1,0
74.38%,email_personal,yes_talk,1,0
73.89%,email_account,email_business,1,0
73.39%,parents_dataent,email_account,0,1
73.39%,influence,email_business,1,0
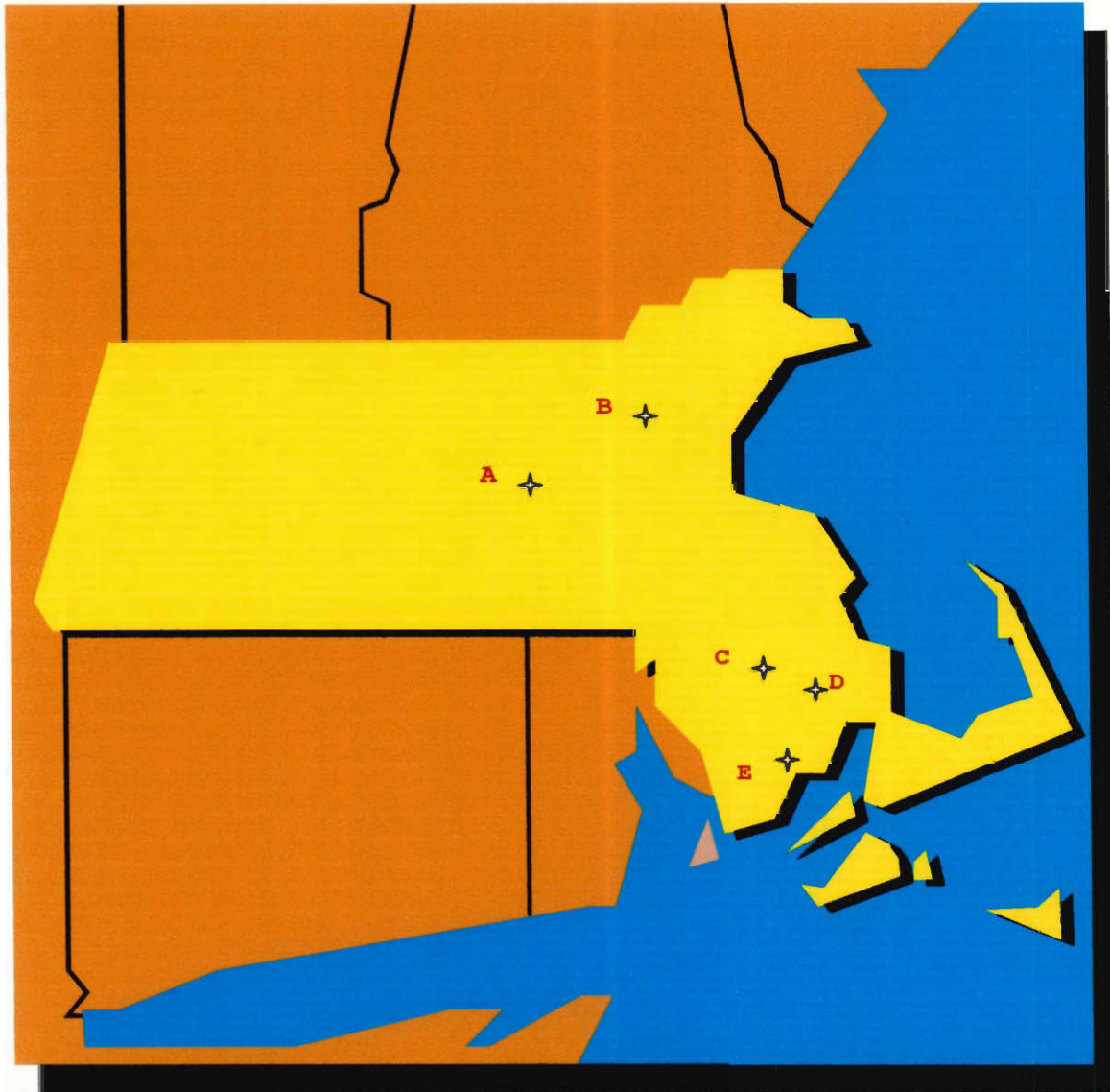73.39%,influence,yes_other,1,0
73.39%,email_account,email_personal,1,1


...


0.49%,nocomp_newspapers,nocomp_videos,3,1
0.49%,nocomp_newspapers,nocomp_videos,5,6
0.49%,nocomp_sports,nocomp_music,1,4
0.49%,nocomp_sports,nocomp_music,1,5
0.49%,nocomp_sports,nocomp_music,6,1
0.49%,nocomp_sports,nocomp_music,6,2
0.49%,nocomp_sports,nocomp_music,6,6
0.49%,nocomp_sports,nocomp_music,7,2
0.49%,nocomp_sports,nocomp_videos,1,6
0.49%,nocomp_sports,nocomp_videos,2,2
0.49%,nocomp_sports,nocomp_videos,2,6
0.49%,nocomp_sports,nocomp_videos,4,1
0.49%,nocomp_sports,nocomp_videos,6,2
0.49%,nocomp_sports,nocomp_videos,7,4
0.49%,nocomp_music,nocomp_videos,1,2
0.49%,nocomp_music,nocomp_videos,2,6
0.49%,nocomp_music,nocomp_videos,5,2
0.49%,nocomp_music,nocomp_videos,6,1
0.49%,nocomp_music,nocomp_videos,6,2
0.49%,nocomp_music,nocomp_videos,6,4
(not a full data listing)
```

Appendix E

Maps

Legend

Key	City	Survey Group
A	Worcester	Test Survey in a WPI CS3043 Class taught by Prof. Michael Ciaraldi
B	Lexington	Lexington Public Schools: 4th, 7th, 10th grades.
C	Bridgewater	Burnell Elementary School: 4th grade.
D	Pembroke Kingston	Silver Lake Regional Junior High Schools: two 7th grade classes Silver Lake Regional High School: two 10th grade classes.
E	Rochester	Olde Rochester Regional High School: two 10th grade classes

Appendix F

Financial Analysis

Travel Expenses

Lexington
```
        Mileage: 111 Miles * $0.30/mile      $ 33.30
        Tolls:                               $  1.20
        Total:                               $ 34.50
```

Bridgewater/Burnell
```
        Mileage: 430 Miles * $0.30/mile      $129.00
        Total:                               $129.00
```

Rochester
```
        Mileage: 220 Miles * $0.30/mile      $ 66.00
        Total:                               $ 66.00
```
Silver Lake
```
        Mileage: 190 Miles * $0.30/mile      $ 57.00
        Total:                               $ 57.00
```

```
Total Travel Expenses:                                   $286.50
```

Printing Expenses

Survey Printing
```
        Copies: 520 impressions @ $0.025     $13.00
        Stapling: 100 sets @ $0.02           $ 2.00
        Total:                               $15.00
```

Brochure Printing
```
        Copies: 40 impressions @ $0.65       $26.00
        Total:                               $26.00
```

Final Report Printing
```
        Copies: 254 impressions @ $0.30      $76.20
        Total:                               $76.20
```

```
Total Printing Expenses:                                 $ 117.20
```

Website Expenses

Domain Name Registration (http://www.project-404.net)
```
        Cost for one year                    $19.00
```

```
Total Website Expenses:                                  $ 19.00
```

Miscellaneous Expenses

Candy for Students
```
        Bags: 9 @ $1.50                      $ 13.50
```

Stationery Supplies
```
        Clips for Surveys: 1 package         $ 3.00
```

```
Total Miscellaneous Expenses:                            $  16.50
```

```
Total Expenses:                                                        $439.20
Total Expected Student Contribution: $100 per term X 3                 $300.00
Total Expected Reimbursement:                                          $139.20
```

Appendix G

Glossary

Broadband – Refers to telecommunication in which a wide range of frequencies is available to transmit information which allows faster transmission than ordinarily possible. Examples of a broadband connection include cable modems and DSL.

Browsing the web – The act of using a web browser such as Internet Explorer or Netscape to view World Wide Web pages.

Cable – Internet access through the cable lines. Typically, a modem is connected to the cable, which is then connected to a computer or network of computers, which then have access to the Internet. Typically much faster than a Dial-up connection.

CGI - (Common Gateway Interface)
A set of rules that describe how a *Web Server* communicates with another piece of software on the same machine, and how the other piece of software (the CGI program) talks to the web server. Any piece of software can be a CGI program if it handles input and output according to the CGI standard.

Computer – **1.** A device that accepts data, processes the data in accordance with a stored program, generates results, and usually consists of input, output, storage, arithmetic, logic, and control units. **2.** A functional unit that can perform substantial computation, including numerous arithmetic operations or logic operations, without human intervention during a run.

Dial-up – A method of connecting to the Internet wherein the computer's model dials up to an access number for a local Internet Service Provider and makes a connection.

Distribution Analysis – An analysis of the various distributions within a raw data set.

Domain Name - The unique name that identifies an Internet site.

DNS - (1) Short for Domain Name System, an Internet service that translates domain names into IP addresses. Because domain names are alphabetic, they're easier to remember. The Internet however, is really based on IP addresses. Every time you use a domain name, therefore, a DNS service must translate the name into the corresponding IP address. For example, the domain name *www.example.com* might translate to *198.105.232.4*.
The DNS system is, in fact, its own network. If one DNS server doesn't know how to translate a particular domain name, it asks another one, and so on, until the correct IP address is returned.

DSL - (Digital Subscriber Line)
A method for moving data over regular phone lines. A DSL circuit is much faster than a regular phone connection, and the wires coming into the subscriber's premises are the same (copper) wires used for regular phone service.

Error 404 – An error message in HTTP wherein a requested object or page was not found. Our project was named after this error because it's a common error that we liked.

File Server – A computer or device on a network that manages network resources and files.

HTML - (HyperText Markup Language)
> The coding language used to create *Hypertext* documents for use on the *World Wide Web*. HTML looks a lot like old-fashioned typesetting code, where you surround a block of text with codes that indicate how it should appear.
> The "hyper" in Hypertext comes from the fact that in HTML you can specify that a block of text, or an image, is linked to another file on the Internet. HTML files are meant to be viewed using a "Web Browser".

Hypertext - Generally, any text that contains links to other documents - words or phrases in the document that can be chosen by a reader and which cause another document to be retrieved and displayed.

internet – (lower case i) Any time you connect 2 or more *networks* together, you have an internet - as in inter-national or inter-state.

The Internet – (upper case i)The vast collection of inter-connected networks that are connected using the *TCP/IP* protocols and that evolved from the *ARPANET* of the late 60's and early 70's.
> The Internet connects tens of thousands of independent networks into a vast global *internet* and is probably the largest *Wide Area Network* in the world.

IP number - (Internet Protocol Number)
> Sometimes called a dotted quad. A unique number consisting of 4 parts separated by dots, e.g. 165.113.245.2
> Every machine that is on the Internet has a unique IP number - if a machine does not have an IP number, it is not really on the Internet. Many machines (especially servers) also have one or more Domain Names that are easier for people to remember.

ISDN - (Integrated Services Digital Network)
> Basically a way to move more dataover existing regular phone lines. ISDN is available to much of the USA and in most markets it is priced very comparably to standard analog phone circuits. It can provide speeds of roughly 128,000 bits-per-second over regular phone lines. In practice, most people will be limited to 56,000or 64,000 bits-per-second.
> Unlike *DSL*, ISDN can be used to connect to many different locations, one at a time, just like a regular telephone call, as long the other location also has ISDN.

LAN - (Local Area Network) A computer network limited to the immediate area, usually the same building or floor of a building.

Laptop – Portable computer that typically weighs less than 10 pounds, and has all the components necessary for use. (screen, processor, memory, mouse, keyboard, etc.)

Left/Right skewed - If the mean is less than the median in a set of data, then the data is right skewed, and if the mean is more than the median in a set of data, then the data is left skewed.

Mac – Short for Macintosh, refers to a brand of personal computers.

Mean – A statistics term referring to the *average* of a set.

Median – A statistics term referring to the halfway point of a set. For example if a set contains {1, 4, 7, 9, 10, 15, 17, 123, 132} the median would be 10, which is the middle number in the set.

Mode – A statistics term referring to the most common element in a set. For example if a set contains {1, 2, 6, 6, 6, 9, 10, 15, 15, 20} the Mode is 6 because it occurs 3 times which was more than any other element.

NCP - Network Control Protocol - A set of rules for formatting, ordering, and error checking data used by the ARPANET and other early forerunners of the Internet.

Network - Any time you connect 2 or more computers together so that they can share resources, you have a computer network. Connect 2 or more networks together and you have an *internet*.

Online – The state of sitting at a computer while said computer is connected to the Internet.

Outlier – A piece of raw data that is inconsistent with all other data gathered, and is not indicative of the group of data as a whole.

PAN – Personal Area Network - the interconnection of information technology devices within the range of an individual person, typically within a range of 10 meters. For example, a person traveling with a laptop, a personal digital assistant (PDA), and a portable printer could interconnect them without having to plug anything in, using some form of wireless technology. Typically, this kind of personal area network could also be interconnected without wires to the Internet or other networks.

PC - Personal Computer, Non-Macintosh, IBM compatibles.

Perl – Practical Extraction and Report Language – A scripting language that excels in text processing.

Posix – Portable Operating System Interface for Computing Environments. - A Federal Information Processing Standard Publication (FIPS PUB 151-1) for a vendor-independent interface between an operating system and an application program, including operating system interfaces and source code functions

REIEFS – Results of Electronic Information Exchange in a Free Society – Electronic information exchange refers to computers on the Internet, and Free Society refers to any society that aspires to give it's citizens basic freedoms.

Scripting Language – A programming language supported by and specific to a particular program. *Note:* A scripting program is normally used to automate complex or advanced features or procedures within the program.

T1 - A *leased-line* connection capable of carrying data at 1,544,000 *bits*-per-second. At maximum theoretical capacity, a T-1 line could move a *megabyte* in less than 10 seconds. That is still not fast enough for full-screen, full-motion video, for which you need at least 10,000,000 bits-per-second. T-1 lines are commonly used to connect large *LAN*s to the *Internet*.

TCP/IP - (Transmission Control Protocol/Internet Protocol)
This is the suite of protocols that defines the *Internet*. Originally designed for the *UNIX* operating system, TCP/IP software is now included with every major kind of computer operating system. To be truly on the *Internet*, your computer must have TCP/IP software.

UDP – (User Datagram Protocol)
One of the protocols for data transfer that is part of the *TCP/IP* suite of protocols. UDP is a "stateless" protocol in that UDP makes no provision for acknowledgement of packets received.

Unix - A computer operating system (the basic software running on a computer, underneath things like word processors and spreadsheets). Unix is designed to be used by many people at the same time (it is multi-user) and has *TCP/IP* built-in. It is the most common operating system for *servers* on the *Internet*.

W3C - an organization to develop and promote open standards on the Internet. Comprised of an international consortium of companies, it is the governing body behind HTTP and HTML/XML standards.

WAN - (Wide Area Network)
Any *internet* or *network* that covers an area larger than a single building or campus.

WWW - (World Wide Web)
World Wide Web (or simply Web for short) is a term frequently used (incorrectly) when referring to "The Internet", WWW has two major meanings:
First, loosely used: the whole constellation of resources that can be accessed using *Gopher, FTP, HTTP,telnet, USENET, WAIS* and some other tools.

Second, the universe of hypertext *servers (HTTP servers)*, more commonly called "web servers", which are the servers that serve *web pages* to web *browsers*.

Xanadu – A project that Ted Nelson started in 1960 which envisions a more complex World Wide Web system than hypertext with bi-directional links, and automatic version and rights management. More information is available at www.xanadu.net

16

Acknowledgements