2015

Digital Forensics Tool Interface Visualization

Roberto A. Altiero
Nova Southeastern University, baltiero@yahoo.com

Digital Forensics Tool Interface Visualization

by

Robert A. Altiero

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
Computer Information Systems

Graduate School of Computer and Information Sciences

Nova Southeastern University

2015

We hereby certify that this dissertation, submitted by Robert Altiero, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

Maxine S. Cohen, Ph.D. — Date
Chairperson of Dissertation Committee

Eric S. Ackerman, Ph.D. — Date
Dissertation Committee Member

Gary C. Kessler, Ph.D. — Date
Dissertation Committee Member

Approved:

Eric S. Ackerman, Ph.D. — Date
Dean, Graduate School of Computer and Information Sciences

Graduate School of Computer and Information Sciences
Nova Southeastern University

2015

An Abstract of a Dissertation Submitted to Nova Southeastern University
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Digital Forensics Tool Interface Visualization

by
Robert A. Altiero
January 2015

Recent trends show digital devices utilized with increasing frequency in most crimes committed. Investigating crime involving these devices is labor-intensive for the practitioner applying digital forensics tools that present possible evidence with results displayed in tabular lists for manual review. This research investigates how enhanced digital forensics tool interface visualization techniques can be shown to improve the investigator's cognitive capacities to discover criminal evidence more efficiently. This paper presents visualization graphs and contrasts their properties with the outputs of The Sleuth Kit (TSK) digital forensic program. Exhibited is the textual-based interface proving the effectiveness of enhanced data presentation. Further demonstrated is the potential of the computer interface to present to the digital forensic practitioner an abstract, graphic view of an entire dataset of computer files. Enhanced interface design of digital forensic tools means more rapidly linking suspicious evidence to a perpetrator.

Introduced in this study is a mixed methodology of ethnography and cognitive load measures. Ethnographically defined tasks developed from the interviews of digital forensics subject matter experts (SME) shape the context for cognitive measures. Cognitive load testing of digital forensics first-responders utilizing both a textual-based and visualized-based application established a quantitative mean of the mental workload during operation of the applications under test. A *t*-test correlating the dependent samples' mean tested for the null hypothesis of less than a significant value between the applications' comparative workloads of the operators. Results of the study indicate a significant value, affirming the hypothesis that a visualized application would reduce the cognitive workload of the first-responder analyst. With the supported hypothesis, this work contributes to the body of knowledge by validating a method of measurement and by providing empirical evidence that the use of the visualized digital forensics interface will provide a more efficient performance by the analyst, saving labor costs and compressing time required for the discovery phase of a digital investigation.

Acknowledgements

I am forever grateful to my lovely wife Joann for quietly enduring this eight-year journey of sacrificed evenings, weekends, and vacations while I researched.

Additionally, I would like to thank my dissertation chair, Dr. Maxine Cohen, for her patience, direction, and even keel throughout my course work and dissertation process. Also, I would like to thank my dissertation committee members, Dean Eric Ackerman and Dr. Gary Kessler, for their always timely and thoughtful review of this work.

Moreover, I would like to thank business owners Mr. Brent Snyder and Mr. Victor Holt for time, tuition, and encouragement. Further thank yous to my supervisors, Ms. Gina Nairn and Ms. Julia Brandt, for encouragement to complete this task.

Table of Contents

List of Tables

Tables

List of Figures

Figures

Chapter 1

Introduction

Background

Osborne, Turnbull, and Slay (2010) chronicle digital forensics tools as vital for providing analysts the utility for detecting and discovering digital evidence of a crime, identifying two such industry-standard tools as the EnCase Forensics (Guidance Software) and the AccessData Forensic Toolkit (FTK). However, these forensic tools require an inordinate amount of human intervention for evidence value to be determined. The most recently released upgrade of the FTK introduces a visualization module—although it is accompanied by no supporting empirical visualization documentation. As these existing forensic tools provide little guidance in the discovery of evidence, their effectiveness depends upon the experience of the practitioner (Jankun-Kelly, Franck, Wilson, Carver, Dampier, & Swan, 2008). Additionally, ever-increasing numbers of cases of computer-generated crime render inversely proportionate the resource of professionals available to detect and solve the infractions and those cases to be solved (Neufeld, 2010). Compounding the problem of this disadvantageous position of professionals aiming to detect evidence and solve cases are methodologies that prove inadequate as they do not scale to the increasing volumes of crimes and digital evidence. Efficacy of forensic software tools such as those mentioned above potentially depends

more upon the tool interface and less upon cognitive decisions of the analyst whose access to information visualization will enhance the process of evidence discovery. Integrating visualization into the interface design of forensic tools condenses the analysis phase of investigations by utilizing the practitioner's visual sensing abilities, thereby improving the incident rate of detection and discovery of valuable evidence (Ayers, 2009; Osborne et al.).

Problem Statement

The problem to be resolved is that forensic tools currently in use require an inordinate amount of human effort for value of evidence to be determined. At the root of this problem is the fact that computer forensic technologies' textual-based interfaces, as described by Osborne et al. (2010), overburden the investigator's cognitive capacity to gather evidence rapidly and efficiently from the information systems suspected of having been used with criminal intent. Moreover, Osborne et al. and Hargreaves and Patterson (2012) surmise that the circumstance of the continually increasing storage capacity of computer systems compounds with that of additional, varied types of digital devices to thwart forensic analysts' attempts to discover forensics evidence with their obsolete manual analysis processes. Whereas, previously, a forensics analyst had to comb through mere megabytes of information to locate evidence of inappropriate acts and behaviors, one may now have to look through terabytes of data to retrieve evidence. Visually enhanced tools could provide the opportunity to amplify thousands or millions of files—currently consigned to painstaking analysis by textual-based tools—in the process of locating relevant evidence, according to Osborne et al. Yet the arduousness and

limitations of this process diminish the effectiveness of law enforcement, government, and other organizations that need to function in the domain of computer forensics. This study has provided subjects a prototype visualized interface forensics tool and contrasted their results with those of their use of textual-based interface utilizing The Sleuth Kit (TSK) program, an open-source digital forensics tool with a textual-based interface. TSK has been utilized to measure the effectiveness of enhanced data presentation. Resolving the problems introduced by textual-based interfaces was expected to elevate the investigator's capabilities so as to expedite the discovery of digital evidence (Ayers, 2009; Osborne & Turnbull, 2009; Osborne et al.).

Dissertation Goal

The questions presented below were intended to guide the research and to suggest appropriate research methods, such as ethnographic observations, for discovery of a novel solution to the problem of the analyst's overburdened cognitive capacity in relation to digital forensics tools currently in use. Enhanced user interface of forensics tools' analysis presentation through visualization techniques achieves a dual goal: it not only presents a solution to optimize the digital investigator's cognitive capacity but also develops best practices in contemporary approaches to the application of human–computer interaction (HCI) visualization tactics. This study was developed partly to verify Osborne and Turnbull's (2009) suggestion that HCI techniques applied to develop practices based on visualization have proven capable of enhancing the user interfaces of intrusion detection systems (IDS), antivirus tools, and other anti-malicious software solutions. Visualization facilitates the discovery of evidence from large volumes of data

during the detection process by minimizing the element of human interaction (Osborne & Turnbull).

Ayers (2009) identified the requirements for the next generation of forensics analysis systems as having 1) increased investigation speed, 2) better accuracy, 3) established workflow, and 4) an advanced abstraction layer that improves human capabilities. The current generation of forensics tools presents the file system hierarchy whereby information visualization displays an abstract view of all data under investigation so that a user may obtain knowledge or so that he or she may discover digital evidence. Osborne et al. (2010) explained visualization as an enabler for the investigator to understand large amounts of data. However, they pointed out that few published works detail visualization techniques intended for the frameworks of investigative tools. Presented in this research are a detailed review of relevant contemporary literature and reports of product testing and software prototyping of a visualization solution.

Research Questions

Answers to research questions develop concepts of other researchers, such as Ayers (2009); Osborne and Turnbull (2009); and Osborne et al. (2010). The following questions were meant to guide this research effort and to lead, ultimately, to answers that will begin to resolve the problem of an exorbitant amount of human intervention needed currently in the discovery of digital forensics evidence determination.

1. What are the investigator's primary tasks for evidence identification while operating a traditional digital forensics tool set? Answer is determined by techniques employed in the analysis phase of an investigation for cataloging purposes and by observed

methods of a digital forensics user group for the purpose of ethnographic discovery (Fetterman, 2010).

2. What is the cognitive load of an assessment of human working memory beyond just time and accuracy measures while performing ethnographically discovered techniques of a predefined set of tasks to establish a baseline of evidence identification? Answer is determined by application of predefined, relevant standard tasks identified as ethnographic and implemented for the measurement of a prototype visualized application intended to improve the effectiveness of a digital forensics investigation (Çakir, 1997; Fetterman; Huang, Eades, & Hong, 2009; Saraiya, North, & Duca, 2010).
3. What efficiency level may be attained for a digital investigation improved for the benefit of analysis by application of visualization to predefined tasks? Answer is determined by comparing/contrasting the results of the users' surveys (Çakir, 1997; Fetterman; Huang, Eades, & Hong; Saraiya, North, & Duca).

The working hypothesis of the research is that—by application of advanced graphical or visualization technology to digital forensics tools—the workload of digital forensics analysts will be reduced, enabling them to discover digital evidence more expeditiously during the analysis phase of an investigation than is currently possible. Employing visualization techniques allows analysts not only to view an entire data set under investigation but also to zoom and filter items of interest and then to gain—via the interface—access to specific details on demand of the data under examination (Shneiderman, 2008; Shneiderman, 1996).

Relevance and Significance

The Regional Computer Forensics Laboratories (RCFL) program of the Federal Bureau of Investigation (FBI) provides critical digital forensics expertise, services, and training to thousands of law enforcement officers and hundreds of agencies nationwide. Membership in the RCFL, which is composed of 130 participating agencies from 17 states, requires an investigator to earn FBI certification as a computer forensics examiner. As documented in the RCFL 2012 fiscal year (FY) report, released in May 2013, nearly all criminal investigations involve digital evidence availing prosecutors a window into events that occurred before, after, and sometimes during the execution of criminal acts. Committed by tech-savvy perpetrators, these crimes involve financial schemes, terrorism, child pornography, and gang-related activities—among other types. While digital storage capacity technologies increase annually for consumers, the cost of enhancement of computer electronics decreases; such affordability permits computer use by the masses, including the criminal element. Usually, apprehended criminals are associated with a computer with volumes of information that must be labor-intensively examined by an investigator. Effectiveness of current digital forensic techniques' search approaches is regarded by experts as "poor" (Beebe, Clark, Dietrich, Ko, & Ko, 2011; Al-Zaidy, Fung, & Youssef, 2011; RCFL, 2013).

Within the RCFL, digital crime is known to be far-ranging at home as well as overseas, involving terrorists intent on killing thousands of innocent civilians and destroying valuable infrastructure. According to Richard and Roussev (2006), improved digital forensics tools advance investigating authorities' ability to safeguard property and

even life, especially in time-sensitive situations wherein the examination of digital evidence is critical to solving a crime in progress.

Despite commercial digital forensics tool vendors' introductions of equipment operations and database archives, attempts at improving digital forensics technology have left the architecture of the present generation of tools relatively unchanged from that of earlier versions. While advances may have proven effective wherever sufficient human resources have been made available, neither efficiency nor reliability has improved for the single investigator (Ayers, 2009).

At risk is the populace—through investigators' inabilities to safeguard them from perpetrators utilizing computer systems as a medium for their crimes. Contemporary researchers such as Ayers (2009), Osborne and Turnbull (2009), and Osborne et al. (2010) called for the development of improved methods of visualizing digital forensics data. This work contributes to the knowledge base of digital forensics by developing visualization techniques with measurable outcomes synthesized from methodologies used in the domains of HCI and information security. Findings apply to the field of digital forensics—with resulting presentation techniques and measures not yet determined or known to the domain. The project's results and measures are general enough, however, to apply to other domains within which the presentation of large volumes of data is needed for system users to function with measurable outcomes.

The goal here has been to discover means by which to improve digital forensics interfaces, to identify how implementing visualization enhances interfaces, and to focus on contemporary approaches for the application of HCI visualization tactics. The paper identifies a problem with computer forensics tools currently in use, for which it presents a

potential solution. The problem may be defined as forensics tools' stopping short in the computer forensics evidence-gathering process and leaving evidence validation to human intervention. Endicott-Popovsky and Frincke (2007) stated that the first discoverers of evidence are often network administrators who have applied their cognitive skills to a digital crime scene by combing through textual evidence. The digital forensics crime scene is most often a functional computer network supporting a business operation. Endicott-Popovsky and Frincke refered to these network administrators as "first responders," for it is they who must decide whether to respond to network users or to pursue investigations. Such conditions sometimes overlook prosecutable crimes (Endicott-Popovsky & Frincke).

This work expands the digital forensics science knowledge base by establishing digital forensics processes gained from an ethnographic study and empirical evidence from cognitive measures, offering recommendations for the domain of digital forensics, and demonstrating enhancement of the investigator's final analysis of digital evidence by exploitation of visualization characteristics in support of interface capabilities, having retrieved values and filtered data graphically, as suggested by Ahn, Plaisant, and Shneiderman (2011). Tufte (1990) finely analyzed large amounts of multi-dimensional information and illustrated it graphically in the two-dimensional print medium. His printed images yielded complex timetables of dense patterns of information, exhibiting four or five variants of information for audiences. Richard and Roussev (2006) identified primary operations of a digital forensics investigation depending mostly upon the capture of file-centric evidence. This study presents visually the multi-dimensional attributes of a

file system to aid the investigator in a manner that allows him or her to analyze and identify digital forensics evidence expeditiously.

Researchers such as Ayers (2009) called for advancements in the abstraction of relevant data to comprise an approach outside the viewing of the hierarchical file system currently used in forensics tools. Findings of this research—when utilized as enhancements to digital forensics applications—explain how the investigator's capability to identify evidence may be improved through visualization. Moreover, this work serves as a resource for other domains, such as data mining and information security that may benefit from visualization-enhanced interfaces.

Barriers and Issues

Despite the introduction of paralleled equipment operations and database archives, attempts at improving digital forensics technology have left the architecture of the present generation of tools relatively unchanged (Ayers, 2009). Advances have proven effective only in instances where sufficient human resources have been available, but they have not improved efficiencies or reliability of factors such as audit capabilities.

Technical barriers include the sheer volume of data to be analyzed, a condition attributed to the increase in both storage technology advances and digital device ownership. Difficulty in research is exacerbated by variety in devices; for example, the Microsoft operating systems feature many different files, bins, and categories for data storage, such as the recycle bin and event logs. Accommodating the many categories and locations of varying file system metadata compounds the forensics analyst's workload. Digital forensics investigations, per se, are not particularly technically challenging;

rather, such projects prove to be time-intensive. Hence, automation of the workload on digital forensics professionals proves critical for reducing the time required for digital evidence discovery (Hargreaves & Patterson, 2012).

Increasing the workload of analysts investigating the digital crime scene is the variety of types of digital devices owned by individuals, such as data-capable mobile devices, now found at practically every crime scene and manufactured at an ever-increasing rate. Additionally, these mobile devices, being network devices, introduce concomitant challenges, including privacy issues and technical issues as the devices move into and out of their networks (Mislan, Casey, & Kessler, 2010).

Several papers are dedicated to visualization of digital forensics temporal evidence, such as those of Hargreaves and Patterson (2012) and Olsson and Boldt (2009). However, timelines are just one file attribute evaluated during an investigation. According to the National Institute of Justice Special Report (2004), digital forensics analysis also requires examination of file data: file name, size, and path as well as correlation of relationships among files.

Lastly, the paucity of courses and programs related to digital forensics education creates a shortage of trained investigators (Kessler & Ramsay, 2014). Varying communities of interest—such as business and education as well as law enforcement, judicial departments, policy makers, and other government agencies—are adversely impacted, for each depends upon the identification and delivery of digital evidence. Urgently needed is digital forensics education supporting students of both career development and degree-granting programs (Cooper, Finley, & Kaskenpalo, 2010).

Examination of large amounts of data burdens the practitioner. Glatz (2010) suggested that the cognitive load on the analyst can be reduced through visualization, which provides a method for making accumulated data easier to read. Hargreaves and Patterson (2012) further suggested that applied visualization will reduce the volumes of data to be analyzed through timeline analysis, thereby improving file system metadata examination.

Assumptions, Limitations, and Delimitations

Existing digital forensics research findings clearly imply that using visualization in interface design will reduce the cognitive workload of the digital forensics analyst, as suggested by Osborne, Turnbull, and Slay (2012). An example presented by Jones (2008) illustrated the principle that an investigator must examine each file of a computer system. He suggested that resources are conserved in performing these digital forensics investigations through visualization. Moreover, the digital forensics discipline lacks clear, empirically supported research data proving efficiencies gained through visualized digital forensics tool sets. This study assumes that data yielded by its examination of study participants in the forensics field prove such efficiencies gained.

Beyond the control of this study were resource constraints, such as the schedule availability of digital forensics investigators to participate in the study, a circumstance realized by Barbara (2013) as well. Likewise limiting was unavailability commercially available tools such as FTK do not offer trial versions for testing or for research purposes. Tools used in this study are those freely available as open sources for both digital forensics and visualization demonstrations. Tufte (1990) mentioned numerous options for presenting multi-dimensional data attributes and density of data to illustrate

information techniques. Covering all visualization techniques and technologies possible for the presentation of digital forensics information would have been impossible. That digital evidence presented graphically requires less cognitive effort to be understood by the analyst—as contrasted with evidence derived from textual displays—is the hypothesis of this study, a concept advanced by Osborne, Turnbull, and Slay (2012). Simulated forensics data functioned to preserve the right of confidentiality of both an alleged perpetrator and a victim and to preserve the custody chain of evidence. The hypothesis is demonstrated by one tool's textual representation of digital evidence and by another tool's graphic representation of digital evidence. Limiting the number of tools to two narrows the scope of the study for purposes of manageability and precludes generalizations in results of the study.

Conti (2007) introduced a visualized file system's multi-dimensional attributes acquired by application of the freely available SequoiaView, which features a treemap format, providing big-picture context enabled with interactive application controls for drilling down to file system items of interest. SequoiaView is the graphic interface tool for this study demonstrating visualization. Demonstrating a fully functional digital forensics tool with textual display, this study utilized The Sleuth Kit (TSK) with the Autopsy user interface, an open-source tool usable for in-depth analysis of multiple file system images (Sleuth Kit, 2012).

Definition of Terms

Analyst — Collects, understands, and determines collected digital events as legally admissible evidence (Peisert, Bishop, Karin, & Marzullo, 2007)

Autopsy — The graphical interface to The Sleuth Kit digital forensics tool (Sleuth Kit, 2012).

Cognitive Capacity — The extent of an individual's allocation of cognitive resources used for analytic processing, the primary component being one's working memory (Stanovich & West, 2000)

Cognitive Load Measures — The assessment of cognitive load by measuring mental load, mental effort, and performance (Paas, 1992)

Cognitive Load Theory (CLT) — Developed for the improvement of instructional methods utilizing the learner's limited cognitive processing capacity in acquiring knowledge based on one's limited working memory for processing visual/spatial and auditory/verbal information (Paas, Tuovinen, Tabbers, & van Gerven, 2003)

Digital Evidence — Data preserved that have been identified through discovery by their attributes or recovered deleted files or other information captured from digital media and used to ascertain the truth in the proof or disproof of a crime (Osborne et al., 2010)

Digital Forensics or Computer Forensics — Analysis of an electronic device's current state of stored information in order to solve crimes (Osborne et al., 2010; Peisert, Bishop, & Marzullo, 2008)

Digital Forensic Tools — Software/hardware used by investigators for viewing files or directories as well as unallocated files of a suspect computer system (Carrier, 2003)

Electronic Fingerprint — Result of a hashing algorithm utilized to authenticate that the digital evidence has not been tampered with or altered since being captured during a digital forensics investigation (Kruse & Heiser, 2002).

Ethnography — Originating in anthropology, fieldwork that studies cultural and societal norms from inside their operations (Ormerod et al., 2005)

First-Responder — Practitioner, often a network administrator, who collects digital crime scene data (Endicott-Popovsky & Frincke, 2007)

Graph Visualization — Data presented graphically with nodes representing information intersects and intersect attributes, such as connectors that may be represented by node color and shape (Hansen, Shneiderman, & Smith, 2010; Huang et al., 2005)

Hash — The encryption result of a mathematical algorithm procedure conducted on a device or file utilized as a digital fingerprint, which provides authenticity of evidence gathered during a digital forensic investigation (Kruse & Heiser, 2002)

Human–Computer Interaction (HCI) — A multidisciplinary science focused on social and behavioral sciences, including computer and information technology, concerned with

how devices and systems can be more useful and more readily usable by people (Carroll, 2003)

Information Visualization — "The use of computer-supported, interactive, visual representations of abstract data to amplify cognition" (Carroll, 2003, p. 468)

Parallel Coordinate Visualization — A method of visualization presenting multidimensional data items and displaying data along a polygonal line intersecting the horizontal dimension axes at the position corresponding to the value for the corresponding dimension (Keim, 2002)

SequoiaView — An open-source visualization computer application employing the cushion treemap technique to present the entire content of a hard drive or file system in a single view (SequoiaView, 2014)

SleuthKit — An open-source digital forensics tool utilized to investigate computer disk images for in-depth analysis of a file system (Sleuth Kit, 2012)

Subject Matter Expert (SME) — Specialized practitioners in their domain of expertise; in this study, law enforcement's digital forensics practitioners (Peterson, Raines, & Baldwin, 2007)

Summary

Digital forensics tools are software and other devices that provide analysts an instrument to assist in the discovery of digital evidence located in an array of computer systems. The RCFL 2012 FY report revealed that nearly all current criminal investigations involve such a device. This research builds upon previous works to demonstrate enhancement of the discovery phase of the evidence-identification process through the use of visualization. It illustrates the impact of visualization integrated into the user interface of the digital forensics tool and upon the digital investigation itself. However, as digital forensics remains in its infancy, further development is needed to aid the analyst who may need to assess thousands, millions, or even billions of files to identify digital evidence, an inordinately cognitively challenging task.

Such development is promoted by proof of the impact of visualization on the digital forensics tool to analyze and discover evidence read digitally rather than merely cognitively. Additionally heightened is the understanding of the process whereby visualization reduces analysts' cognitive workload, thereby proving digital evidence detection more efficient.

Chapter 2

Review of the Literature

Theory and Research

Three primary domains of study are included in the research topic: digital forensics, visualization, and HCI. Research methods cannot possibly cover every topic from each of the domains. Regarding the domain of digital forensics alone, Ray and Bradford (2007) provided four models of digital forensics activities pertaining solely to case development. The primary focus for this research is validation of types of improvements to digital forensics tools.

Garfinkel (2010) identified research challenges of digital forensics tools as evidence-oriented design, visibility, filtering and report modeling, the difficulty of reverse engineering, and monolithic applications. Further, he advocated digital forensics tools' displaying all evidence data in a tabular form. Concentration upon evidence visibility presented by the digital forensics tool suggests that the use of a graphical interface for presentation of evidence will reduce the cognitive load of the investigator, for example, by testing a data store simulating digital evidence.

The ubiquity of suspected digital devices used in crime is explicable by their affordability and the fact that the increasingly vast storage capacity of binary data has outpaced the capabilities of digital forensics toolsets and the analysts who operate them.

These key issues, as suggested by Osborne, Turnbull, and Slay (2012), challenge investigations of large quantities of data yet are countered by application of information visualization techniques, which can highlight patterns in digital evidence by both technical analysts and nontechnical investigators. Graphical views condense and present millions of data points of interest in a single display. Such abstract presentations reduce the cognitive demand upon analysts to assimilate data sets vastly larger than those of textual displays (Osborne, Turnbull, & Slay, 2010, 2012).

Visualization's objective is to aid users in the examination, organization, and discernment of large amounts of data (Card & Mackinlay, 1997). Through interactive presentations, visualization boosts the cognitive capability of the analyst to gain knowledge in data being studied. Interactive visualization reveals the existence of relationships within a digital collection of information (Card, Mackinlay, & Shneiderman, 1999). Interaction between increasingly vast abstract datasets and their inherent attributes requires awareness and observations absorbed mainly through the visual sense; information is said to be attained more commonly through the visual sense than through all other senses combined (Card & Mackinlay; Card et al.; Osborne et al., 2012). Card et al. asserted that visualization improves the user's cognition by increasing his or her memory and processing speed, decreasing the time required to search information, using patterns to enhance the detection of information, enabling inference operations by use of perceptual mechanisms for monitoring, and encoding information in an adjustable medium. Information visualization utilizes the capabilities of the human's visual sense, thereby enhancing an analyst's awareness and understanding of abstract

information found in the immense sets of data prevalent in today's computing environments (Heer, Card, & Landay, 2005).

Comer, Gries, Mulder, Tucker, Turner, and Young (1989) referred to human-computer communication as one of the primary disciplines in the academic field of computer science. Today this field of study, referred to as HCI, is dedicated to design, evaluation, and implementation of interactive systems through education and planning. HCI is multi-disciplined, including areas such as psychology, ergonomics, and cognitive sciences. The primary role of HCI in research is to improve the human's experience when interfacing with computing devices. State-of-the-art interfaces improve the user's cognitive abilities by implementation of appealing visual presentations, enabling capabilities through the human's visual perception. The amount of multidimensional data in the modern information system challenges usability and, thus, drives current and future interface design (Ebert, Gershon, & van der Veer, 2012).

The method of study employed in this research draws upon ethnography and cognitive load theory. Ethnography is a strategy providing the researcher insight into the natural environment of the study participants (Creswell, 2009). The first phase of this study was conducted in order to document how textual-based tools are used in the process of evidence collection, and the second phase to demonstrate the application of cognitive load theory to produce empirical information about how a participant's cognitive load has been reduced when he or she has been provided with a visualized display of simulated digital evidence, as recommended by Huang, Eases, and Hong (2009).

According to the literature, digital forensics tools have been outdone by variations of computing devices such as cell phones and these devices' ever-increasing storage

capabilities. Enhanced capabilities of these devices to store digital evidence—combined with faster networking and Internet speeds—result in increasing numbers of digital appliance owners with affordable, high-capacity devices (Beebe, Clark, Dietrich, Ko, & Ko, 2011; Osborne, Turnbull, & Slay, 2010). The aforementioned challenges prompt researchers to suggest information visualization methods as a means of resolution, for such techniques include interface designs that are adjustable as well as interactive, visual, quantitative representations of data featuring shapes, colors, and animation (Osborne, Turnbull, & Slay, 2012).

Information visualization aids the user in exploring, managing, and understanding the increasing quantities of digital information (Toker, Conati, Steichen, & Carenini, 2013). Shneiderman (1996) pointed to research successes in interface visualization design methods for structured database and textual presentations, attributing such successes to designers' having created a visual language in multiple domains wherein users complete visual technology tasks of filtering for information-gathering. Using visualization reduces information overload and anxiety in the user experience in data mining and data warehousing, for example, in digital forensics by the capabilities of both the experienced and non-savvy analyst (Osborne, Turnbull, & Slay, 2012; Shneiderman).

Experts within the information security field of study have concluded that challenges that they face are also faced by experts in the field of digital forensics (Osborne, Turnbull, & Slay, 2010). Information visualization presents graphically to the digital forensics analyst a compressed view of data and sources by the millions, reducing the amount of cognitive effort required by analysts of textual displays (Osborne, Turnbull, & Slay, 2012). Toker, Conati, Steichen, and Carenini (2013) provided details on how

information visualization improves user cognitive abilities in processing large amounts of information. Conti (2007) detailed how visualization relates to information security and explains that evaluation of this type of visualization is lacking. Similarly, he noted the lack of works detailing the impact of visualization on the digital forensics tool interface user community.

The literature fails to account for the impact of visualization upon digital forensics tool sets. Texts and other peer-reviewed works have been published on digital forensics, some addressing visualization, but none presents empirical evidence explaining how to reduce the cognitive workload of the analyst by use of contemporary visualization techniques.

Saltzer and Schroeder (1975) described the mechanics of providing security for computer-based information. Their architecture defined and detailed eight general practices for safeguarding information from security incidents: economy of mechanism; fail-safe defaults; complete mediation; open design; separation of privilege; separation of least privilege; separation of least common mechanism; and psychological acceptability. Ultimately, addressed here is psychological acceptability, an outline of the human interface to the system.

The past decade has yielded numerous frameworks and proposed models for improvements to digital forensics, such as those by Bhat, Rao, Abhilash, Shenoy, Venugopal, and Patnaik (2010) and Digital Forensic Research Workshop (DFRWS) (2001). Some suggested improved graphics to enhance the analyst's capabilities (Ayers, 2009; Osborne & Turnbull, 2009; Richard & Roussev, 2006). Hargreaves and Patterson (2012) and Olsson and Boldt (2009) documented visual representations of digital

evidence timelines. However, few sources explored implementation of visualization for forensics application.

Digital Forensics

Discussion of visualization methods necessitates consideration of what is known about computer or digital forensics. Mohay, Anderson, Collie, McKemmish, and de Vel (2003) defined the process of digital forensics as the science of identifying, evaluating, and presenting digital evidence. There exists well-defined means and processes by which to accomplish the goals of this branch of science—as there are with any other. Just as in other fields, one of the required and less glamorous steps in computer forensics is tediously documenting everything from whatever may have led to an investigation to how it has been conducted and what has been discovered. Selamat, Sahib, Hafeizah, Yusof, and Abdollah, (2013) presented the framework for an investigation derived from the DFRWS (2001) digital forensics process categories as follows: 1) identification, 2) preservation, 3) collection, 4) examination, 5) analysis, and 6) presentation. They defined the framework as varying by organizational policy and digital medium/device type.

The terms *digital forensics* and *computer forensics*, applied interchangeably (Peisert, Bishop, & Marzullo, 2008), are defined as a branch of forensic science wherein investigative results are used to prove or refute accusations of crime committed by means of digital devices. The prosecution in most criminal cases depends heavily upon physical evidence; in cases of digital crimes, physical evidence may be nonexistent though

potentially incriminating evidence is stored in digital logs or binary form (Bhat et al., 2010; Peisert et al., 2008; Trček, Abie, Skomedal, & Starc, 2010).

Advances in communications and computer equipment have greatly impacted both personal and professional contemporary lifestyles. Commonplace services as well as public infrastructure depend upon these systems. As these conveniences advance, so does a range of malicious activities, including both natural disasters and human misuses of computer systems. Many daily activities have moved into cyberspace—along with those of the criminal element of society. Trček et al. (2010) described how obtaining evidence of cyberspace crimes is particularly challenging to investigators as there has been a shift in criminal investigations from reliance upon witnesses; confessions; and, most demanding, no physical evidence—to reliance upon digital evidence. Without physical evidence, digital crimes exist only as binary information found in a wide range of digital media such as magnetic storage devices, semiconductors, and optical media (Trček et al., 2010).

With much critical evidence being found in computer logs, digital evidence is replacing physical evidence. Peisert et al. (2008) presented FBI findings claiming that in 2002 half of criminal cases investigated involved a computer, and the RCFL 2013 report revealed that in 2012 all of its participating agents' investigations involved a computer. In the recent past, digital evidence was seized in a single device whereas today evidence is often located on critical live networks that cannot be easily secured, for example, air traffic control systems (Trček et al., 2010).

Digital Forensics Process

Digital forensics' existing methods and principal strategies aim to secure admissible evidence, to preserve the evidence, to identify and extract pertinent data, and to present documentation of interpreted computer data (Bhat et al., 2010; Trček et al., 2010; Wang, 2007). This fundamental procedure as a framework for digital forensics investigation is shown in the DFRWS technical report to have been scientifically derived. The procedure is one of the first sets of guidelines outlining how to derive evidence from digital sources. However, digital forensics methods are developing only gradually; hence, they are not as mature as physical evidence frameworks, for example, those found in DNA-based forensics (Peisert et al., 2008; Trček et al., 2010).

The first step to every digital forensics investigation is acquisition of the digital evidence, which necessitates copying the original data and storing the digital evidence by following a prescribed chain of custody consisting of positive controls that formalize procedures in sealing, archiving, and documenting the process. Handling the evidence, once it has been gathered, poses other challenges, including analysts' conforming to protocol regarding collection, identification, chain of custody, transportation, and storage of evidence (Kruse & Heiser, 2002; Trček et al., 2010).

Maintaining integrity of evidence necessitates protection of the evidence from environmental factors and keeping the evidence consistent with its source of origin. Kruse and Heiser (2002) stated that keeping an "electronic fingerprint" of an entire drive or file can be achieved by using a cryptographic technique called a *hash*. The hash value is produced during the initial collection process: a time stamp and key validate the evidence's authenticity. The same technique and algorithm used to produce the hash are

used to test the evidence in the future to ensure that it has retained its original form. Challenges such as these (and others) must be taken into account when evidence is gathered (Kruse & Heiser, 2002; Trček et al., 2010).

When acquiring evidence, an investigator must be flexible in adapting to presenting anomalies. Evidence that may reside in computer memory could become corrupt if the computer under investigation were to be shut down. Live or volatile evidence, such as information residing in memory, presents a significantly greater number of challenges to the forensics analyst than does evidence found in non-volatile disk or flash storage devices. Volatile evidence cannot be cloned as can traditional, static evidence. Additionally, even with non-volatile evidence-gathering operations, restrictions may be imposed in instances of a system's being critical for business operations so that a system manager or a business owner may be reluctant to shut down the system for the purpose of gathering evidence networks. Mobile devices provide additional challenges due to their networking properties that potentially change internal evidence as they move into and out of their coverage zones (Ieong & Leung, 2007; Mislan, Casey, & Kessler, 2010).

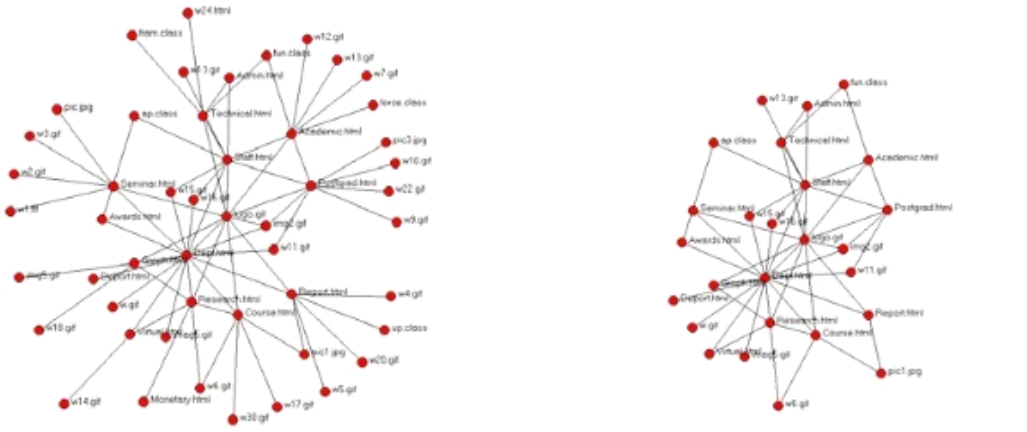
Once the evidence has been preserved, the job of the analyst is to trace the evidence from the victim and link it to the perpetrator. The objective is to discover a chain of events validating the criminal activities. Digital evidence may be used to support the identification of a crime committed with a digital device or to corroborate a traditional crime (Trček et al., 2010).

The analyst's results are used to develop a presentation that clearly links the source of the evidence to the crime and explains how it relates to the perpetrator. An effective hypothesis of guilt or no guilt is derived for the purpose of convincing members of a

court proceeding (Wang, 2007). Difficulties do arise in courts, which—being made up of lawyers and judges—often fail to understand interworking of computer systems (Bhat et al., 2010; Trček et al., 2010).

Visualization

Huang, Eades, and Lai (2005) chronicled the many works that present detailed visualization models outlining processes, designs, and guidelines. One such model is that of Shneiderman (1996, 2008), who summarized visual design with a mantra: "overview first, zoom and filter, then details on demand." From the user's perspective, the mantra names tasking capabilities, which may additionally include identifying relationships, extracting data subsets, and tracing actions, for example, *undo*. These tasks may be specific to associated data type characteristics. Similarly, Chi (2000) described critical visual operations as a workflow pipeline with an underlying data structure that itself remains unchanged while an analytical transformation for up to 36 visualization techniques is performed. Graph visualization is well suited to the presentation of file system or relational data in visual form whereby nodes represent entities, and connections between the entities represent relationships with limitations (Hansen et al., 2010; Huang et al., 2005). Hansen et al. stated that node analysis was first studied in 1736 by mathematician Leonhard Euler and revisited by Paul Erdos and Alfred Renyi, who developed Graph Theory in the 1950s. Figures 1 (a) and (b) illustrate simple graphs of relational information datasets with visualized and unwanted information filtered out for close analysis.



(a) A graph with 50 nodes. (b) A graph filtered with 27 nodes.

Figure 1. Simple graphs representing relational information (Huang et al., 2005)

Tufte (2000) observed that scientific visualization has for centuries been able to present in maps and statistical graphics the fundamental context of size, orientation, and label. He explained how to improve images with dimensions of direction, timelines that may be added to graphics (depicting movement), and temporal sequences (providing additional quantitative order to visual presentations).

Visualization of Data

Visualization has for some time been looked to as a means of analyzing large amounts of accumulated computer data. Ferster (2013) saw the human limitations of analyzing large amounts of raw data possessing multidimensional relationships. Researched visualization techniques are found to be useful in fraud detection and business data mining; even today's social media networks, such as Twitter and Facebook, are being visually analyzed. Data are continually being collected in our daily lives, for example, during credit card transactions and telephone calls. These data are automatically accumulated and stored by computer systems in ever-growing volumes. Specifications of

these stored transactions provide multiple dimensions of information (Hansen et al., 2010; Keim, 2002).

The visualization of information wherein the data are deficient of two-dimensional (2D) and three-dimensional (3D) properties is inherently difficult to map to digital displays. However, identified display methods for visualizing information, such as x-y plots, line plots, and histograms, are available. Display methods are beneficial to the analyst of information whenever a dataset overview is needed to present graphically an entire single-dimensional dataset. Although they are limited in their dimensional capacity and relatively ineffective in small datasets, many different techniques have been developed to represent graphically datasets in multidimensional fashion wherein only single-dimensional data are available. These methods include approaches such as parallel coordinate visualization and dense pixel displays' recursive pattern techniques (Keim, 2002).

Parallel coordinate visualization represents multidimensional data elements, which were explained by Inselberg and Dimsdale (1990) as each data dimension's being represented by a vertical line and each data item represented with a horizontal polygonal line. The polygonal line intersections exhibit relationships among the other pieces of data. Vidmar (2007) described parallel coordinates as useful in data mining large datasets, frequently found in biomedical research. Figure 2 represents a parallel coordinate plot and associated data set created in Microsoft Excel.

	A	B	C	D
1	x1	x2	x3	etc
2	2	1	7	9
3	3	4	3	6
4	2	5	6	5
5	3	4	7	8
6	1	2	3	9

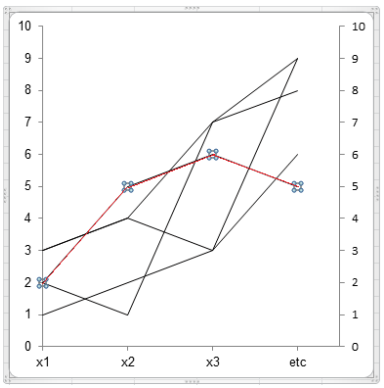


Figure 2. Parallel coordinate plot and associated data set

Keim (2000) described in theory dense pixel techniques as mapping data elements by dimension value of colored pixels and grouping the pixels into adjacent areas belonging to dimensional sectors. One pixel represents one datum value, thus limiting the largest amount of data as those that can be presented by a certain number of pixels within a display screen (Inselberg & Dimsdale, 1990; Keim).

Finding and correlating valuable information proves difficult by use of textual-based interfaces, wherein only a fraction of the data is displayed. As noted by Shneiderman, Dunne, Sharma, and Wang (2012), it is challenging to explore with frequent scrolling and make sense of millions of data items when only hundreds are presented. The prospect is limited for finding unknown or new hypotheses from small amounts of data. Visualization uses graphical technologies to present large amounts of data (Osborne et al., 2010). However, when data exploration using visualization techniques adheres to the information-seeking mantra presented by Shneiderman (1996), Shneiderman (2008), and Shneiderman et al.—overview, zoom, and filter—likelihood of the discovery of otherwise invisible data is significantly increased (Keim, 2002).

Visual data exploration was explained by Keim (2002) and Ferster (2013) as needing the human's flexibility, creativity, pattern-recognition capabilities, and knowledge to be effective in mining the volumes of storage capacity of the modern computer system. The visualization objective is to take advantage of the human's perceptual capabilities in the discovery of useful information within very large data sets. Textual-based displays underutilize humans' abilities. Visualization techniques, on the other hand, enhance humans' perceptual capabilities to rapidly recognize and recall images—detecting changes in size, color, shape, movement, or texture—according to Shneiderman (1996).

When properly implemented, visualization allows the human to interact directly with large datasets. Even when the analyst is not familiar with the data, this direct interaction becomes useful in facilitating the capability to rapidly change data-exploration goals. These visual tools and techniques outstrip automated data-mining techniques and statistics, especially when the data under examination are noisy—even when the analyst has little understanding of intricate algorithms or statistics. Exploration of digital information comprised of large datasets is faster when presented visually and often produces better results (Keim, 2002; Shneiderman, 1996).

Palomo, North, Elizondo, Luque, and Watson (2012) employed visualization to analyze large data sets of network traffic logs to discover both human and machine anomalies. Their work presented two visualizing techniques to analyze data, replacing often error-prone and time-consuming manual processes. Information flows in the human memory as a three-part system: 1) sensory registers receive information such as the visual and auditory, 2) short-term memory processes the information as strategies and decisions in working memory; short-term memory is limited in capabilities, and 3) long-term

memory stores information to be retrieved for later use (Atkinson & Shiffrin, 1971). Huang et al. (2009) viewed the working memory as being responsible for processing a limited number of cognitive tasks.

Visualization provides cognitive support to data analysts, according to Huang et al. (2009). This support—visual representation of data—reduces cognitive process workloads by reducing the demand on the human memory, visualization functioning as an (external) extension of the memory.

Visualization and Forensics

Shneiderman, Plaisant, Cohen, and Jacobs (2010) described visual presentations as being easier to comprehend than textual displays. Interactive, compact presentations capable of visual data-mining enable the human perceptual system to answer even those questions that have not been asked. Increased volumes of data increase demands on investigators in event correlation in digital forensics evaluations (Osborne & Turnbull, 2009). Visualization techniques have the potential to integrate human perception into the data exploration process of large datasets, according to Keim (2002). As previously mentioned, digital forensics tools are available both commercially and as open-source products. EnCase forensic software is a commercially available tool touted by its producer as the premier computer forensics application on the market. EnCase enables investigators to explore and preserve evidence in file format. The open-source tool TSK provides a library of command line tools for the investigation of file system data useful to finding evidence. The Autopsy Forensic Browser incorporates a graphical interface to the command line digital investigation tools in TSK. Together, Autopsy and TSK allow a

computer file system to be investigated. However, the Autopsy graphical interface, similar to EnCase, is a textual-based interface with limited capacity for presenting large amounts of information to the investigator (see Figure 3) (Sleuth Kit, 2012). Augmenting the digital forensics interface with visualization increases the investigator's working memory, thus increasing processing capacity.

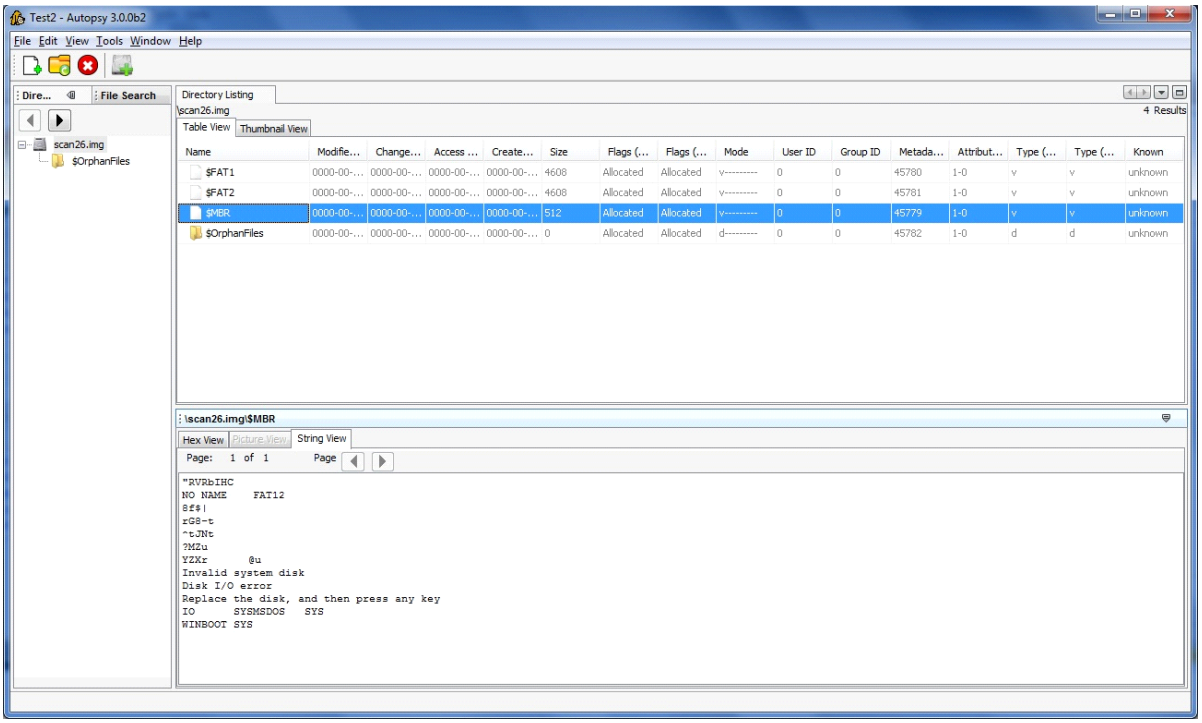


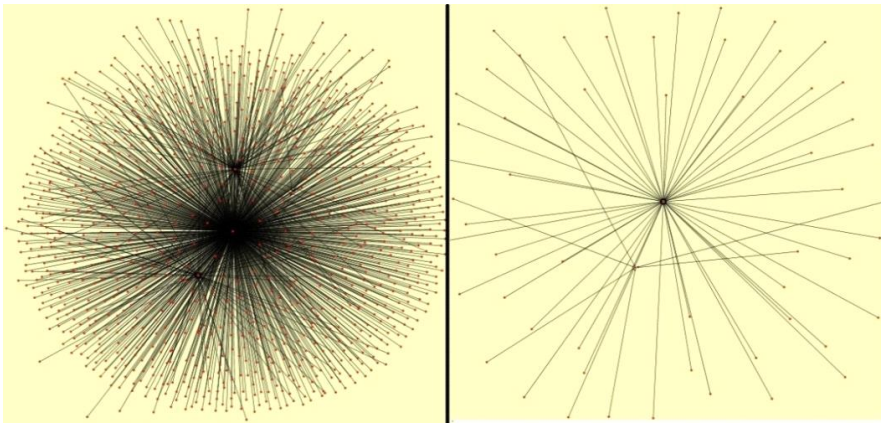
Figure 3. Sleuth Kit-Autopsy interface Digital Forensics Application—textual-based interface (2012).

Jones (2008) pointed out that gathering digital forensics data for a particular case is both time-consuming and expensive. However, he cited work by other researchers, such as Conti and Dean, who demonstrated how visualization increases the analyst's efficiency by speeding up the process of file content examination. He explained that visualization interface techniques provide the opportunity for previously overlooked or undiscovered information to be identified. Additionally, Jones held that visualization is capable of aiding in piecing together fragments of data stored in computer memory.

The basic graph as seen in Figures 1 a) and 1 b) is limited to only a few hundred nodes at best (50 shown here). However, information visualization applications often need to address thousands or millions of nodes. Two main techniques to overcome the limitation in node presentation were identified by Huang and Eades (1998) as clustering and navigation. Clustering creates super nodes that summarize like nodes. Navigation requires interactivity between the graphic and the user. Shneiderman (1996) and Shneiderman et al. (2010) explained that the interactivity allows the user to zoom in for analysis of a subset of nodes. For demonstration, used in this report is the open-source tool NodeXL, an extension of the Microsoft Excel spreadsheet tool used for network analysis for learning, concept presentation, and visualization (Hansen et al., 2010). Figures 4 a) and b) exemplify how a complete dataset and a subset of nodes may be presented graphically. Fu, Hong, Nikolov, Shen, Wu, and Xu (2007) explained how malicious intrusion upon a corporate email system is recorded and traceable in server log files for analysis. However, text-based log events may be made to expose graphically results that exhibit unexpected findings, which—when introduced textually—may actually go unnoticed or may otherwise be discounted as noise. Huang, Zhang, Nguyen, and Wang (2011) employed the clustering type of visualization as an experiment showing effective use for analysis of network security. Their experiment displayed a data file representing 4100 spam emails that originate from 450 locations. Clustered structures demonstrated graphically were meant to enable the analyst to identify unusual events and certain types of spam email attacks, spam emails being reviled for clogging email systems and robbing networks of bandwidth capacity.

The graphical techniques presented by Huang et al. (2011) for evaluating spam email datasets form the bases for the thesis in this work: digital forensics tools benefit from the addition of interactive graphical displays to their interfaces. Relational properties of file-centric digital evidence—such as file type, creator, size, date created/accessed, etc.—lend themselves neatly to graphical presentations. According to Osborne and Turnbull (2009), those graphical enhancements provide an abstract representation of data that increases the analyst's capacity to obtain knowledge. Moreover, data abstraction increases the cognitive absorption of information while amplifying relationships within the datasets. Filtering easily removes or masks irrelevant data while enhancing or highlighting information of interest.

Digital Corpora (2011) provides randomly generated file sets for researchers of simulated digital forensics evidence. These files, meant for testing tools and practicing digital forensics analysis, are arranged with the relational attributes of filename, last modified, date, time, size, and description. Figure 4 a), again using NodeXL, visualizes digital forensics evidence as a dataset of 1000 files simulated by Digital Corpora. In Figure 4 b) the file set has been filtered to display only those files with attributes of computer graphics and utility. Lastly, Figure 5 introduces details on demand by presenting information about a single file from the simulated investigation.



a) Entire evidence dataset of 1000 files is visualized.

b) The dataset appears filtered, rendering only those files of interest.

Figure 4. NodeXL visualization of simulated digital forensics evidence (Hansen et al., 2010).

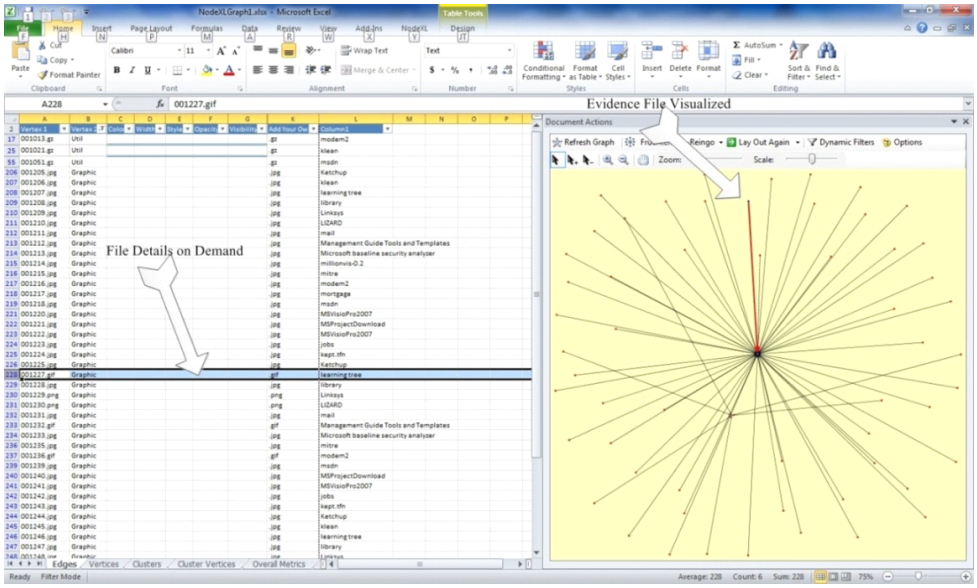


Figure 5. NodeXL detail on demand of a single simulated digital forensics evidence file (Hansen et al., 2010).

Schrenk and Poisel (2011) stated, “Digital crimes are increasing, [sic] so is the need for improvements in digital forensics. Visualization allows for displaying big amounts of data at once, so a forensic investigator is able to maintain an overlook about the whole

case" (p. 758). Peisert et al. (2008) presented FBI findings revealing that in 2002 half of the cases investigated involved a computer, and RCFL (2013) reported that all 2012 investigations involved a computer.

As computer software improves with advances in display resolution, increasingly significant becomes the question of how visual presentation improves the effectiveness of the interface over textual-based information. Research findings show how digital forensics tool interfaces incorporating these visualization techniques improve the capability of the investigator to identify and gather digital evidence more effectively during the analysis process of a criminal investigation. Today's digital forensics investigative technologies are over-reaching the investigator's ability to identify evidence rapidly and efficiently from digital devices suspected of having been used during the execution of a crime. Previous research has showed how the advancement of digital forensics tools improves investigations, especially when the event is time-sensitive (Osborne & Turnbull, 2010; Richard & Roussev, 2006; Shneiderman, 1996).

Visualization in computer applications provides humans the ability to take advantage of their inherent physiological capability to process large amounts of information visually. Visualization techniques and technologies were applied in multiple domains, including medical, military, and business (Krasser, Conti, Grizzard, Gribschaw, & Owen, 2005; Osborne & Turnbull, 2009). Krasser et al. proved how IT security tools, such as IDS, which generate large amounts of network security data, waste humans' resources, capabilities, and time in effectively analyzing large numbers of network traffic patterns. Large sets of digital forensic data may be better examined through visualization, enabling the investigator to rapidly scan these data sets.

Visualized information processes graphic evidence, aiding the analysis phase of an investigation by identifying that which may be suitable for conveyance to a colleague or to a court of law. Other graphical techniques may potentially be used to examine simultaneously multiple devices containing evidence, such as mobile phones and other forms of portable digital storage devices. Similar techniques were successful in the domain of security to visualize large amounts of network data or logging information found in firewalls. The increased volumes of data and number of devices require enhanced capabilities of tools, and analysts require new techniques in evidence discovery (Osborne & Turnbull, 2009).

Entire datasets of digital evidence can be prototyped, presented, and filtered to provide the user with interactive details on demand about digital evidence for analysis. These prototyped principles follow Shneiderman's (1996) information-seeking mantra, incorporating visualization into a digital forensics frontend application. Osborne et al. (2010) chronicled digital forensics tools as critical in providing analysts the utility for detecting and discovering digital evidence of a crime. Increased volumes of data increase demands on investigators in event correlation and in analyzing other relational attributes found in digital evidence (Osborne & Turnbull, 2009). Visualization techniques have the potential to integrate human perception into the data exploration process of large data sets, accordingly reducing the burden of analysis on the investigative practitioner (Keim, 2002; Osborne & Turnbull).

Summary

A considerable amount of research detailed digital forensics tools and processes. Additionally, several researchers called for the improvement of digital forensics tools through the use of visualization (Hoelz & Ralha, 2013). Visualization design principles, when applied to the software interface, offer capabilities to present very large datasets to the user, bringing resolution to multi-dimensional information (Shneiderman, 2008; Tufte, 1990).

The hypothesis tested by this research is that the cognitive load of the practitioner is reduced whenever visualization has been integrated into the digital forensics application interface. Huang, Eades, and Hong (2009) validated the premise that cognitive load is consistently reduced in applications into which visualization has been integrated, proving such applications superior to textual-based interfaces. In this study ethnographic research identifies the primary duties of the digital forensics analyst during the analysis phase of investigative discovery while cognitive load is measured for purposes of comparison of mental workloads as reported in the user experience of the visualized interface and the textual-based interface.

This research contributes to the body of knowledge by validating a method of measurement and by providing empirical evidence consistent with theory introduced and hypothesis asserted in this study: the use of the visualized digital forensics interface provides a more efficient workload for the analyst, saving labor costs and compressing an analyst's time in the digital investigation discovery phase.

Chapter 3

Methodology

Overview of Research Methodology

This study is based on a mix of ethnographic observations and cognitive load measures. Traditional ethnography as described by Fetterman (2010) was established within the social sciences for research to be conducted in a natural setting rather than under simulated conditions. Fetterman pointed out that ethnography establishes the context of the human environment as it occurs in nature. In their study of technologies, Ormerod et al. (2005) described ethnography as an effective way to discover best practices and common patterns in work processes: the perspective of research participants ethnographically describes the social context in their settings. Although such studies have been conducted for many decades, these types of studies are not problem-free. Presenting study findings to system designers, for example, has proven to be difficult because findings are often not orderly or clearly stated; erratic options prove difficult for engineers to understand. Such problems result from misunderstood observations or disruption of normal operations (Hughes, King, Rodden, & Andersen, 1995).

Cognitive load theory (CLT) is based on the principle of cognitive architecture that working memory is limited. According to Paas, Tuovinen, Tabbers, and van Gerven

(2003) an appraisal of cognitive load may be accomplished by measuring "mental load, mental effort and performance" (p. 64). Both analytical and empirical methods have been researched in measuring cognitive load. Sweller (1988) used analytical methods such as expert opinion and task analysis to measure cognitive load. Empirical methods using rating scales similar to self-ratings are described by Paas et al. (2003) as the context for CLT whereby people report with accuracy their mental burden. Huang et al. (2009) defined cognitive load—also referred to as memory demand—as a measure of the required amount of cognitive capacity needed to perform a given chore.

Specific Research Methods Employed

This study followed ethnographic study guidelines—merged to measure the use of cognitive resources of the participants. Guidelines for an ethnographic study, according to Rose, Shneiderman, and Plaisant (1995), include preparation, actual field study, analysis, and reporting. Preparation includes identifying specific culture and policies of the study participants' organization, becoming accustomed to the digital forensics tools and the tools' history within the organization, preparing goals and questionnaires for the study, and gaining physical access to conduct the study. The field study itself—through observation or interviews in the workplace or home rather than a laboratory setting—is meant to develop amicable relationships among the researcher, managers, and users. The researcher must uncover from the study participants' surroundings, environmental clues and must document findings.

Cognitive load was self-measured by the study participants' using Paas' (1992) seven-point Likert scale in context for the CLT. The scale of numerical values,

distributed from one to seven, was used to record perceived mental effort in increments from “very low mental effort” to “very difficult.” Paas et al. (2003) reported that subjects were quite capable when reporting their own expended mental effort. Although suggestive, this scale has been found to be reliable, non-intrusive, and sensitive to variations in mental effort, as documented by Huang et al. (2009).

Analyzing the findings required the assembly and aggregation of testing results by calculating the statistical values from the study’s CLT findings. Textual data were compiled by group or class as comments, needs, or suggestions. Rose et al. (1995) recommended that numerical information from Likert self-assessments present the mean of the tasks by the relevant systems tested and reviewed for patterns. The results may be seen in Appendix A of this document.

Rose et al. (1995) identified finalization of the process as interpreting observations and distilling the goals and procedures found by the study. Variations upon reporting must consider audiences and their needs. Finally, reports were to be prepared and results presented.

The aforementioned process of preparation and organization seems requisite for any study. However, in this instance, a need prevailed to reevaluate the series of actions in order to achieve results for individual circumstance. Careful modeling of how the research is conducted positively influences both users and managers. Attention to details, such as use of the vernacular in communication, establishes trusting relationships within the observed community. See Appendix B for the study process flow.

The accumulation of findings—participants’ comparison of textual and visualized interfaces—defined their accounts of differences and likenesses between interface types.

A summary and final report explaining that the hypothesis (*H*) of an improvement employing the visualization approach has been realized concludes the study, as prescribed by Saraiya, North, and Duca (2010).

Instrument Development and Validation

Shneiderman and Plaisant (2006) suggested that the ethnographic strategy holds many benefits when thoughtfully planned and executed, for instance, by building study participants' trust and confidence in the researcher. With the researcher being resident for the study, camaraderie among the subjects and researcher naturally develops, providing meaningful input, in this case, for software application interface design. The abovementioned ethnographic evaluation guidelines by Shneiderman and Plaisant relate to an interface redesign study conducted by Rose et al. (1995).

Research conducted by Shneiderman and Plaisant (2006) presented "multi-dimensional in-depth long-term case studies (MILCs)" of information visualization. The MILCs are an ethnographic method that establishes evaluation goals for visualization research studies. With studies differing, MILC guidelines are meant to be flexible to accommodate the individual research project. By presenting an entire dataset of evidence to be examined, this study-(focusing on visualization) evaluated how the digital forensics application with a textual-based interface may bolster the digital forensics analyst's abilities. Following the basic information exploration approach outlined by Shneiderman (1996), the procedure displayed what was initially text-based information, first, presenting an overview of the entire dataset; second, providing an interface capable of zooming and filtering for items of interest; and then discovering the ability to drill down

to specific data for details. For this study—to gather information about understanding and refining digital forensics tools and improving the achievements of the analyst—MILCs seem well suited. See Appendix B for the study's process.

Population

Participants were selected according to minimum standards set for this study: 1) subject matter experts (SME) or digital forensics experts and 2) those of the first-responder community, such as network administrators and security professionals.

Participant SMEs of the first phase of the study were an ethnographic research group whose input helped to develop the primary discovery tasks in how to perform the digital forensics investigation. The SMEs group was comprised of digital forensics specialists who are experienced in law enforcement or other investigative professions wherein digital forensics is their primary duty. Additionally, according to the U.S. Bureau of Labor Statistics (2012), private detectives or investigators of computer crime are college-educated or have completed dedicated on-the-job training. Pendergast (2010) described digital forensics professionals as experts in operating systems and may have a specialty niche area such as mobile devices. Furthermore, Pendergrast reconized professional digital forensics certification as a differentiator when searching for employment and a confidence booster for the practitioner. The task development SME study—comprised of a small sample size, two or three participants—emulated a model by Hughes, King, Rodden, and Andersen (1995), wherein few participants were engaged in an ethnographic study.

For the second part of the study, a group of first-responders established an experimental group. These participants modeled the work of Shneiderman and Plaisant (2006), employing various levels of skill and education in order to avoid homogeneity in the result set. Specifically, the first-responders were selected according to U.S. Department of Defense (DoD) (2012) workforce guidelines for persons who collect data from intrusion detection systems, firewalls, and network traffic and system logs—using a variety of tools to analyze events that occur within their environment. The first-responders' minimum qualifications were two years' experience using analysis tools tactics, techniques, and procedures for evaluating computer event information. Additionally, these first-responders hold, minimally, professional certification—such as A+, Security+, Network+, Certified Information Systems Security Professional (CISSP), or Systems Security Certified Practitioner (SSCP)—as outlined in guidelines for personnel with privileged system access in the U.S. DoD (2012) workforce. These participants' having a general knowledge of the legal aspects of digital evidence collection (20 first-responders forming an experimental group with similar skill sets) ensured effectiveness of the evaluation, advised by Cohen (1988).

Criteria for the first-responders to participate in this study were that they be network administrators, as modeled by Endicott-Popovsky and Frincke (2007); first-responders in the digital forensics evidence-gathering processes; and information security professionals especially trained in IDS and digital forensics. Shneiderman and Plaisant (2006) suggested that domain knowledge levels of participants vary to provide contrasting perspectives. Initial testing was scheduled face-to-face and in advance with selected first-responders for tool introductions and testing. If clarification was not needed, subsequent

interviews were not conducted. The user community defined success in various ways when accomplishing their assigned tasks. In this case, a variety of participants from the evidence-gathering community provided that diverse perspective.

In order to collect data for this research, a simplified version of the MILCs was implemented in the following steps:

1. The University Institutional Review Board Memorandum provides approval for the protocols to be utilized while testing human subjects for this study (see Appendix C).
2. Specifications for the research were outlined and distributed in advance to the participants in the Nova Southeastern University Institutional Review Board (IRB) consent for participation form (see Appendix D).
3. Research was conducted.
4. Three SMEs with specific qualifications provided input through ethnographic interviews related to digital forensics.
5. Experimental group of 20 first-responders answered questionnaires (see Appendix E).
6. SME interviews and first-responder questionnaires were aggregated with findings.

Research Design

This research tested the hypothesis that the digital forensics analysts operate at reduced cognitive load, yet demonstrate improving performance, when using an interface employing visualization rather than a textual-based interface.

The ethnographic segment of this study provided the method to answering research question one, seen below, through interviews with three SMEs. As the portion of the study measuring cognitive load contrasted visualized and textual-based applications, each

of 20 first-responder participants provided answers to the following research questions two and three.

1. What are the investigator's primary tasks for evidence identification while operating a traditional digital forensics tool set?
2. What is the cognitive load of an assessment of human working memory beyond just time and accuracy measures while performing ethnographically discovered techniques of a predefined set of tasks to establish a baseline of evidence identification?
3. What efficiency level may be attained for a digital investigation improved for the benefit of analysis by application of visualization to predefined tasks?

Each SME and first-responder participant spent minimally one hour providing input for the study. Some spent up to one and a half hours providing input.

Arithmetic Mean

The arithmetic mean is the most widely used method for testing how a population leans toward a hypothesis, according to Cohen (1988). The sample of *n* cases of a population, randomly selected, is tested to form around a mean for a researcher to prove that a null hypothesis (H_0) exists. In the case of this research, *n* cases or how many first-responders are needed for testing to prove H_0. H_0 for this study is affirmed if there is no cognitive load difference between the mean of the user experience of the textual-based interface digital forensics tool set (m_A) and the visualized (m_B) or H_0: $m_A - m_B = 0$. However, research significantly rejecting H_0 proves that a phenomenon of the hypothesis

(*H*) exists or, in this case, that the visualized display reduces the cognitive load on the tested mean of the population, stated as (H_0: m_A - $m_B \neq 0$) (Cohen, 1988).

In order for *H* to be acceptable and ensure that the occurrence is not by chance, there must be a statistically significant departure from the mean of H_0. The criteria of significance set for rejecting H_0 is referred to as the alpha level (α). According to Eng (2003), α is most often set to .05 or 5 %. The smaller the α, the larger is the sample or *n* cases needed for a credible study result. With α set to .05, the risk is small, the risk being false rejection of H_0 5 % of the time—called a type I error. The selected beta (β) level provides protection against the type II error or false acceptance of H_0. Traditionally, β is set by the statistical power of (*P*) where ($P = 1 - \beta$). Araujo and Frøyland (2007) considered .80 or 80% a suitable *P* or β of .20 or 20%.

_t_-test

The *t*-test was selected as the experimental design for testing H_0 in this study. Simply, the *t*-test compares two means, often in before-and-after studies, either independently or dependently. This type of testing contrasts two means to see if they differ significantly from one another (Urdan, 2010). The most widely used *t*-test is the independent test whereby, for example, grade observations of 50 girls and 50 boys in an elementary school are compared. Another type of *t*-test, the paired or dependent sample test, was used in this study. Urdan described the dependent *t*-test as testing determining the difference in the means to check a sample taken from a single population. In this case, cognitive load from a sample population utilizing both a textual-based display and a visualized display was tested as participants performed a given task, creating a distribution of scores. Both pre-

test and post-test averages on a single sample may generate a distribution of scores (Urdan).

The probability of the rejection of H_0 in this study shows a reduction in cognitive load for the participants rather than showing a more centralized distribution. In normal distributions, whenever the mean or median may fall on either side of the distribution, a two-tailed test is conducted. In this case, a directional relationship was expected; therefore, the experiment relies upon a single-tailed test. In order to compensate for a skewed distribution where the mean is found on a single side of the resulting distribution, the significance of α for accepting or rejecting H_0 is halved or changed from .1 to .05 or 5%, ensuring a strong argument for rejection (Prajapati, Dunne, & Armstrong, 2010; Urdan, 2010).

Sample

In order for first-responder participants to validate this study, the sample size estimation and power analysis required a sample sufficient to detect real effect of the research (Prajapati, Dunne, & Armstrong, 2010). Cohen (1988) described standard effect size "*ES* index" as small, medium, and large or *d* value. Cohen's conventional framework for power recommends *ES* be set .2 for small, .5 for medium, and .8 for large. The smaller effect sizes require a larger sample size. Cohen provided tables as a primary utility in developing sample size, employing the aforementioned factors: significance criterion value of α set to .05; the *ES* index of a large *d* value set to .80; and the power of *P* value set to .80. See the α of .05 set for the table demonstrated in Table 1, depicting how Cohen developed a sample size *n* of 20—also utilized in this study.

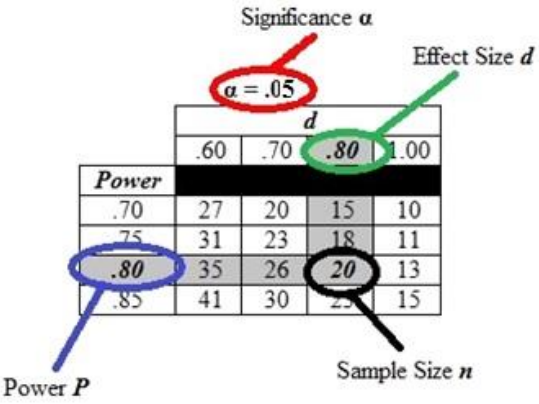


Table 1. Power of *t*-test Sample Size (Cohen, 1988)

Data Collection Procedures

Immediately following completion of each of the ethnographically discovered tasks, the first-responder participants were asked to complete a self-report using a questionnaire designed to assess their invested mental effort (Paas & van Merriënboer, 1994). In order for the participants to maintain anonymity, each study subject's questionnaire was distinctly marked by an assigned ID number (see Appendix E) based on the order of their testing—one through 20. Participants were questioned according to a Likert scale to assess and report mental effort required to perform the assigned tasks. The seven-point Likert response scale ranges from 1-2 (low mental effort) to 3-5 (neither easy nor difficult) to 6-7 (very difficult). Subjective ratings of cognitive load were chosen because they are easy to implement, do not interfere with the primary task, and have been used successfully in previous CLT research (Paas & van Merriënboer). See questionnaire in Appendix E.

Resources

Resources of facility, equipment, and software were provided by the researcher whereas resources needed for the deployment of a prototype application were acquired as cost-free, open-source materials. The three SMEs and 20 first-responder human participants were not compensated.

Summary

The research design reflects a method based on CLT to present empirical evidence demonstrating the hypothesis (*H*) in determining that the use of visualization in the identification of digital evidence reduces the digital forensics application user's cognitive load. The goals of this study are explicitly defined in research questions as recommended by Eisenhardt (1989), which lay out the specific path for the researcher to focus efforts. Ethnographic interviews of three SMEs were performed in this study to develop evaluation tasks of primary file-centric activities found in evidence discovery of a digital forensics investigation. These primary investigative activities were the base for the evaluation measures in comparing the visualized and textual-based digital forensics applications. This research adopted the self-appraisal aspect of a system interface determined by users' observation levels of cognitive load patterns (Rose et al., 1995).

A *t*-test as described by Cohen (1988), the design methodology used in this study, is intended to compare two sample means testing for the null hypothesis (H_0). In order to reject or accept H_0 or H there must be a significant set criteria or alpha level (α) of deference when testing two means—in this case, the mean cognitive load when

comparing the visualized and textual-based digital forensics applications**.** To ensure this high level of reliability of confidence, this study's appropriate sample size subject population selection is set to 20, following the procedure for a power analysis, as suggested by Cohen.

Data collection for this test was achieved by compiling the results of 20 first-responders performing the primary file-centric investigative tasks on both the visualized and textual-based digital forensics applications. Following execution of each of the ethnographically discovered tasks, the first-responder participants were asked to complete a self-report using a questionnaire designed to assess their invested mental effort (Paas & van Merriënboer, 1994). The results were compiled and the mean comparisons tested for the null hypothesis (H_0).

Chapter 4

Results

Research

The basis of this research is the premise that the cognitive load of the digital forensics analyst can be reduced by his or her use of a visualized display rather than a traditional textual-based presentation. To test this hypothesis, as discussed in Chapter 3, the research was conducted by means of two separate methods and two groups of participants. To preserve the privacy of those supporting the study, each was assigned a participant ID provided by the researcher as indicated on an associated questionnaire (see Appendix E).

The first group of participants were SMEs, experts in the field of digital forensics, possessing at a minimum a professional certification of CCE. Based on the CCE, the SMEs are product tool vendor neutral and have proven to be proficient in the digital crime examination and analysis. The SMEs participated in ethnographic interviews for the purpose of generating the primary tasks to be employed for the second group of participants. The SME participants hold various positions within the digital forensics community: policy developer, department director, and malware reverse engineer. However, each is an expert in the investigative field of digital forensics and has performed many digital forensics criminal investigations.

Tasks resulting from the SME ethnographic interviews provided a list of items applied as part of a simulated criminal investigation performed in this study. The identified tasks are those file-centric activities often performed by digital investigators, see Table 2.

Task	Task Description
Task 1	Locate files with the .jpg extension
Task 2	Locate the file named Kittie.jpg
Task 3	Date range to establish timeline
Task 4	Identify the size of a directory structure
Task 5	Identify the largest file

Table 2. Task List

The second group of study participants were tasked to test the role of the first-responders of a digital investigation for this study. To qualify, this group of study subjects all perform some level of network administration in their daily work activities, being persons who collect data from intrusion detection systems, firewalls, and network traffic and system logs to analyze events occurring within their environment. The first-responders' minimum qualifications were two years of experience analyzing computer event information and holding professional certification—such as A+, Security+, Network+, Certified Information Systems Security Professional (CISSP), or Systems Security Certified Practitioner (SSCP). The tasks performed by the subjects are those often accomplished by digital investigators during an actual computer system examination, mentioned previously as having been identified by the SME study group. The first-responders examined a flash drive simulating a storage device utilized by a perpetrator as part of a criminal act.

Testing one's cognitive load is accomplished simply by the test-taker's assessing the difficulty of a task performed. The research shows that individuals being tested provide the best measure for estimating cognitive load (Paas, 1992). The first-responders established the experimental group. These participants modeled the work of Shneiderman and Plaisant (2006), that is, employing various levels of skill and education in order to avoid homogeneity in the result set. Furthermore, these participants possess general knowledge of the legal aspects of digital evidence collection.

Data Analysis

The experimental group of 20 participants with similar skill sets ensured effectiveness of the evaluation (Cohen, 1988). Data were collected as a dependent-sample *t*-test of subjects' repeated measures to assess the cognitive load induced when answering the questions found in the questionnaire (see Appendix E) and measured against the dataset for each of the requisite digital forensics applications (two conditions on one measure). The applications, one visualized and one textual-based, provided the prototypes for the experiment. Final testing results are that the mean varies between the paired/matched observations, differing significantly from zero. That is, the dependent-sample *t*-test procedure detected a significant difference between the means of the two variables, in this study the variables being the subject results from operation of the opposing applications. The participants were assessed under the dual conditions for a single task, paired on the questionnaire developed from aspects of Shneiderman's (1996, 2008) visual design mantra.

The evaluation's *t* value ($t = 2.55$) being above the critical value of t_0 ($t_0 = 1.79$)

supports the hypothesis (*H*) that cognitive load on the digital forensics investigator is reduced significantly when operating a digital forensics application into which has been incorporated a visualized user interface. Conversely, the null hypothesis (H_0) was not supported: no change or an insignificant *t* value ($t <= 1.79$) occurs in the cognitive load for an analyst operating a visualized user interface as compared to cognitive load while operating a textual-based interface. The assessment concluded in the difference in mean values or delta (δ), which must result in a value of greater statistical significance than the confidence level alpha (α) of .05 established for this study in Chapter 3 in order to reject H_0. The testing results' raw data are recorded in the tables in Appendix A and the study's (*t*) distribution is detailed in the sections to follow.

Findings

The findings result from data collected from the 20 first-responder subjects. First-responders are similar in that they met minimum computer information system qualifications for investigating system logs—in addition to meeting experience and professional certification requirements. Participant demographics—such as geographic location, gender, age, or education—were not considered in this study.

The study results are based on the impacted participant cognitive load as measured by five tasks performed and judged by the participants, see Table 2. The measures are listed on the study questionnaire's seven-point Likert scale, ranging from "low mental effort" (1) to "very difficult" (7); the questionnaire is available in Appendix E of this work. Each first-responder participant was given study instructions, shown in Appendix F, for step-by-step performance of each of the five tasks on the identical datasets by utilizing,

independently, both the SequoiaView visualized analysis tool and The Sleuth Kit digital forensics tool set; instructions also explain the study's purpose and specify criteria for participant selection.

The cognitive workload impact differential between the textual-based and the visualized digital forensics tools on the 20 study participants represents a measure of difficulty in their performance of the five investigative tasks required by this study, see Table 2. The study participants answered a total of 14 questions referencing the applications' presentation following performance of each of the investigative tasks according to the instructions guide for this study. The mean of The Sleuth Kit results is 2.49 and that of the SequoiaView is 2.08 on a scale of one to seven, a reduction of 0.42 or nearly a 16% reduction in workload. The final differential value of the digital forensics tools' calculated cognitive load mean of all of the study responses' comparative results is recorded in the histogram, Figure 6, with the vertical axis representing the degree of difficulty as found in the respondent questionnaire, see Appendix E.

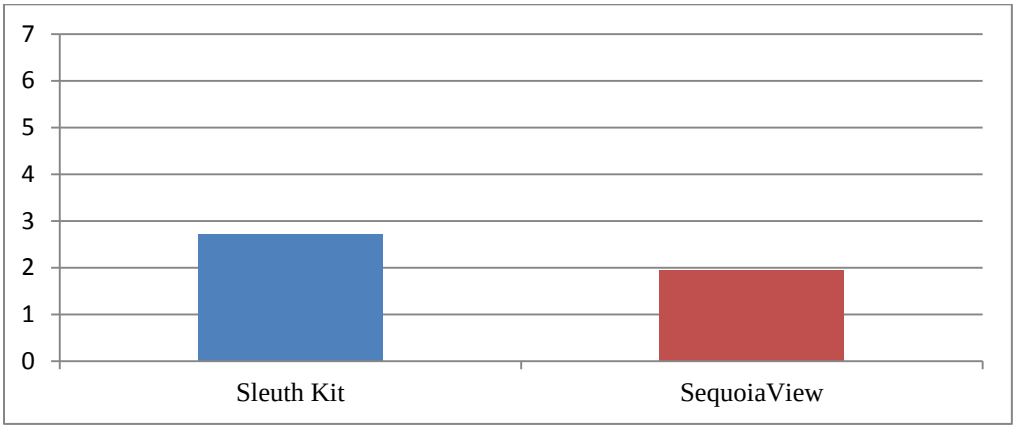


Figure 6. Mean Comparative Results

Measuring Mental Effort

The method for measuring mental effort empirically was accomplished by the study participants' self-rating the amount of effort required to perform the given tasks for the study. The participants' self-rating activity consisted of answering, in order, each of the 14 questions, following task completion, to register their perception of how easily they could identify patterns and clusters of information visually from the given application's presentation. The presentations represented a simulation of a suspected file system found on a thumb drive consistent with an investigation's file-centric properties. The rating scales provided a report of mental burden in locating suspected files from a perpetrator's illegal activities and the amount of cognitive capacity needed to perform given chores.

The sum mean value of all results from the participant surveys for each of the task area's cognitive load mean values is presented in Table 3. Each of the deltas (δ) for the task comparisons between the applications support the hypothesis (*H*) with each of the resulting factors demonstrating that the visualized application reduced the workload with each task performed. Proportionately, these factors, when evaluated, determine the impact of visualization in reducing the cognitive workload of the analyst.

Task	Sleuth Kit - μ1	SequoiaView – μ2	Deltas (δ) = μ1 – μ2
Task 1	2.86	2.10	0.76
Task 2	2.70	1.82	0.88
Task 3	2.71	1.89	0.82
Task 4	2.60	2.00	0.60
Task 5	2.71	1.90	0.81

Table 3. Task Mean Values

The histogram (Figure 7) displays the distribution of mean values for each task of all results from each of the participant surveys assessing the cognitive workload on the

vertical axis representing the degree of difficulty by the interface under test as found in the respondent questionnaire, see Appendix E.

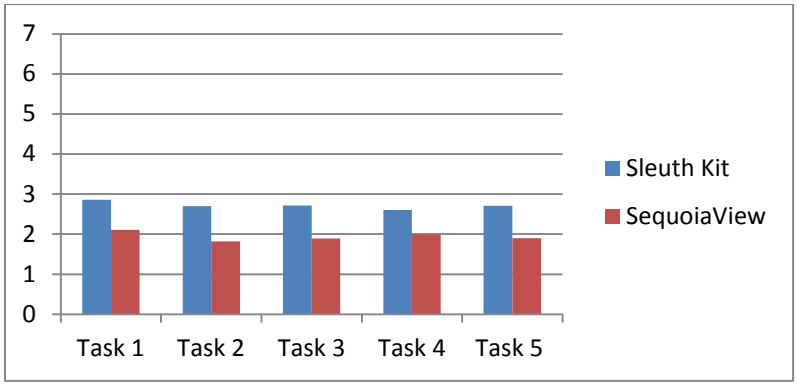


Figure 7. Total Subject Mean Difficulty by Task and Interface

The questionnaire results represent distributed values based on the respondents' perception of cognitive impact as related to visualization principles employed by the application under test. When mean values were compiled, the deltas (δ) from survey questions corresponded to the applications, supporting the hypothesis (*H*) with each visualized application's resultant mean values in the reduced mental workload among the first-responders across all questions. Proportionately, these factors, when evaluated, determined the impact of visualization in reducing the cognitive workload of the analyst; questionnaire survey questions' mean value and correlative-result-related deltas (δ) are presented in Table 4.

Question	Sleuth Kit - μ1	SequoiaView – μ2	Deltas (δ) = μ1 – μ2
Question 1	2.68	1.79	0.89
Question 2	2.51	1.55	0.96
Question 3	2.53	1.59	0.94
Question 4	2.32	1.57	0.75

Question	Sleuth Kit - μ1	SequoiaView – μ2	Deltas (δ) = μ1 – μ2
Question 5	2.63	2.05	0.58
Question 6	2.93	2.82	0.11
Question 7	2.61	2.06	0.55
Question 8	4.3	1.75	2.55
Question 9	2.62	2.05	0.57
Question 10	2.4	2.01	0.39
Question 11	2.56	2.15	0.41
Question 12	2.44	2.2	0.24
Question 13	2.74	1.96	0.78
Question 14	2.74	1.62	1.12

Table 4. Survey Questions' Mean Value and Correlative-Result-Related Deltas (δ)

The histogram (Figure 8) displays the distribution of values for degree of difficulty total mean by the interface under test for cognitive workload reported by first-responder subjects 1-20. Data showing the cognitive load mean for subjects by interface for each task are shown in Table 5.

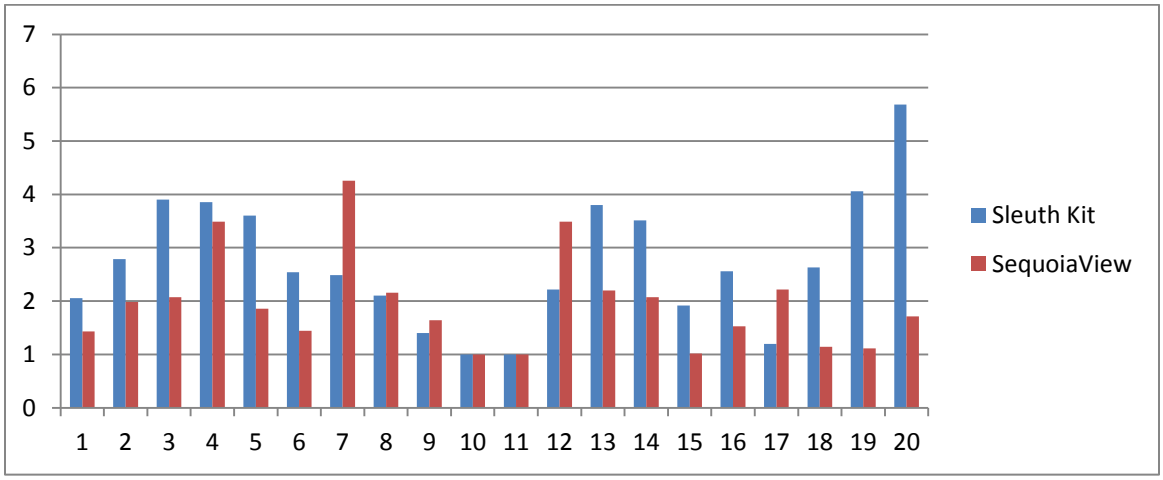


Figure 8. Total Subject Mean by Interface

Subject	Interface	Task 1	Task 2	Task 3	Task 4	Task 5	Mean
1	SequoiaView	1.36	1.58	1.36	1.43	1.43	1.43
1	Sleuth Kit	1.86	2.28	1.93	2.14	2.07	2.06
2	SequoiaView	2.14	1.64	2.21	2.07	1.86	1.98
2	Sleuth Kit	3.14	3	2.86	2.5	2.43	2.78
3	SequoiaView	2.29	2.07	2.29	1.86	1.86	2.07
3	Sleuth Kit	3.93	3.93	3.86	3.57	4.21	3.9
4	SequoiaView	3.36	3.21	3.79	3.43	3.64	3.49
4	Sleuth Kit	4	4	3.79	3.64	3.86	3.86
5	SequoiaView	2.43	1.71	1.71	1.71	1.71	1.86
5	Sleuth Kit	3.5	3.57	3.79	3.57	3.57	3.6
6	SequoiaView	1.5	1.43	1.43	1.43	1.43	1.44
6	Sleuth Kit	3	2.43	2.57	2.21	2.5	2.54
7	SequoiaView	4.36	4.36	4.14	4.29	4.14	4.26
7	Sleuth Kit	2.79	2.57	2.64	2.21	2.21	2.49
8	SequoiaView	2.57	2.21	2.07	1.86	2.07	2.16
8	Sleuth Kit	2.5	2.21	1.79	2	2	2.1
9	SequoiaView	1.93	1.79	1.36	1.5	1.64	1.64
9	Sleuth Kit	1.64	1.29	1.14	1.43	1.5	1.4
10	SequoiaView	1	1	1	1	1	1
10	Sleuth Kit	1	1	1	1	1	1
11	SequoiaView	1	1	1	1	1	1
11	Sleuth Kit	1	1	1	1	1	1
12	SequoiaView	5	3.64	3.64	3.57	3.5	3.87
12	Sleuth Kit	2.14	2.57	2.3	2	2	2.21
13	SequoiaView	2	2	2	3	2	2.2
13	Sleuth Kit	4	3	4	4	4	3.8
14	SequoiaView	3.07	3	3	3	3	3.01
14	Sleuth Kit	3.57	3.57	3.5	3.43	3.5	3.51
15	SequoiaView	1.07	1	1	1	1	1.01
15	Sleuth Kit	2.14	1.93	2	1.64	1.86	1.91
16	SequoiaView	2	2	2	2	1	1.8
16	Sleuth Kit	2.5	2.29	2.21	3.14	2.64	2.56
17	SequoiaView	3	2	6.29	3.5	3.29	3.61
17	Sleuth Kit	2	1	1	1	1	1.2
18	SequoiaView	1.14	1.14	1.14	1.14	1.14	1.14
18	Sleuth Kit	1.29	1	1	1	1	1.06
19	SequoiaView	1.43	1	1	1.14	1	1.11
19	Sleuth Kit	1.29	1	1	1	1	1.06
20	SequoiaView	2.5	1.43	1.21	2.21	1.21	1.71
20	Sleuth Kit	5.79	5.07	6	5.79	5.79	5.69

Table 5. Cognitive Load Mean for Subjects by Interface

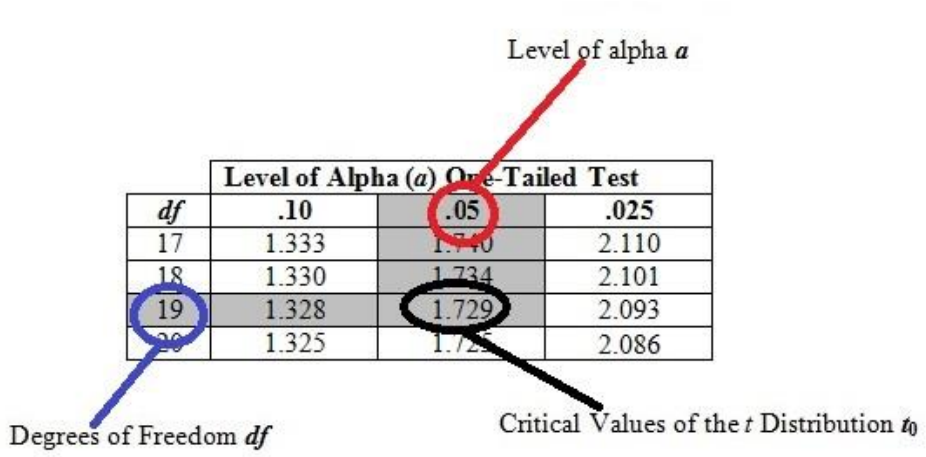
The questionnaire provided a method to collect cognitive load data empirically. The *t*-test was performed on the results of 2800 data points recorded in Appendix A. Results support the study hypothesis (*H*). The visualized digital forensics interface reduced the cognitive workload on the analysts.

Differences Between Sets of Conditions

Dependent-samples were used since the same subjects were tested and compared on both the textual-based and visualized applications. The dependent-samples *t*-test was performed on the calculated means of the cognitive load data collected from the questionnaire respondent results to determine whether the use of the visualized application produced a reduced cognitive workload for the first-responders compared to that produced by the textual-based presentation. The final analysis results are to follow in the remainder of this section. Microsoft Excel was used to calculate the mean and to produce the charts and result tables for the study. The analysis in the previous section of this document established mean comparison deltas (δ) for the study research questions, study tasks performed, and subject testing—overall each being positive; therefore, the testing for the null hypothesis (H_0) establishes a calculated mean as a single direction: the testing was conducted on the positive side of the rejection region (a one-tailed test).

The one-tailed dependent-sample *t*-test was conducted to compare the observed samples of cognitive load for first-responders to determine if there was a true difference under the testing conditions between the textual-based and visualized interfaces. In view of the statistical change in confidence due to varying sample sizes, the number of study participants is taken into account by calculating the degrees of freedom (*df*) as

recommended by Urdan (2010). The degrees of freedom (*df*) are calculated by adding together the sample size from the observation and subtracting 1 ($df = N - 1$), in this case 20 - 1 = 19 or $df = 19$. With the value of degrees of freedom (*df*) known, the critical values of the *t* distribution or the value of statistical significance may be determined by considering the confidence level of alpha (α) set to .05. According to Urdan's (2010) table of critical values of the *t* distribution or the level of statistical significance, (t_0) may be derived. With the degrees of freedom (*df*) being 19 and an alpha (α) level of .05, the level of statistical significance ($t_0 = 1.729$) is identified; see Table 6 (Urdan, 2010).



	Level of Alpha (*a*) One-Tailed Test		
df	**.10**	**.05**	**.025**
17	1.333	1.740	2.110
18	1.330	1.734	2.101
19	1.328	1.729	2.093
20	1.325	1.725	2.086

Table 6. Critical Values of the *t* Distribution (Urdan, 2010)

When calculating the final study results, according to Urdan (2010), typical variation or standard deviation between the scores of the respondents occurs. As with degrees of freedom (*df*), this variance is impacted by the number of study participants, see Equation 1 for the standard deviation formula as presented by Urdan (2010). The difference (*D*) is the difference between study scores; see Table 7 for study results, deltas (δ) or (*D*), sums, means, and D^2 (Urdan, 2010).

$$s_D = \sqrt{\frac{\sum D^{2-} \frac{(\sum D)^2}{N}}{N-1}} \quad (1)$$

Subject	Sleuth Kit μ1 Subject Mean	SequoiaView μ2 Subject Mean	Deltas (δ) μ1 – μ2 or (*D*)	Deltas $(\delta)^2$ or $(D)^2$
Subject 1	2.057	1.43	0.63	0.40
Subject 2	2.79	1.99	0.80	0.64
Subject 3	3.90	2.07	1.83	3.34
Subject 4	3.86	3.49	0.37	0.14
Subject 5	3.60	1.86	1.74	3.04
Subject 6	2.54	1.44	1.10	1.21
Subject 7	2.49	4.26	-1.77	3.14
Subject 8	2.10	2.16	-0.06	0.003
Subject 9	1.40	1.64	-0.24	0.059
Subject 10	1.00	1.00	0.00	0
Subject 11	1.00	1.00	0.00	0
Subject 12	2.21	3.49	-1.27	1.62
Subject 13	3.80	2.20	1.60	2.56
Subject 14	3.51	2.07	1.44	2.08
Subject 15	1.91	1.04	0.90	0.81
Subject 16	2.56	1.53	1.03	1.06
Subject 17	1.20	2.21	-1.01	1.03
Subject 18	2.63	1.14	1.49	2.21
Subject 19	4.06	1.11	2.94	8.66
Subject 20	5.69	1.71	3.97	15.77
Sum (Σ)	54.30	38.81	15.49	47.76
Mean (Σ/20)	2.72	1.94	0.77	

Table 7. Study Results, Deltas (δ) or (*D*), Sums, Means, and D^2

With the mean of the subject mean differences known, as well as the number of study participants, the standard deviation is calculated, see Equation 2. Similar to the calculation of standard deviation, a standard error of the differences between the means ($\overline{s_D}$) is needed to calculate the final value of *t.* The standard deviation ($s_D = 1.37$)

known, the standard error of the differences between the means ($\overline{s_D}$) is found with the formula; calculations are seen in Equation 3 as presented by Urdan (2010).

(2)

$$s_D = \sqrt{\frac{\sum D^{2-} \frac{(\sum D)^2}{N}}{N-1}} \quad s_D = \sqrt{\frac{47.75 - \frac{(15.49)^2}{20}}{20-1}} \quad s_D = \sqrt{\frac{47.75 - \frac{239.94}{20}}{19}}$$

$$s_D = \sqrt{\frac{47.75 - 12}{19}} \quad s_D = \sqrt{\frac{35.75}{19}} \quad s_D = \sqrt{1.88} \quad s_D = 1.37$$

(3)

$$\overline{s_D} = \frac{s_D}{\sqrt{N}} \quad \overline{s_D} = \frac{1.37}{\sqrt{20}} \quad \overline{s_D} = \frac{1.37}{4.47} \quad \overline{s_D} = 0.306$$

With the standard error of the differences between the means ($\overline{s_D}$ = .306), the mean value of the questionnaire results for The Sleuth Kit ($\overline{\mu 1}$ = 2.715), and the mean value of the questionnaire results for the SequoiaView ($\overline{\mu 2}$ = 1.94) known, the value of *t* can be calculated, see Equation 4. The paired-sample *t*-test was conducted to compare the relative cognitive workloads of digital forensics first-responders operating, first, a textual-based interface and then a visualized prototyped interface. Urdan (2010) stated, "the difference between the means divided by the standard error of the difference between the means produce the *t* value" (p. 101).

(4)

$$t = \frac{\overline{\mu 1} - \overline{\mu 2}}{\overline{s_D}} \quad t = \frac{2.72 - 1.94}{.306} \quad t = \frac{2.72 - 1.94}{.306} \quad t = \frac{.78}{0.306}$$

$$t = 2.55$$

The observed value of *t* (*t* = 2.55) falls within the region of rejecting the null-hypothesis (H_0): above the critical value of the (*t*) distribution (t_0 = 1.729). Significant difference in the conditions is indicated by this study's rejection of the null hypothesis of no difference between the first-responders' cognitive load operating textual-based and visualized digital forensics interface on study results bell curve (*t* = 0), see Figure 9.

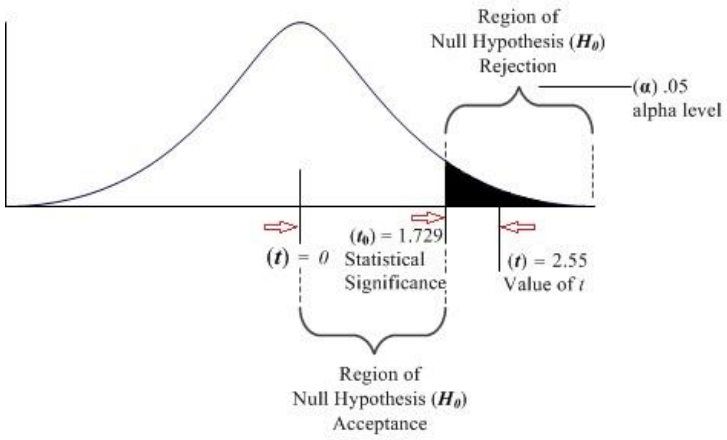


Figure 9. Study Results Bell Curve (Urdan, 2010)

Summary of Results

This research shows through empirical discovery that the visualized interface display implemented in the digital forensics process significantly reduced the cognitive workload impact compared to the workload imposed by the textual-based interface. These findings

support the hypothesis (*H*) and do not support the null hypothesis (H_0) that there would not be a significant change in the analysts' cognitive load.

The study results focus upon five file-centric tasks, see Table 2, identified during ethnographic interviews with digital forensics SMEs. Instructions to perform each task were given to 20 first-responders testing the visualized and textual-based interfaces and comparing the cognitive workload experienced while performing the tasks. Following each task, study participants responded to each of the 14 questions about interface design and also self-rated the cognitive workload of each task.

A *t*-test was conducted, testing the distribution between two dependent means, in this case the cognitive load results when comparing a visualized and a textual-based interface. Calculations of the 2800 data points recorded by the researcher as results of the questionnaire tallies produced the mean values for this study. The *t*-test analysis strategy reveals the difference between the two means, including an alpha level ($\alpha = .05$) to ensure significant contrasts between the applications under test. Additionally, to ensure test validity, a sample size of 20 was used to be consistent with Cohen's (1988) "effective size" to show a sufficient sample for the alpha level (α) set for the research. Microsoft Excel was used to calculate and present the findings graphically in this chapter and to store the study test results for analysis throughout the study.

From the sample size of 20, the degrees of freedom (*df = 19)* are obtained. The correlation between degrees of freedom (*df = 19)* and alpha level ($\alpha = .05$) enables the critical value of the *t* distribution or statistical significance ($t_0 = 1.729$) to be derived by Urdan's (2010) methodology. Utilizing the mean differences (δ) or (D), the sum of the differences squared $(\Sigma D)^2$, the squared sum of the differences (ΣD^2), and the degrees of

freedom (*df = 19)*, the standard deviation ($s_D = 1.37$) is calculated. With the degrees of freedom (*df = 19)*, the standard deviation ($s_D = 1.37$), and the standard error of the differences between the means ($\overline{s_D}$ = .306) is obtained. Finally, the calculated standard error of the differences between the means ($\overline{s_D}$ = .306), the sum of the mean questionnaire value for The Sleuth Kit ($\overline{\mu 1}$ = 2.715), and the sum of the mean of the questionnaire results for the SequoiaView ($\overline{\mu 2}$ = 1.94), *t* is presented as (*t* = 2.55). Since the value (*t* = 2.55) is greater than the statistical significance (t_0 = 1.729), the null hypothesis (H_0) is not supported and the hypothesis (*H*) is supported with a significant change, demonstrating decreased cognitive load of analysts utilizing the visualized digital forensics interface (Urdan, 2010).

Chapter 5

Conclusions, Implications, Recommendations, and Summary

Conclusions

Currently, digital forensic tools utilize a textual-based interface such as The Sleuth Kit. Even though commercial vendors (such as FTK) have integrated visualization into their product lines, the domain of digital forensics lacks empirical evidence to support the claim that visualization reduces the cognitive workload of the analyst when operating a digital forensics interface utilizing a visualized component. The hypothesis (*H*) of this work asserts that the visualized interface display integrated into the digital forensics process significantly reduces the analyst's cognitive workload compared to that required by the textual-based interface. Presented here is a potential solution to the problem of analyst overload: use of a visualized interface to reduce the cognitive workload of the analyst during the evidence-gathering process of the digital forensics investigation.

In this study three digital forensics SMEs identified five file-centric tasks described as commonly being performed during a digital forensics investigation, enabling the researcher to develop relevant scenarios for this study, see Table 2. The tasks were presented to qualified digital forensics first-responders for analysis—with each first-responder testing both a visualized and textual-based interface—to determine the cognitive workload required for the investigative outcomes to be achieved for the

researcher's examination. The first-responder study subjects followed the order of testing guided by the study instructions in Appendix F.

Results obtained by use of the visualized digital forensics tool—contrasted with those of the textual-based interface—reveal that the latter (Sleuth Kit) was outperformed by the prototype (SequoiaView) application's visualized presentation. The visualized display performed consistently to fulfill the predetermined tasks with a mental workload on the first-responder lower than that imposed by the textual-based display. Questionnaires based on basic visual design principles captured cognitive load data after each common task scenario was performed for a quantitative assessment of results. Compilation of the first-responder data indicates reduced workload on the analyst.

Research Questions

Eisenhardt (1989) described research as an explicit forward-moving process. Identified by the researcher in Chapter 1 are three research questions addressing the specific research objective explicitly defined in this dissertation:

1. What are the investigator's primary tasks for evidence identification while operating a traditional digital forensics tool set?
2. What is the cognitive load of a human's working memory beyond just time and accuracy measures while performing ethnographically discovered techniques of a predefined set of tasks to establish a baseline of evidence identification?
3. What efficiency level may be attained for a digital investigation improved for the benefit of analysis by application of visualization to predefined tasks?

Through ethnographic investigation research, Question 1 was answered with five primary file-centric tasks, see Table 2, for evidence identification by analysts operating a

digital forensics tool set. These five tasks were performed by the first-responder experimental group in testing the visualized and the textual-based tools to develop a mean-based comparison. The researcher detected a significant difference between sets of results pertaining to the study participants' cognitive workload while simulating the analysis phase of an investigation.

In order to answer Question 2 of the study, the cognitive load of the study's first-responder participants was measured by capturing an empirical assessment of their human working memory through self-evaluation. These participants performed a baseline of predefined, evidence-identification tasks established through ethnographically discovered techniques of the SME study. The identical baseline tasks were performed on both a visualized and a textual-based application, concluding with intent to improve the effectiveness of a digital forensics investigation.

The final research question was satisfied by the researcher's identifying the efficiency level attained for a digital investigation improved by benefit of visualization having been applied to an interface while first-responders performed predefined tasks. Question 3's answer derived from comparing/contrasting the results of the user surveys: analysis of the compiled results revealed a difference considered to be statistically significant—referencing the 95% confidence level of .05 value or greater level of alpha (*a*)—established in Chapter 3.

Alternatives

Many alternatives were considered for each step of this study that may have led to additional explanations of findings and additional areas for research, including the research methods for the dissertation itself, digital forensics tools and prototypes, and

statistical measures for result analysis. Initial research methods such as a purely ethnographic study were considered. Fundamentally, the ethnographic study would have worked. However, controlling the number of participants and tools would have been difficult. Also considered was the programming of a prototype tool from scratch. An originally programmed tool would have worked although the availability of the open source tools satisfied the tool requirement. The prospect of statistical analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) were considered but given up for the more appropriate *t*-test, which is most frequently used by behavioral scientists in measuring mean values, according to Urdan (2010).

Strengths

The primary strengths of this work stem from sound academic practices and policies as described in the Nova Southeastern Graduate School of Computer and Information Sciences *Dissertation Guide*. Key to this dissertation is an explicitly stated research and methodical process defining criteria for research questions, data capture and collection methods, and final project conclusion. Eisenhardt (1989) described research as an iterative process, taking into account previous knowledge, in the form of a detailed literature review. Hence, both seminal and contemporary materials from peer-reviewed sources were referenced, and the researcher ensured that identical testing was performed on both types of interfaces and that all tasking related to the same simulated dataset to provide an unbiased environment for the study.

With the study's taking place in the Washington DC area, qualified digital forensics professionals and first-responders were available. The metropolitan area, having a large

technologically sophisticated population base, provided well-educated, experienced, and certified professionals to participate in the study.

The ethnographic portion of the study provided an opportunity to define the human group of digital forensics SMEs. Not only is their expertise in the area of investigative processes noted but also the types of criminal activities they have investigated are identified as unrelated discovery and the history of their personal career paths was learned. The comparison of means in testing the visualized and textual-based interfaces with the paired variables of study participants, data, and tasks lent itself well to the dependent sample *t*-tests, advocated by Urdan (2010).

Weaknesses

Weaknesses of this study may have resulted from both internal and external dependencies. The internal dependencies include, for example, sample size, sample demographics, and formats for presenting multidimensional data visually in a two-dimensional display. According to Cohen (1988) a larger sample size will always increase the viability and accuracy of a study due to a smaller error rate or a chance of falsely rejecting the null hypothesis (H_0). Additionally, demographics were not collected for this study, such as geographic location, gender, age, and years of experience. The researcher did find that the most senior study subjects were very comfortable in either the textual-based or the visualized environments. The researcher observed this phenomenon, and it may be seen in the study results where the subjects responded with 1's or "low mental effort" for both interfaces for all tasks and questions. Lastly, only node and treemap visualized presentations were discussed in this study. Tufte (1990) discussed

graphically presenting multidimensional, quantitative data in a two-dimensional environment in many varying formats for multiple domains of study, such as chemistry and astronomy.

External dependencies that may have impacted outcomes include economic considerations and the single-researcher point of view. Due to resource constraints, this research utilized only open-source tools for testing. Digital forensics tools are commercially available, but not available for trial licensing or educational use. Key to this research has been thoughtful input from the dissertation committee. However, for obvious reasons, the dissertation process restricts input to the study content from other researchers, limiting the study development process to a narrow view.

Limitations

There are several limitations to this study. For instance, data used for subject testing were simulated, the testing itself was limited in scope, and, again, open-source tools were used. Simulated perpetrator case data used for testing were not actual crime scene data. The simulated data were used to enforce a chain-of-custody element for actual case data, as mentioned in Chapter 1, protecting both a potential victim and perpetrator. The scope in testing for the study was limited by the participants in completing individual tasks. Measures for task performance, such as time and accuracy, were not developed or used. Because tasks utilized for establishing the cognitive workload built upon the previous tasks, they were conducted in the same order for each participant. Additionally, the applications tested were not alternated but were utilized in the same order by all

participants. Though the researcher is thankful for their availability, the tools used lacked some obviously desirable features, such as a back button for application navigation.

Implications

With digital evidence playing a constant role in criminal investigations, according to Regional Computer Forensics Laboratories (RCFL) (2013), this study is both timely and relevant. This work complements the previous work of others while strengthening the body of knowledge of digital forensics and human–computer interaction (HCI).

To date, no published empirical documentation accounts for how visualization impacts the cognitive workload of the digital forensics practitioner in the investigative analysis process. The results of this study reach across many domains where digital forensics tools are utilized and provide solid footing for enhancement of these tools. Additional relevance of this work is that it exemplifies a clear analysis method for future research by the HCI researcher, providing a template for comparing and contrasting application interface design techniques.

Recommendations

This study provides five real-world tests for examining and measuring outcomes in the testing or prototyping of digital forensics tools. Its findings are applicable for the betterment of the digital forensics analyst's efficiencies and, in turn, better protect the populace. Future work stemming from this research might benefit from the following recommendations:

1. Determine the most efficient way to present file-centric data visually to the digital forensics practitioner.

2. Improve the digital forensics tool for optimum demographic analysis.
3. Incorporate findings from this work and those of future projects into open-source and commercial digital forensics tool developers' tool sets.
4. Acquire funding for testing commercial tool sets.

Summary

Digital forensics tools are software and other instruments that assist analysts in the discovery of digital evidence located in an array of computer systems. According to the RCFL 2012 FY report, nearly all current criminal investigations involve such a device. This research builds upon previous works, such as that of Garfinkel (2010), to demonstrate enhancement in the discovery phase of the evidence-identification process through the use of visualization. To date, understanding of the process is so limited as to preclude clear identification of the impact of visualization integrated into the user interface of the digital forensics tool or upon the digital investigation itself. One reason for this dearth of understanding is that the domain of digital forensics remains in its infancy stage. During an investigation the analyst may need to assess thousands, millions, or even billions of files to identify digital evidence, an inordinately cognitively challenging task.

It is expected that a significant degree of understanding of the impact of visualization on the digital forensics tool may be obtained through analysis and discovery of evidence read digitally rather than merely cognitively. Additionally expected is heightened understanding of the process whereby visualization reduces analysts' cognitive workload, thereby proving digital evidence detection more efficient.

A considerable amount of research details digital forensics tools and processes. Additionally, several researchers have called for the improvement of digital forensics tools through the use of visualization (Hoelz & Ralha, 2013). Visualization design principles, when applied to the software interface, offer capabilities to present very large datasets to the user, bringing resolution to multi-dimensional information (Shneiderman, 2008; Tufte, 1990).

The hypothesis (*H*) in this research is that the cognitive load of the practitioner will be reduced by incorporating visualization into the digital forensics application interface. Huang, Eades, and Hong (2009) validated the premise that cognitive load is consistently reduced in applications into which visualization has been integrated, proving such applications superior to textual-based interfaces. In this study ethnographic research identifies the primary duties of the digital forensics analyst during the analysis phase of investigative discovery while the cognitive load is measured by means of the user's comparison of the visualized interface with the textual-based interface.

This research contributes to the body of knowledge in its field by validating a method of measurement and by providing empirical evidence consistent with the hypothesis (*H*): use of the visualized digital forensics interface will provide a more efficient performance by the analyst, saving labor costs and compressing time required for the digital investigation discovery phase.

The purpose of this research is to provide empirical evidence to determine whether the use of visualization in the identification of digital evidence will reduce the digital forensics application user's cognitive load. However, visualization may have a negative impact in the legal environment or when presented in another public forum (John, 2012).

This research adopted the self-appraisal aspect of a system interface determined by users' observation levels of cognitive load patterns, suggested by Rose et al. (1995).

The methodology design is intended to provide results ensuring a high level of reliability. Confidence in the study is derived from appropriate sample size, subject population selection, and procedures followed for a power analysis, as suggested by Cohen (1988).

This research shows through empirical discovery that the visualized interface display implemented in the digital forensics process significantly reduces the cognitive workload compared to that reported for the textual-based interface. These findings support the hypothesis (H) and invalidate the null hypothesis (H_0): there would not be a significant change in the analyst's cognitive load.

The study results focus upon five file-centric tasks (see Table 2) identified during ethnographic interviews with digital forensics SMEs. Richard and Roussev (2006) defined file-centric activities as primary labors of the digital forensics investigation. Instructions to perform each task were given to 20 professional, qualified first-responders who tested the visualized and textual-based interfaces by comparing the cognitive workloads experienced—according to interface type—while performing the tasks. Following each task, study participants responded to a questionnaire with a total of 14 interface design self-rating questions to score respective workloads.

A *t*-test was conducted to identify the mean values of over 2800 data points recorded by the researcher as questionnaire tallies. The *t*-test analysis strategy provided the difference between two means to include an alpha (α) level of .05 to ensure a significant difference among the results of the applications under test. Additionally, to ensure test

validity, a sample size of 20 was used to be consistent with Cohen's concept (1988) of "effective size" to show a sufficient sample for the alpha level (α) set for the research. Microsoft Excel was utilized in this study for storing, calculating, and presenting graphically the test result findings.

From the sample size of 20, the degrees of freedom (*df = 19)* are obtained. The correlation between degrees of freedom (*df = 19)* and alpha level (α = .05) enables the critical value of the *t* distribution or statistical significance (t_0 = 1.729) to be derived from Urdan's (2010) formula that allows one to determine statistical significance. Utilizing the mean differences (δ) or (*D*), the calculated sum of the differences squared $(\Sigma D)^2$, the squared sum of the differences (ΣD^2), and the degrees of freedom (*df = 19),* the standard deviation ($s_D = 1.37$) is calculated. With the degrees of freedom (*df = 19)* and the standard deviation ($s_D = 1.37$), the standard error of the differences between the means ($\overline{s_D}$ = .306) is obtained. Finally, the standard error of the differences between the means ($\overline{s_D}$ = .306), the sum of the mean questionnaire value for The Sleuth Kit ($\overline{\mu 1}$ = 2.715), and the sum of the mean value of the questionnaire results for the SequoiaView ($\overline{\mu 2}$ = 1.94) produce a *t* of (*t* = 2.55). Since the value (*t* = 2.55) is greater than the statistical significance (t_0 = 1.729), the null hypothesis (H_0) is invalidated and the hypothesis (*H*) is supported with a significant change, demonstrating the analysts' cognitive load being decreased while utilizing the visualized digital forensics interface.

This study provides five real-world tasks for examining and measuring outcomes in the testing or prototyping of digital forensics tools, see Table 2. Its findings are applicable for the betterment of digital forensics analysts' efficiencies and, in turn, better protection of the populace. Future researchers may benefit from this research by

continuing to determine efficient ways to present file-centric data visually to the digital forensics practitioner. Additionally, developers of open-source and commercial digital forensics tools should incorporate findings from this work and those of future projects into their tool sets to improve the digital forensics tool to be optimized for the analysts' distinguishing personal demographics. Finally, future research would benefit from the inclusion of commercial tool sets with acquired grants or funding.

Appendix A

Raw Data

Data and Mean for Subjects by Question and Interface

Subject	Interface	Question	Task 1	Task 2	Task 3	Task 4	Task 5	Mean
1	SequoiaView	1	1	1	1	1	1	1
1	SequoiaView	2	1	1	1	1	1	1
1	SequoiaView	3	2	3	1	1	1	1.6
1	SequoiaView	4	1	1	1	1	1	1
1	SequoiaView	5	2	2	2	2	2	2
1	SequoiaView	6	2	2	2	2	2	2
1	SequoiaView	7	2	2	1	2	2	1.8
1	SequoiaView	8	1	1	1	1	1	1
1	SequoiaView	9	1	1	1	1	1	1
1	SequoiaView	10	2	1	2	2	2	1.8
1	SequoiaView	11	1	2	2	2	2	1.8
1	SequoiaView	12	1	2	2	2	2	1.8
1	SequoiaView	13	1	2	1	1	1	1.2
1	SequoiaView	14	1	1	1	1	1	1
1	SequoiaView	Mean	1.36	1.58	1.36	1.43	1.43	1.43
1	Sleuth Kit	1	1	1	1	1	1	1
1	Sleuth Kit	2	1	3	1	1	1	1.4
1	Sleuth Kit	3	3	4	3	3	4	3.4
1	Sleuth Kit	4	1	1	1	1	1	1
1	Sleuth Kit	5	1	1	1	1	1	1
1	Sleuth Kit	6	2	2	2	2	2	2
1	Sleuth Kit	7	2	2	2	2	2	2
1	Sleuth Kit	8	5	7	7	7	7	6.6
1	Sleuth Kit	9	1	1	1	1	1	1
1	Sleuth Kit	10	1	1	1	1	1	1
1	Sleuth Kit	11	2	2	1	1	1	1.4
1	Sleuth Kit	12	1	2	2	2	2	1.8
1	Sleuth Kit	13	1	1	1	2	1	1.2
1	Sleuth Kit	14	4	4	3	5	4	4
1	Sleuth Kit	Mean	1.86	2.28	1.93	2.14	2.07	2.06
2	SequoiaView	1	1	1	2	1	1	1.2
2	SequoiaView	2	1	1	2	1	2	1.4
2	SequoiaView	3	1	1	1	1	1	1
2	SequoiaView	4	1	1	2	1	1	1.2

2	SequoiaView	5	4	2	2	2	2	2.4
2	SequoiaView	6	3	2	7	7	7	5.2
2	SequoiaView	7	2	4	3	1	1	2.2
2	SequoiaView	8	1	1	2	2	2	1.6
2	SequoiaView	9	2	1	2	2	2	1.8
2	SequoiaView	10	4	1	2	2	1	2
2	SequoiaView	11	3	1	2	2	2	2
2	SequoiaView	12	5	3	2	3	2	3
2	SequoiaView	13	1	2	1	2	1	1.4
2	SequoiaView	14	1	2	1	2	1	1.4
2	SequoiaView	Mean	2.14	1.64	2.21	2.07	1.86	1.98
2	Sleuth Kit	1	3	1	3	2	1	2
2	Sleuth Kit	2	4	2	2	3	2	2.6
2	Sleuth Kit	3	2	2	1	3	1	1.8
2	Sleuth Kit	4	2	2	2	2	1	1.8
2	Sleuth Kit	5	2	3	4	4	2	3
2	Sleuth Kit	6	5	6	6	6	5	5.6
2	Sleuth Kit	7	3	3	3	3	3	3
2	Sleuth Kit	8	5	3	2	2	5	3.4
2	Sleuth Kit	9	2	3	3	1	1	2
2	Sleuth Kit	10	4	4	3	2	3	3.2
2	Sleuth Kit	11	3	5	3	2	4	3.4
2	Sleuth Kit	12	3	3	4	2	3	3
2	Sleuth Kit	13	2	2	2	1	1	1.6
2	Sleuth Kit	14	4	3	2	2	2	2.6
2	Sleuth Kit	Mean	3.14	3	2.86	2.5	2.43	2.78
3	SequoiaView	1	2	1	1	1	2	1.4
3	SequoiaView	2	1	1	1	1	1	1
3	SequoiaView	3	2	2	3	2	2	2.2
3	SequoiaView	4	3	2	2	1	1	1.8
3	SequoiaView	5	2	3	2	2	2	2.2
3	SequoiaView	6	4	3	4	4	3	3.6
3	SequoiaView	7	2	3	3	2	3	2.6
3	SequoiaView	8	1	1	1	1	1	1
3	SequoiaView	9	4	3	3	3	2	3
3	SequoiaView	10	1	1	2	1	1	1.2
3	SequoiaView	11	2	2	2	2	2	2
3	SequoiaView	12	3	2	2	2	2	2.2
3	SequoiaView	13	3	3	3	3	2	2.8
3	SequoiaView	14	2	2	3	1	2	2
3	SequoiaView	Mean	2.29	2.07	2.29	1.86	1.86	2.07
3	Sleuth Kit	1	5	5	4	6	5	5
3	Sleuth Kit	2	4	4	2	4	4	3.6
3	Sleuth Kit	3	6	4	5	5	6	5.2
3	Sleuth Kit	4	2	4	3	2	3	2.8
3	Sleuth Kit	5	2	3	4	4	4	3.4
3	Sleuth Kit	6	4	4	4	4	4	4
3	Sleuth Kit	7	4	4	3	3	3	3.4
3	Sleuth Kit	8	6	6	6	3	6	5.4
3	Sleuth Kit	9	3	5	5	4	4	4.2
3	Sleuth Kit	10	2	3	2	3	5	3

3	Sleuth Kit	11	3	3	3	3	4	3.2
3	Sleuth Kit	12	4	2	4	2	4	3.2
3	Sleuth Kit	13	5	3	3	3	2	3.2
3	Sleuth Kit	14	5	5	6	4	5	5
3	Sleuth Kit	Mean	3.93	3.93	3.86	3.57	4.21	3.9
4	SequoiaView	1	2	3	6	4	4	3.8
4	SequoiaView	2	2	3	2	4	4	3
4	SequoiaView	3	2	3	4	2	4	3
4	SequoiaView	4	5	2	4	5	4	4
4	SequoiaView	5	3	3	6	3	4	3.8
4	SequoiaView	6	3	4	2	2	4	3
4	SequoiaView	7	4	4	4	4	4	4
4	SequoiaView	8	4	3	4	5	4	4
4	SequoiaView	9	2	2	2	2	2	2
4	SequoiaView	10	4	5	5	2	4	4
4	SequoiaView	11	5	5	5	4	4	4.6
4	SequoiaView	12	4	2	2	2	2	2.4
4	SequoiaView	13	2	2	2	4	3	2.6
4	SequoiaView	14	5	4	5	5	4	4.6
4	SequoiaView	Mean	3.36	3.21	3.79	3.43	3.64	3.49
4	Sleuth Kit	1	4	5	4	4	4	4.2
4	Sleuth Kit	2	4	4	4	4	4	4
4	Sleuth Kit	3	4	4	4	3	3	3.6
4	Sleuth Kit	4	4	4	3	3	4	3.6
4	Sleuth Kit	5	5	6	4	4	4	4.6
4	Sleuth Kit	6	4	4	4	4	4	4
4	Sleuth Kit	7	4	4	4	4	4	4
4	Sleuth Kit	8	6	6	6	6	6	6
4	Sleuth Kit	9	2	2	2	2	2	2
4	Sleuth Kit	10	4	3	4	3	4	3.6
4	Sleuth Kit	11	5	4	4	5	5	4.6
4	Sleuth Kit	12	2	2	2	2	2	2
4	Sleuth Kit	13	4	4	4	3	4	3.8
4	Sleuth Kit	14	4	4	4	4	4	4
4	Sleuth Kit	Mean	4	4	3.79	3.64	3.86	3.86
5	SequoiaView	1	1	1	1	1	1	1
5	SequoiaView	2	1	1	1	1	1	1
5	SequoiaView	3	5	1	1	1	1	1.8
5	SequoiaView	4	7	1	1	1	1	2.2
5	SequoiaView	5	1	1	1	1	1	1
5	SequoiaView	6	5	5	5	5	5	5
5	SequoiaView	7	1	1	1	1	1	1
5	SequoiaView	8	1	1	1	1	1	1
5	SequoiaView	9	1	1	1	1	1	1
5	SequoiaView	10	1	1	1	1	1	1
5	SequoiaView	11	1	1	1	1	1	1
5	SequoiaView	12	1	1	1	1	1	1
5	SequoiaView	13	7	7	7	7	7	7
5	SequoiaView	14	1	1	1	1	1	1
5	SequoiaView	Mean	2.43	1.71	1.71	1.71	1.71	1.86
5	Sleuth Kit	1	6	6	6	6	6	6

5	Sleuth Kit	2	3	3	6	4	4	4
5	Sleuth Kit	3	1	1	1	1	1	1
5	Sleuth Kit	4	4	5	4	4	4	4.2
5	Sleuth Kit	5	5	5	5	5	5	5
5	Sleuth Kit	6	5	5	6	6	6	5.6
5	Sleuth Kit	7	1	1	1	1	1	1
5	Sleuth Kit	8	7	7	7	7	7	7
5	Sleuth Kit	9	1	1	1	1	1	1
5	Sleuth Kit	10	3	3	3	3	3	3
5	Sleuth Kit	11	1	1	1	1	1	1
5	Sleuth Kit	12	1	1	1	1	1	1
5	Sleuth Kit	13	7	7	7	7	7	7
5	Sleuth Kit	14	4	4	4	3	3	3.6
5	Sleuth Kit	Mean	3.5	3.57	3.79	3.57	3.57	3.6
6	SequoiaView	1	1	1	1	1	1	1
6	SequoiaView	2	1	1	1	1	1	1
6	SequoiaView	3	1	1	1	1	1	1
6	SequoiaView	4	1	1	1	1	1	1
6	SequoiaView	5	2	2	2	2	2	2
6	SequoiaView	6	3	3	3	3	3	3
6	SequoiaView	7	1	1	1	1	1	1
6	SequoiaView	8	1	1	1	1	1	1
6	SequoiaView	9	1	2	2	2	2	1.8
6	SequoiaView	10	2	2	1	1	1	1.4
6	SequoiaView	11	1	1	2	2	2	1.6
6	SequoiaView	12	2	1	1	1	1	1.2
6	SequoiaView	13	3	2	2	2	2	2.2
6	SequoiaView	14	1	1	1	1	1	1
6	SequoiaView	Mean	1.5	1.43	1.43	1.43	1.43	1.44
6	Sleuth Kit	1	1	2	1	1	1	1.2
6	Sleuth Kit	2	1	1	1	1	1	1
6	Sleuth Kit	3	3	2	1	2	1	1.8
6	Sleuth Kit	4	3	1	1	1	1	1.4
6	Sleuth Kit	5	4	2	6	3	6	4.2
6	Sleuth Kit	6	1	2	2	2	2	1.8
6	Sleuth Kit	7	3	1	1	1	1	1.4
6	Sleuth Kit	8	6	6	6	6	6	6
6	Sleuth Kit	9	6	3	3	3	4	3.8
6	Sleuth Kit	10	1	1	2	1	1	1.2
6	Sleuth Kit	11	6	6	6	3	4	5
6	Sleuth Kit	12	1	1	1	1	1	1
6	Sleuth Kit	13	3	3	2	3	3	2.8
6	Sleuth Kit	14	3	3	3	3	3	3
6	Sleuth Kit	Mean	3	2.43	2.57	2.21	2.5	2.54
7	SequoiaView	1	4	4	5	4	6	4.6
7	SequoiaView	2	6	4	2	4	3	3.8
7	SequoiaView	3	2	2	2	5	3	2.8
7	SequoiaView	4	2	2	2	5	2	2.6
7	SequoiaView	5	6	6	6	3	6	5.4
7	SequoiaView	6	7	7	7	7	6	6.8
7	SequoiaView	7	5	5	2	2	2	3.2

7	SequoiaView	8	1	2	1	2	2	1.6
7	SequoiaView	9	3	4	6	6	6	5
7	SequoiaView	10	6	6	6	2	3	4.6
7	SequoiaView	11	6	6	6	6	6	6
7	SequoiaView	12	6	5	6	6	6	5.8
7	SequoiaView	13	6	6	6	6	6	6
7	SequoiaView	14	1	2	1	2	1	1.4
7	SequoiaView	Mean	4.36	4.36	4.14	4.29	4.14	4.26
7	Sleuth Kit	1	3	2	3	2	2	2.4
7	Sleuth Kit	2	1	1	2	1	1	1.2
7	Sleuth Kit	3	5	1	1	3	3	2.6
7	Sleuth Kit	4	1	2	1	1	1	1.2
7	Sleuth Kit	5	1	4	2	1	1	1.8
7	Sleuth Kit	6	3	2	4	2	2	2.6
7	Sleuth Kit	7	4	4	2	2	2	2.8
7	Sleuth Kit	8	6	6	6	6	6	6
7	Sleuth Kit	9	1	1	2	1	1	1.2
7	Sleuth Kit	10	2	1	1	1	1	1.2
7	Sleuth Kit	11	2	1	3	1	1	1.6
7	Sleuth Kit	12	4	4	3	3	3	3.4
7	Sleuth Kit	13	5	6	6	6	6	5.8
7	Sleuth Kit	14	1	1	1	1	1	1
7	Sleuth Kit	Mean	2.79	2.57	2.64	2.21	2.21	2.49
8	SequoiaView	1	1	1	1	1	2	1.2
8	SequoiaView	2	2	2	2	2	2	2
8	SequoiaView	3	1	1	1	1	1	1
8	SequoiaView	4	1	1	2	1	2	1.4
8	SequoiaView	5	3	3	2	2	1	2.2
8	SequoiaView	6	4	4	3	3	3	3.4
8	SequoiaView	7	3	3	3	2	3	2.8
8	SequoiaView	8	2	2	2	2	2	2
8	SequoiaView	9	3	3	3	3	3	3
8	SequoiaView	10	4	3	3	3	3	3.2
8	SequoiaView	11	4	3	2	2	2	2.6
8	SequoiaView	12	4	2	3	2	3	2.8
8	SequoiaView	13	2	1	1	1	1	1.2
8	SequoiaView	14	2	2	1	1	1	1.4
8	SequoiaView	Mean	2.57	2.21	2.07	1.86	2.07	2.16
8	Sleuth Kit	1	3	3	2	2	2	2.4
8	Sleuth Kit	2	4	3	2	2	2	2.6
8	Sleuth Kit	3	4	1	1	1	1	1.6
8	Sleuth Kit	4	2	5	2	1	2	2.4
8	Sleuth Kit	5	2	2	2	3	2	2.2
8	Sleuth Kit	6	4	3	3	3	3	3.2
8	Sleuth Kit	7	2	3	2	3	3	2.6
8	Sleuth Kit	8	3	3	2	2	3	2.6
8	Sleuth Kit	9	3	2	2	3	3	2.6
8	Sleuth Kit	10	1	1	2	2	2	1.6
8	Sleuth Kit	11	1	2	2	2	2	1.8
8	Sleuth Kit	12	2	1	1	2	1	1.4
8	Sleuth Kit	13	2	1	1	1	1	1.2

8	Sleuth Kit	14	2	1	1	1	1	1.2
8	Sleuth Kit	Mean	2.5	2.21	1.79	2	2	2.1
9	SequoiaView	1	1	1	1	1	1	1
9	SequoiaView	2	1	1	1	1	1	1
9	SequoiaView	3	2	1	1	1	1	1.2
9	SequoiaView	4	1	1	1	1	1	1
9	SequoiaView	5	4	2	1	1	1	1.8
9	SequoiaView	6	2	1	1	1	1	1.2
9	SequoiaView	7	4	4	2	1	1	2.4
9	SequoiaView	8	4	4	1	4	4	3.4
9	SequoiaView	9	1	2	1	2	1	1.4
9	SequoiaView	10	1	1	1	1	4	1.6
9	SequoiaView	11	1	1	2	1	4	1.8
9	SequoiaView	12	3	4	4	4	1	3.2
9	SequoiaView	13	1	1	1	1	1	1
9	SequoiaView	14	1	1	1	1	1	1
9	SequoiaView	Mean	1.93	1.79	1.36	1.5	1.64	1.64
9	Sleuth Kit	1	1	1	1	1	1	1
9	Sleuth Kit	2	1	1	1	2	1	1.2
9	Sleuth Kit	3	1	1	1	1	1	1
9	Sleuth Kit	4	1	1	1	1	2	1.2
9	Sleuth Kit	5	2	1	1	1	2	1.4
9	Sleuth Kit	6	1	1	1	1	2	1.2
9	Sleuth Kit	7	2	2	1	2	1	1.6
9	Sleuth Kit	8	4	2	2	4	4	3.2
9	Sleuth Kit	9	1	2	2	1	1	1.4
9	Sleuth Kit	10	2	1	1	2	2	1.6
9	Sleuth Kit	11	1	2	1	1	1	1.2
9	Sleuth Kit	12	4	1	1	1	1	1.6
9	Sleuth Kit	13	1	1	1	1	1	1
9	Sleuth Kit	14	1	1	1	1	1	1
9	Sleuth Kit	Mean	1.64	1.29	1.14	1.43	1.5	1.4
10	SequoiaView	1	1	1	1	1	1	1
10	SequoiaView	2	1	1	1	1	1	1
10	SequoiaView	3	1	1	1	1	1	1
10	SequoiaView	4	1	1	1	1	1	1
10	SequoiaView	5	1	1	1	1	1	1
10	SequoiaView	6	1	1	1	1	1	1
10	SequoiaView	7	1	1	1	1	1	1
10	SequoiaView	8	1	1	1	1	1	1
10	SequoiaView	9	1	1	1	1	1	1
10	SequoiaView	10	1	1	1	1	1	1
10	SequoiaView	11	1	1	1	1	1	1
10	SequoiaView	12	1	1	1	1	1	1
10	SequoiaView	13	1	1	1	1	1	1
10	SequoiaView	14	1	1	1	1	1	1
10	SequoiaView	Mean	1	1	1	1	1	1
10	Sleuth Kit	1	1	1	1	1	1	1
10	Sleuth Kit	2	1	1	1	1	1	1
10	Sleuth Kit	3	1	1	1	1	1	1
10	Sleuth Kit	4	1	1	1	1	1	1

10	Sleuth Kit	5	1	1	1	1	1	1
10	Sleuth Kit	6	1	1	1	1	1	1
10	Sleuth Kit	7	1	1	1	1	1	1
10	Sleuth Kit	8	1	1	1	1	1	1
10	Sleuth Kit	9	1	1	1	1	1	1
10	Sleuth Kit	10	1	1	1	1	1	1
10	Sleuth Kit	11	1	1	1	1	1	1
10	Sleuth Kit	12	1	1	1	1	1	1
10	Sleuth Kit	13	1	1	1	1	1	1
10	Sleuth Kit	14	1	1	1	1	1	1
10	Sleuth Kit	Mean	1	1	1	1	1	1
11	SequoiaView	1	1	1	1	1	1	1
11	SequoiaView	2	1	1	1	1	1	1
11	SequoiaView	3	1	1	1	1	1	1
11	SequoiaView	4	1	1	1	1	1	1
11	SequoiaView	5	1	1	1	1	1	1
11	SequoiaView	6	1	1	1	1	1	1
11	SequoiaView	7	1	1	1	1	1	1
11	SequoiaView	8	1	1	1	1	1	1
11	SequoiaView	9	1	1	1	1	1	1
11	SequoiaView	10	1	1	1	1	1	1
11	SequoiaView	11	1	1	1	1	1	1
11	SequoiaView	12	1	1	1	1	1	1
11	SequoiaView	13	1	1	1	1	1	1
11	SequoiaView	14	1	1	1	1	1	1
11	SequoiaView	Mean	1	1	1	1	1	1
11	Sleuth Kit	1	1	1	1	1	1	1
11	Sleuth Kit	2	1	1	1	1	1	1
11	Sleuth Kit	3	1	1	1	1	1	1
11	Sleuth Kit	4	1	1	1	1	1	1
11	Sleuth Kit	5	1	1	1	1	1	1
11	Sleuth Kit	6	1	1	1	1	1	1
11	Sleuth Kit	7	1	1	1	1	1	1
11	Sleuth Kit	8	1	1	1	1	1	1
11	Sleuth Kit	9	1	1	1	1	1	1
11	Sleuth Kit	10	1	1	1	1	1	1
11	Sleuth Kit	11	1	1	1	1	1	1
11	Sleuth Kit	12	1	1	1	1	1	1
11	Sleuth Kit	13	1	1	1	1	1	1
11	Sleuth Kit	14	1	1	1	1	1	1
11	Sleuth Kit	Mean	1	1	1	1	1	1
12	SequoiaView	1	7	1	7	7	7	5.8
12	SequoiaView	2	6	1	1	1	2	2.2
12	SequoiaView	3	1	1	1	1	1	1
12	SequoiaView	4	2	1	1	3	1	1.6
12	SequoiaView	5	7	6	1	3	3	4
12	SequoiaView	6	7	7	7	7	7	7
12	SequoiaView	7	7	7	6	2	3	5
12	SequoiaView	8	1	7	7	4	3	4.4
12	SequoiaView	9	7	7	7	7	7	7
12	SequoiaView	10	7	2	3	3	4	3.8

12	SequoiaView	11	7	2	3	4	4	4
12	SequoiaView	12	7	7	3	4	3	4.8
12	SequoiaView	13	1	1	1	1	1	1
12	SequoiaView	14	3	1	3	3	3	2.6
12	SequoiaView	Mean	5	3.64	3.64	3.57	3.5	3.87
12	Sleuth Kit	1	1	1	1	1	1	1
12	Sleuth Kit	2	1	1	1	1	1	1
12	Sleuth Kit	3	2	1	1	1	1	1.2
12	Sleuth Kit	4	1	1	1	1	1	1
12	Sleuth Kit	5	1	1	1	1	1	1
12	Sleuth Kit	6	1	1	1	1	1	1
12	Sleuth Kit	7	4	3	4	2	2	3
12	Sleuth Kit	8	7	7	7	7	7	7
12	Sleuth Kit	9	1	4	4	4	4	3.4
12	Sleuth Kit	10	1	3	2	2	2	2
12	Sleuth Kit	11	1	3	1	2	2	1.8
12	Sleuth Kit	12	4	3	4	2	2	3
12	Sleuth Kit	13	4	4	2	1	1	2.4
12	Sleuth Kit	14	1	3	3	2	2	2.2
12	Sleuth Kit	Mean	2.14	2.57	2.3	2	2	2.21
13	SequoiaView	1	2	2	2	3	2	2.2
13	SequoiaView	2	2	2	2	3	2	2.2
13	SequoiaView	3	2	2	2	3	2	2.2
13	SequoiaView	4	2	2	2	3	2	2.2
13	SequoiaView	5	2	2	2	3	2	2.2
13	SequoiaView	6	2	2	2	3	2	2.2
13	SequoiaView	7	2	2	2	3	2	2.2
13	SequoiaView	8	2	2	2	3	2	2.2
13	SequoiaView	9	2	2	2	3	2	2.2
13	SequoiaView	10	2	2	2	3	2	2.2
13	SequoiaView	11	2	2	2	3	2	2.2
13	SequoiaView	12	2	2	2	3	2	2.2
13	SequoiaView	13	2	2	2	3	2	2.2
13	SequoiaView	14	2	2	2	3	2	2.2
13	SequoiaView	Mean	2	2	2	3	2	2.2
13	Sleuth Kit	1	4	3	4	4	4	3.8
13	Sleuth Kit	2	4	3	4	4	4	3.8
13	Sleuth Kit	3	4	3	4	4	4	3.8
13	Sleuth Kit	4	4	3	4	4	4	3.8
13	Sleuth Kit	5	4	3	4	4	4	3.8
13	Sleuth Kit	6	4	3	4	4	4	3.8
13	Sleuth Kit	7	4	3	4	4	4	3.8
13	Sleuth Kit	8	4	3	4	4	4	3.8
13	Sleuth Kit	9	4	3	4	4	4	3.8
13	Sleuth Kit	10	4	3	4	4	4	3.8
13	Sleuth Kit	11	4	3	4	4	4	3.8
13	Sleuth Kit	12	4	3	4	4	4	3.8
13	Sleuth Kit	13	4	3	4	4	4	3.8
13	Sleuth Kit	14	4	3	4	4	4	3.8
13	Sleuth Kit	Mean	4	3	4	4	4	3.8
14	SequoiaView	1	2	2	2	2	2	2

14	SequoiaView	2	3	3	3	3	3	3
14	SequoiaView	3	3	3	3	3	3	3
14	SequoiaView	4	3	3	3	3	3	3
14	SequoiaView	5	4	3	3	3	3	3.2
14	SequoiaView	6	4	4	4	4	4	4
14	SequoiaView	7	3	3	3	3	3	3
14	SequoiaView	8	2	2	2	2	2	2
14	SequoiaView	9	4	3	3	3	3	3.2
14	SequoiaView	10	3	4	4	4	4	3.8
14	SequoiaView	11	4	4	4	4	4	4
14	SequoiaView	12	3	3	3	3	3	3
14	SequoiaView	13	2	2	2	2	2	2
14	SequoiaView	14	3	3	3	3	3	3
14	SequoiaView	Mean	3.07	3	3	3	3	3.01
14	Sleuth Kit	1	4	4	4	4	4	4
14	Sleuth Kit	2	3	3	3	3	3	3
14	Sleuth Kit	3	3	3	3	3	3	3
14	Sleuth Kit	4	3	3	3	3	3	3
14	Sleuth Kit	5	2	2	2	2	2	2
14	Sleuth Kit	6	4	4	4	4	4	4
14	Sleuth Kit	7	3	3	2	2	2	2.4
14	Sleuth Kit	8	6	6	6	6	6	6
14	Sleuth Kit	9	5	5	5	5	5	5
14	Sleuth Kit	10	4	4	4	4	4	4
14	Sleuth Kit	11	4	4	4	4	4	4
14	Sleuth Kit	12	3	3	3	3	3	3
14	Sleuth Kit	13	2	2	2	2	2	2
14	Sleuth Kit	14	4	4	4	3	4	3.8
14	Sleuth Kit	Mean	3.57	3.57	3.5	3.43	3.5	3.51
15	SequoiaView	1	1	1	1	1	1	1
15	SequoiaView	2	1	1	1	1	1	1
15	SequoiaView	3	1	1	1	1	1	1
15	SequoiaView	4	1	1	1	1	1	1
15	SequoiaView	5	1	1	1	1	1	1
15	SequoiaView	6	2	1	1	1	1	1.2
15	SequoiaView	7	1	1	1	1	1	1
15	SequoiaView	8	1	1	1	1	1	1
15	SequoiaView	9	1	1	1	1	1	1
15	SequoiaView	10	1	1	1	1	1	1
15	SequoiaView	11	1	1	1	1	1	1
15	SequoiaView	12	1	1	1	1	1	1
15	SequoiaView	13	1	1	1	1	1	1
15	SequoiaView	14	1	1	1	1	1	1
15	SequoiaView	Mean	1.07	1	1	1	1	1.01
15	Sleuth Kit	1	3	1	2	1	2	1.8
15	Sleuth Kit	2	3	3	2	2	3	2.6
15	Sleuth Kit	3	4	2	3	2	2	2.6
15	Sleuth Kit	4	1	1	3	2	2	1.8
15	Sleuth Kit	5	1	1	1	1	1	1
15	Sleuth Kit	6	1	2	1	1	1	1.2
15	Sleuth Kit	7	3	3	2	3	3	2.8

15	Sleuth Kit	8	1	4	3	2	2	2.4
15	Sleuth Kit	9	1	3	2	2	2	2
15	Sleuth Kit	10	1	1	1	1	1	1
15	Sleuth Kit	11	1	1	1	1	1	1
15	Sleuth Kit	12	3	2	3	2	2	2.4
15	Sleuth Kit	13	4	1	2	1	2	2
15	Sleuth Kit	14	3	2	2	2	2	2.2
15	Sleuth Kit	Mean	2.14	1.93	2	1.64	1.86	1.91
16	SequoiaView	1	2	1	1	1	1	1.2
16	SequoiaView	2	2	1	1	1	1	1.2
16	SequoiaView	3	2	2	1	2	3	2
16	SequoiaView	4	1	1	1	1	2	1.2
16	SequoiaView	5	2	1	1	1	2	1.4
16	SequoiaView	6	2	2	1	2	3	2
16	SequoiaView	7	2	2	1	2	4	2.2
16	SequoiaView	8	1	1	1	3	1	1.4
16	SequoiaView	9	2	1	1	3	2	1.8
16	SequoiaView	10	1	1	1	2	1	1.2
16	SequoiaView	11	1	1	1	1	1	1
16	SequoiaView	12	1	2	1	3	4	2.2
16	SequoiaView	13	1	1	1	1	1	1
16	SequoiaView	14	1	1	1	2	3	1.6
16	SequoiaView	Mean	1.5	1.29	1	1.79	2.07	1.53
16	Sleuth Kit	1	2	2	2	2	2	2
16	Sleuth Kit	2	2	2	1	3	2	2
16	Sleuth Kit	3	1	2	2	3	3	2.2
16	Sleuth Kit	4	1	1	2	3	3	2
16	Sleuth Kit	5	2	2	2	3	3	2.4
16	Sleuth Kit	6	2	3	3	4	4	3.2
16	Sleuth Kit	7	3	3	2	3	3	2.8
16	Sleuth Kit	8	5	4	4	4	3	4
16	Sleuth Kit	9	3	3	3	4	3	3.2
16	Sleuth Kit	10	3	2	2	3	2	2.4
16	Sleuth Kit	11	3	2	2	3	2	2.4
16	Sleuth Kit	12	4	2	3	4	4	3.4
16	Sleuth Kit	13	2	2	1	3	2	2
16	Sleuth Kit	14	2	2	2	2	1	1.8
16	Sleuth Kit	Mean	2.5	2.29	2.21	3.14	2.64	2.56
17	SequoiaView	1	4	1	6	2	4	3.4
17	SequoiaView	2	2	1	7	4	1	3
17	SequoiaView	3	1	1	7	6	5	4
17	SequoiaView	4	3	1	6	1	2	2.6
17	SequoiaView	5	4	1	6	2	4	3.4
17	SequoiaView	6	4	1	6	1	1	2.6
17	SequoiaView	7	4	3	1	3	4	3
17	SequoiaView	8	1	2	7	3	1	2.8
17	SequoiaView	9	2	2	7	3	2	3.2
17	SequoiaView	10	5	5	7	7	7	6.2
17	SequoiaView	11	5	4	7	7	7	6
17	SequoiaView	12	3	4	7	4	4	4.4
17	SequoiaView	13	1	1	7	2	2	2.6

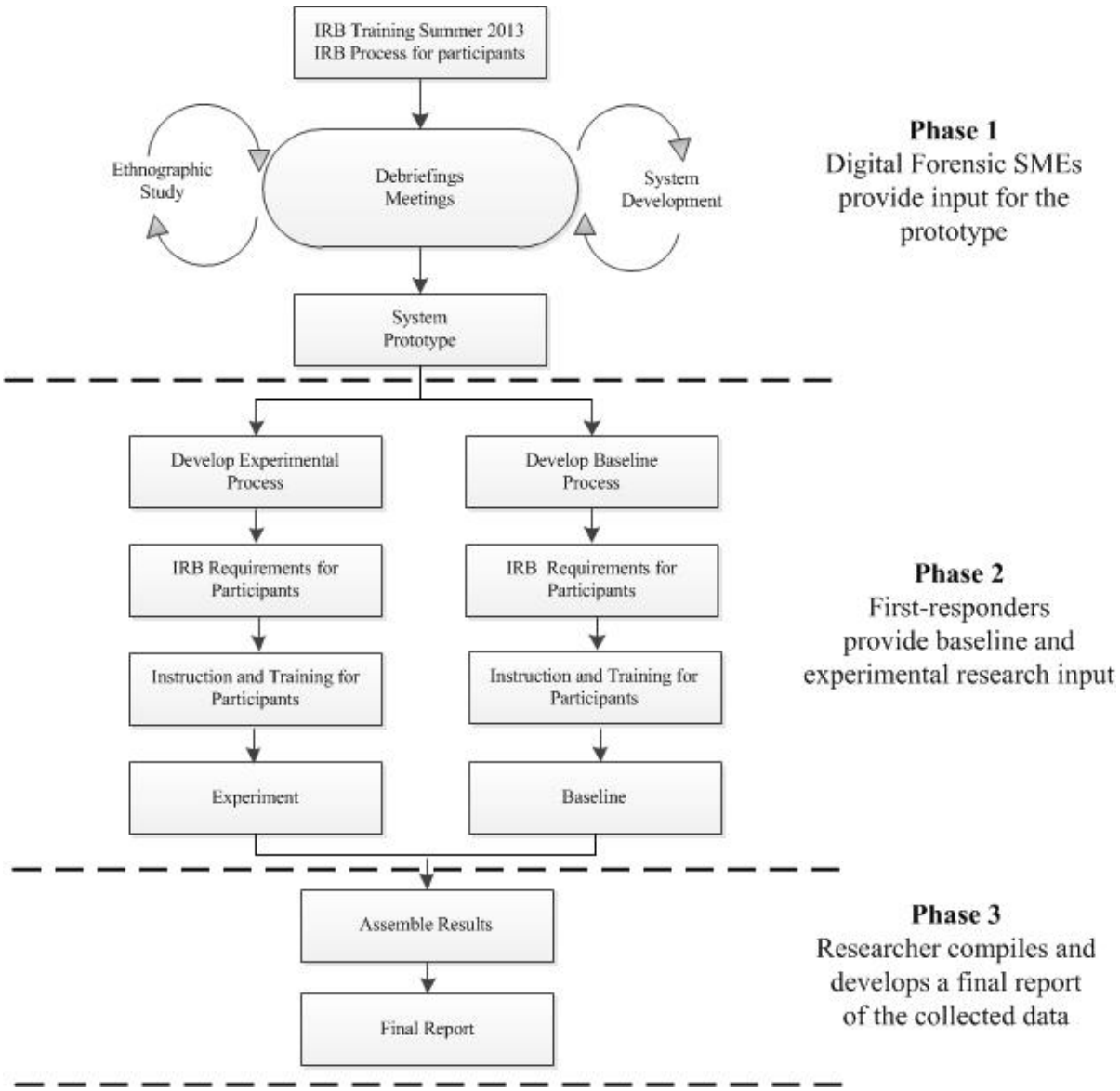
17	SequoiaView	14	3	1	7	4	2	3.4
17	SequoiaView	Mean	3	2	6.29	3.5	3.29	3.61
17	Sleuth Kit	1	2	1	1	1	1	1.2
17	Sleuth Kit	2	2	1	1	1	1	1.2
17	Sleuth Kit	3	2	1	1	1	1	1.2
17	Sleuth Kit	4	2	1	1	1	1	1.2
17	Sleuth Kit	5	2	1	1	1	1	1.2
17	Sleuth Kit	6	2	1	1	1	1	1.2
17	Sleuth Kit	7	2	1	1	1	1	1.2
17	Sleuth Kit	8	2	1	1	1	1	1.2
17	Sleuth Kit	9	2	1	1	1	1	1.2
17	Sleuth Kit	10	2	1	1	1	1	1.2
17	Sleuth Kit	11	2	1	1	1	1	1.2
17	Sleuth Kit	12	2	1	1	1	1	1.2
17	Sleuth Kit	13	2	1	1	1	1	1.2
17	Sleuth Kit	14	2	1	1	1	1	1.2
17	Sleuth Kit	Mean	2	1	1	1	1	1.2
18	SequoiaView	1	1	1	1	1	1	1
18	SequoiaView	2	1	1	1	1	1	1
18	SequoiaView	3	1	1	1	1	1	1
18	SequoiaView	4	1	1	1	1	1	1
18	SequoiaView	5	1	1	1	1	1	1
18	SequoiaView	6	3	3	3	3	3	3
18	SequoiaView	7	1	1	1	1	1	1
18	SequoiaView	8	1	1	1	1	1	1
18	SequoiaView	9	1	1	1	1	1	1
18	SequoiaView	10	1	1	1	1	1	1
18	SequoiaView	11	1	1	1	1	1	1
18	SequoiaView	12	1	1	1	1	1	1
18	SequoiaView	13	1	1	1	1	1	1
18	SequoiaView	14	1	1	1	1	1	1
18	SequoiaView	Mean	1.14	1.14	1.14	1.14	1.14	1.14
18	Sleuth Kit	1	1	1	1	1	1	1
18	Sleuth Kit	2	1	1	1	1	1	1
18	Sleuth Kit	3	1	1	1	1	1	1
18	Sleuth Kit	4	1	1	1	1	1	1
18	Sleuth Kit	5	3	1	1	1	1	1.4
18	Sleuth Kit	6	3	1	1	1	1	1.4
18	Sleuth Kit	7	1	1	1	1	1	1
18	Sleuth Kit	8	1	1	1	1	1	1
18	Sleuth Kit	9	1	1	1	1	1	1
18	Sleuth Kit	10	1	1	1	1	1	1
18	Sleuth Kit	11	1	1	1	1	1	1
18	Sleuth Kit	12	1	1	1	1	1	1
18	Sleuth Kit	13	1	1	1	1	1	1
18	Sleuth Kit	14	1	1	1	1	1	1
18	Sleuth Kit	Mean	1.29	1	1	1	1	1.06
19	SequoiaView	1	4	1	1	3	1	2
19	SequoiaView	2	4	1	1	1	1	1.6
19	SequoiaView	3	1	1	1	1	1	1
19	SequoiaView	4	1	1	1	1	1	1

19	SequoiaView	5	1	1	1	1	1	1
19	SequoiaView	6	1	1	1	1	1	1
19	SequoiaView	7	1	1	1	1	1	1
19	SequoiaView	8	1	1	1	1	1	1
19	SequoiaView	9	1	1	1	1	1	1
19	SequoiaView	10	1	1	1	1	1	1
19	SequoiaView	11	1	1	1	1	1	1
19	SequoiaView	12	1	1	1	1	1	1
19	SequoiaView	13	1	1	1	1	1	1
19	SequoiaView	14	1	1	1	1	1	1
19	SequoiaView	Mean	1.43	1	1	1.14	1	1.11
19	Sleuth Kit	1	3	1	1	1	1	1.4
19	Sleuth Kit	2	3	1	1	1	1	1.4
19	Sleuth Kit	3	1	1	1	1	1	1
19	Sleuth Kit	4	1	1	1	1	1	1
19	Sleuth Kit	5	1	1	1	1	1	1
19	Sleuth Kit	6	1	1	1	1	1	1
19	Sleuth Kit	7	1	1	1	1	1	1
19	Sleuth Kit	8	1	1	1	1	1	1
19	Sleuth Kit	9	1	1	1	1	1	1
19	Sleuth Kit	10	1	1	1	1	1	1
19	Sleuth Kit	11	1	1	1	1	1	1
19	Sleuth Kit	12	1	1	1	1	1	1
19	Sleuth Kit	13	1	1	1	1	1	1
19	Sleuth Kit	14	1	1	1	1	1	1
19	Sleuth Kit	Mean	1.29	1	1	1	1	1.06
20	SequoiaView	1	2	1	1	2	1	1.4
20	SequoiaView	2	2	1	1	2	1	1.4
20	SequoiaView	3	2	1	1	3	1	1.6
20	SequoiaView	4	3	1	1	3	1	1.8
20	SequoiaView	5	4	1	1	3	1	2
20	SequoiaView	6	2	1	1	2	1	1.4
20	SequoiaView	7	3	2	1	3	2	2.2
20	SequoiaView	8	3	2	2	3	2	2.4
20	SequoiaView	9	2	2	2	2	2	2
20	SequoiaView	10	2	1	1	1	1	1.2
20	SequoiaView	11	3	2	1	2	1	1.8
20	SequoiaView	12	3	3	2	2	1	2.2
20	SequoiaView	13	2	1	1	1	1	1.2
20	SequoiaView	14	2	1	1	2	1	1.4
20	SequoiaView	Mean	2.5	1.43	1.21	2.21	1.21	1.71
20	Sleuth Kit	1	6	5	7	6	6	6
20	Sleuth Kit	2	6	5	7	6	6	6
20	Sleuth Kit	3	6	5	7	6	6	6
20	Sleuth Kit	4	6	4	7	6	5	5.6
20	Sleuth Kit	5	6	5	6	5	6	5.6
20	Sleuth Kit	6	6	4	5	6	5	5.2
20	Sleuth Kit	7	7	6	6	6	6	6.2
20	Sleuth Kit	8	7	7	6	7	7	6.8
20	Sleuth Kit	9	6	5	6	6	6	5.8
20	Sleuth Kit	10	4	5	6	6	6	5.4

20	Sleuth Kit	11	5	5	6	5	6	5.4
20	Sleuth Kit	12	6	5	6	5	6	5.6
20	Sleuth Kit	13	4	5	4	5	5	4.6
20	Sleuth Kit	14	6	5	5	6	5	5.4
20	Sleuth Kit	Mean	5.79	5.07	6	5.79	5.79	5.69

Appendix B

The Study Process Flow



Appendix C

Institutional Review Board Memorandum

NOVA SOUTHEASTERN UNIVERSITY
Office of Grants and Contracts
Institutional Review Board

MEMORANDUM

To: Robert Altiero

From: Ling Wang, Ph.D.
Institutional Review Board

Date: Nov. 25, 2013

Re: *Digital Forensics Tool Interface Visualization*

IRB Approval Number: wang09151306

I have reviewed the above-referenced research protocol at the center level. Based on the information provided, I have determined that this study is exempt from further IRB review. You may proceed with your study as described to the IRB. As principal investigator, you must adhere to the following requirements:

1) CONSENT: If recruitment procedures include consent forms these must be obtained in such a manner that they are clearly understood by the subjects and the process affords subjects the opportunity to ask questions, obtain detailed answers from those directly involved in the research, and have sufficient time to consider their participation after they have been provided this information. The subjects must be given a copy of the signed consent document, and a copy must be placed in a secure file separate from de-identified participant information. Record of informed consent must be retained for a minimum of three years from the conclusion of the study.

2) ADVERSE REACTIONS: The principal investigator is required to notify the IRB chair and me (954-262-5369 and 954-262-2020 respectively) of any adverse reactions or unanticipated events that may develop as a result of this study. Reactions or events may include, but are not limited to, injury, depression as a result of participation in the study, life-threatening situation, death, or loss of confidentiality/anonymity of subject. Approval may be withdrawn if the problem is serious.

3) AMENDMENTS: Any changes in the study (e.g., procedures, number or types of subjects, consent forms, investigators, etc.) must be approved by the IRB prior to implementation. Please be advised that changes in a study may require further review depending on the nature of the change. Please contact me with any questions regarding amendments or changes to your study.

The NSU IRB is in compliance with the requirements for the protection of human subjects prescribed in Part 46 of Title 45 of the Code of Federal Regulations (45 CFR 46) revised June 18, 1991.

Cc: Protocol File

3301 College Avenue • Fort Lauderdale, FL 33314-7796 • (954) 262-5369
Fax: (954) 262-3977 • Email: *irgta@nsu.nova.edu* • Web site: www.nova.edu/cwis/ogc

Appendix D

Consent Form for Participation

Consent Form for Participation in the Research Study Entitled

Digital Forensics Tool Interface Visualization

Funding Source: None

IRB protocol #: wang09151306

Principal investigator

Robert Altiero, PhD Candidate

7765 Sutton Ct

Port Tobacco, MD 20677

(301)751-6419

Co-investigator

Maxine S. Cohen, PhD

Graduate School of Computer and Information Sciences

3301 College Avenue

Fort Lauderdale-Davie, FL 33314-7796

(954) 262-2072

For questions/concerns about your research rights, contact:
Human Research Oversight Board (Institutional Review Board or IRB)
Nova Southeastern University
(954) 262-5369/Toll Free: 866-499-0790
IRB@nsu.nova.edu

Site Information:
Site 1: Southern Maryland Business Center, 10665 Stanhaven Place Suite 300A, White Plains, MD 20695

What is the study about?

The purpose of this research is to investigate how a digital forensics tool interface enhanced with visualization techniques may show improved capabilities to an investigator's cognitive capacities in the discovery of criminal evidence.

Why are you asking me?

You have been selected to participate due to your qualifications as a subject matter expert (SME) or as a first-responder. Criteria:

1) SME-digital forensic experts

-- Work in the Washington, DC Metropolitan Area

-- Digital forensics is their primary duty

-- Experienced in law enforcement or other investigative professions

-- College-educated or have completed dedicated on-the-job training

-- Experts in the Microsoft Windows operating systems

-- A specialty niche area such as mobile devices

-- Professional digital forensic certification such as the Certified Computer Examiner (CCE)

2) First-responder, network administrators, or information security information system professionals

-- Work in the Washington, DC Metropolitan Area

-- Two years of experience using analysis tools, tactics, techniques, and procedures for

evaluating computer event information

-- Professional certification—such as Security+, A+, Network+, Certified Information Systems Security Professional (CISSP) or Systems Security Certified Practitioner (SSCP)

What will I be doing if I agree to be in the study?

As a study participant SME, you will be answering open-ended questions about your professional environment in developing a digital forensics investigator's primary tasks for evidence identification while operating a traditional digital forensics tool set. These techniques will simulate the analysis phase of an investigation for digital forensics and will represent the primary tasks to be followed in the final phases of this research.

As a study participant first-responder, you will be instructed to perform the digital forensics investigator's primary tasks for evidence identification while operating a traditional digital forensics tool set in addition to using a visualized interface. These techniques are meant to simulate the analysis phase of an investigation for digital forensics.

Cognitive load, an assessment of human working memory, will be measured using a Likert scale while you are performing a predefined set of tasks. Determined, is the measurement of a prototype visualized application intended to improve the effectiveness of a digital forensics investigation. Efficiency levels will be compared and contrasted for results of the users' surveys to determine the benefit of a visualized application.

Is there any audio or video recording?

There are no audio or video recordings associated with this research.

What are the dangers to me?

Foreseeable risks or discomforts are minimal associated with this research. However, foreseen risks are listed below. The procedures or activities in this study may have unknown or unforeseeable risks.

If you have any questions about the research or your research rights or a research-related injury, please contact Robert Altiero, principal investigator or Maxine Cohen, advisor. You may also contact the IRB at the numbers indicated above with questions about your research rights.

Risk/Discomfort:	Violation of Privacy
Likelihood:	Low
Magnitude/Duration:	Low/Minimal

Risk Minimization: Participant surveys will not identify participants by name. Participant personnel information will not be stored electronically and will be destroyed 36 months after research completion.

Risk/Discomfort: Legal Risks

Likelihood: Low

Magnitude/Duration: Low/Minimal

Risk Minimization: Since this is a digital forensic research project data is needed. Victim and perpetrator data will be simulated in order not to compromise victim or accused's case rights.

Risk/Discomfort: Psychosocial Stress

Likelihood: Low

Magnitude/Duration: Low/Survey completion time—one hour

Risk Minimization: The researcher will ensure that the participants understand that research participation is voluntary and that they are informed of consent. All risk will be addressed and discussed with the participants, including psychosocial stress. Additionally, all meetings will be scheduled at the participants' convenience, and a description of the research and surveys will be given to the participants in advance.

Are there any benefits for taking part in this research study?

There are no direct benefits. However, by participating in this study you will be assisting the researcher in contributing to the body of knowledge by validating the method of measure and by providing empirical evidence consistent with the betterment of digital forensics tools.

Will I get paid for being in the study? Will it cost me anything?

There are no costs to you or payments made for participating in this study.

How will you keep my information private?

All identifying documentation leading to the subject's name will be destroyed by shredding after 36 months following the study's completion. Subject identification will not be used to identify questionnaires; rather an ID will be used to determine respondent.

What if I do not want to participate or I want to leave the study?

You have the right to leave this study at any time or to refuse to participate. If you do decide to leave or you decide not to participate, you will not experience any penalty or loss of services you have a right to receive. If you choose to withdraw, any information collected about you **before** the date you leave the study will be kept in the research records for 36 months from the conclusion of the study, but you may request that it not be used.

Other Considerations:

If significant new information relating to the study becomes available, which may relate to your willingness to continue to participate, this information will be provided to you by the investigators.

Voluntary Consent by Participant:

By signing below, you indicate that

- this study has been explained to you
- you have read this document or it has been read to you
- your questions about this research study have been answered
- you have been told that you may ask the researchers any study-related questions in the future or contact them in the event of a research-related injury
- you have been told that you may ask Institutional Review Board (IRB) personnel questions about your research rights
- you are entitled to a copy of this form after you have read and signed it
- you voluntarily agree to participate in the study entitled Digital Forensics Tool Interface Visualization

Participant's Signature: _____ Date: _____

Participant's Name: _____ Date: _____

Signature of Person Obtaining Consent: _____

Date: _____

Appendix E

Questionnaire

Participant ID:

Date:

Interface Used:

SequoiaView

Autopsy

Task #:

Please indicate in one column with an "X" the numeral that most accurately represents your experience as you performed the presented task.

		Low Mental Effort			**Neither Easy nor Difficult**		**Very Difficult**	
	Question	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	I was able to gain an overview of the entire data set.							
2	I was able to zoom in on items of interest.							
3	I was able to filter out irrelevant items.							
4	I was able to select an item or group of items to get the details that I needed.							
5	I was able to view relationships							

	among files.							
6	I was able to keep a history of actions through *undo* when necessary.							
7	I was able to identify sub-collections of the query parameters.							
8	I was able to develop a three-dimensional relationship of the file's size relative to other files present.							
9	I was easily able to identify file temporal information.							
10	I was easily able to identify file hierarchy in parent-child relationships.							
11	I was able to view relationships among files.							
12	I was able to identify sub-collections and query taxonomy parameters once a file of interest had been located.							
13	I was able to adjust filtering ranges dynamically.							
14	The level of mental effort difficulty to complete the task.							

Questions 1-14 represent the basic design guidelines for the interface presenting visual information as presented by Shneiderman (1996).

Reference

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Visual Languages*, IEEE Symposium, 336-343. Boulder, CO: IEEE. doi:10.1109/VL.1996.545307

Appendix F

Study Instructions

Background:
Recent trends show digital devices utilized with increasing frequency in most crimes committed. Investigating crime involving these devices is labor-intensive for the practitioner applying digital forensics tools that present possible evidence with results displayed in tabular lists for manual review. This research investigates how enhanced digital forensics tool interface visualization techniques can be shown to improve the investigator's cognitive capacities to discover criminal evidence more efficiently.

Purpose:
The primary operations of a digital forensic investigation often depend upon the capture of file-centric evidence. This research presents visualization graphs and contrasts their properties with the outputs of The Sleuth Kit (TSK) contemporary digital forensic program's textual-based interface in order to prove the effectiveness of enhanced data presentation. There is potential for the computer interface to present to the digital forensic practitioner an abstract, graphic view of an entire dataset of computer files. Enhanced interface design of digital forensic tools means more rapidly linking suspicious evidence to a perpetrator.

Your Role:
You will play the role of the digital investigator in this study by performing the five primary tasks listed below. These tasks are often performed by digital investigators. Your investigation will be to examine a flash drive suspected of being a storage device utilized by a perpetrator, see table 2.

Task 1: Locate files with the .jpg extension

Task 2: Locate the file named Kittie.jpg

Task 3: Date range to establish timeline

Task 4: Identify directory structure that takes up the most space

Task 5: Identify the largest file

Testing your cognitive load is simply answering how difficult a task is to perform. The research shows that you are the best measure for estimating your cognitive load.

SequoiaView

The visualized application used for this study is SequoiaView. SequoiaView provides a global view of an entire selected file structure utilizing a treemap presentation. Treemaps provide a visual representation of all files and directories simultaneously and are only limited by the available screen space. Treemaps present a solution for efficient use of the available space.

SequoiaView Launch Sequence

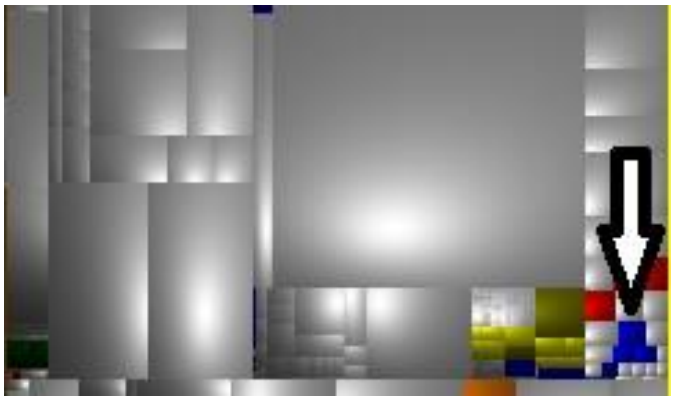
1. Open SequoiaView by clicking the **Start** button. Select **All Programs,** in the list of program results, click **SequoiaView**. Select the **SequoiaView** button to launch the application.
2. Close the **Scan** window. Select the **Browse** button and folders, and then click Select Removable Drive F: Removable Disk (F:) .
3. The **browse for folder** widow will open, again select Removable Disk F: Removable Disk (F:) and click **OK**.

   Each file and folder is represented in a single display as tiled squares and rectangles. The file types are represented by the tile color.

TASK INSTRUCTIONS:

Task 1: Locate files with the **.jpg** extension.
There are two options for identifying the file type .jpg: by using the presentation and 2) using Display the Filter option. You will perform both options.

1. Identify the files with '.jpg' extension by locating those tiles colored **blue**. Note: All graphic files are varying shades of blue.



Mousing over the tiles reveals the file name. By **right clicking** on the tiles, the file properties are displayed, and traversing up and down directories is also possible.

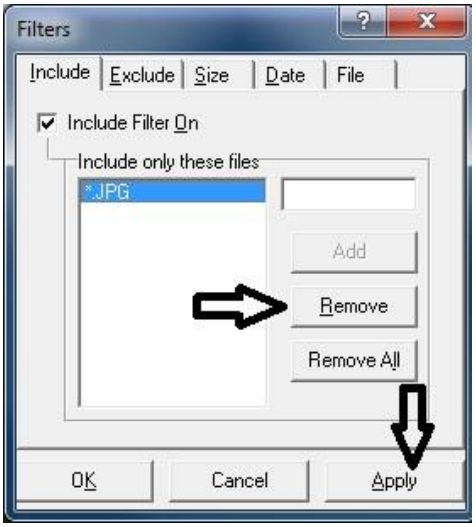
2. On the menu bar select the arrow on the **filter options**  . When the filter dialog box pops up, click the **include filter on** checkbox to enable the filter, and specify '*.jpg' in the **edit box**, and click **add**. Now click **Apply.**

Again, by **mousing** over the tiles reveals the file name. **Right clicking** on the tiles the displays file properties; traversing up and down directories is also possible.

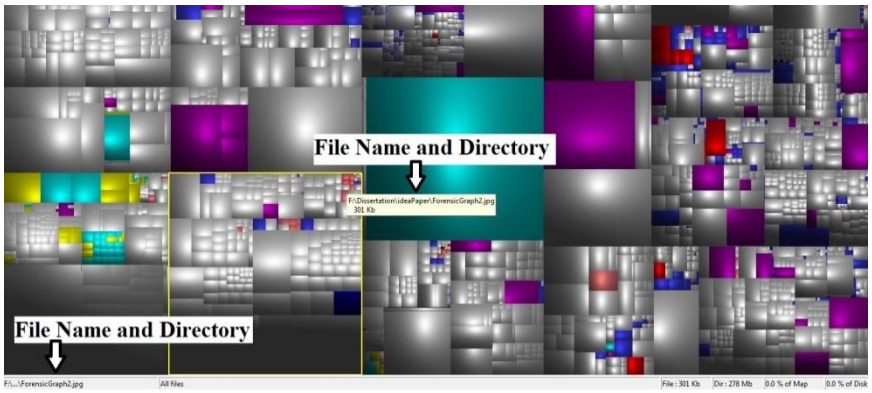
You have completed **Task 1.** Please complete the questionnaire.

Task 2: Locate the file named **Kittie.jpg**

1. Perform the **SequoiaView Launch Sequence** above as needed. Just as with Task 1, there are two options to completing **Task 2**: 1) using the presentation—remember that files with the '.jpg' extension are represented with blue tiles and 2) using the filter option. You will perform both options. Return to the **filter options**  ; when the filter dialog box pops up, click '*.jpg' in the **edit box** and click **Remove**. Now click **Apply.**



1. Locate the file Identified as '**Kittie.jpg**' by locating its **blue**-colored tile.



Mouse over the tiles until locating the file named '**Kittie.jpg**'. **Mousing** over the tiles, the file name causes a directory to pop up; and the name of the file is also located in the lower left information bar. **Right clicking** on the tiles displays the file properties, and traversing up and down directories is also possible.

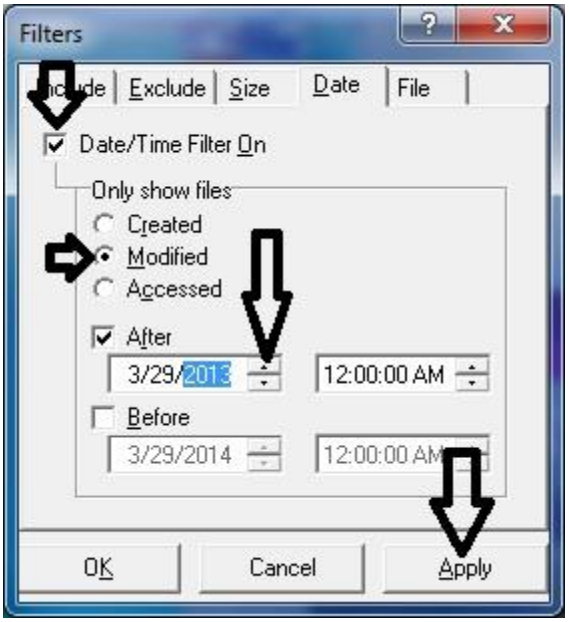
2. On the menu bar select the **down arrow** on the filter options  . When the filter dialog box pops up, click the **include filter on** checkbox to enable the filter, specify '**Kittie.jpg**' in the **edit box,** and click **add**. Now click **Apply.**

Mousing over the tile presents the file name **Right clicking** on the tiles displays the file properties; traversing up or opening the file is also possible.

You have completed **Task 2.** Please complete the questionnaire.

Task 3: Date range to establish timeline
Perform the **SequoiaView Launch Sequence** above as needed. **Task 3** establishes a timeline by applying the date/time filter option and locating suspected files utilizing the presentation. Utilizing the date/time filter eliminates any files presented outside the selected range.

1. On the menu bar select the **down arrow** on the filter options  . When the filter dialog box pops up, select the **Date** tab then click the **Date/Time filter on** checkbox to enable the filter; next click the **Modified** radio button and now click the **After** checkbox**.** Finally, click on the **year** and with the **down arrow** select **2013**. Click **Apply.**

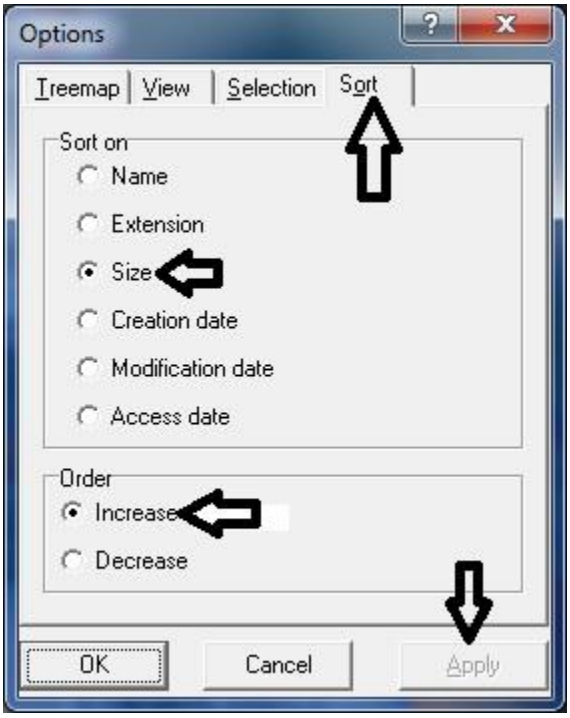


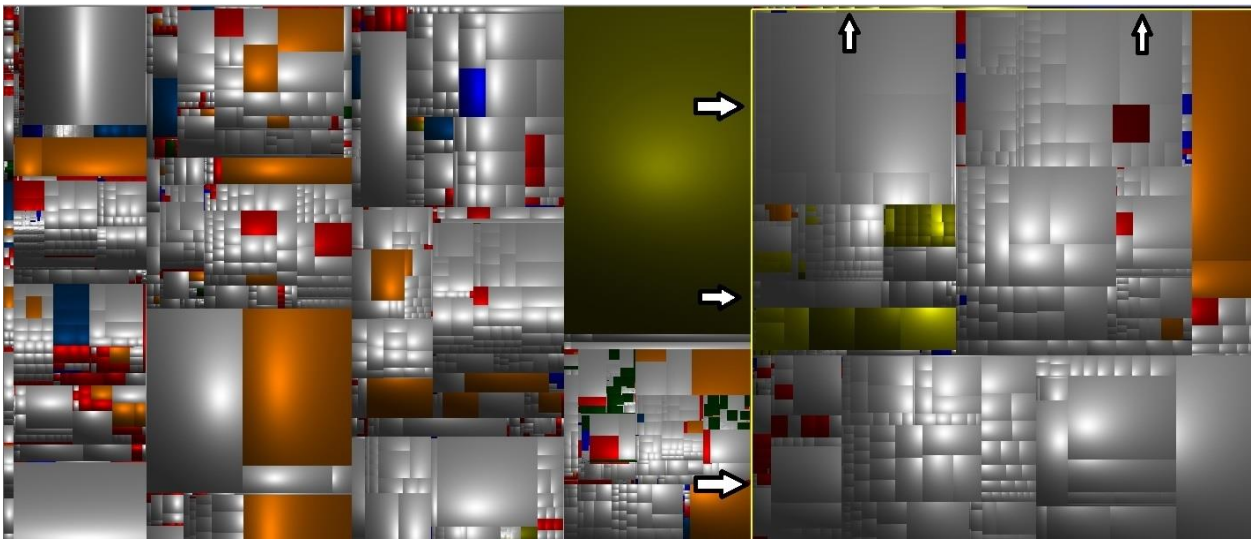
Once again, **Mouse** over the tiles until locating the file named '**Kittie.jpg**'. **Mousing** over the tiles the file name displays a directory; pops up is the name of the file is also located in the lower left-hand portion of the information bar. **Right clicking** on the tiles displays the file properties; traversing up and down directories is also possible.

You have completed **Task 3.** Please complete the questionnaire.

Task 4: Directory structure that takes up the most space
Perform the **SequoiaView Launch Sequence** above as needed. **Task 4** establishes the largest file directories. By enabling the **Treemap Sort Options**, you may order the presentation by directory size. The graphic display presents root level directories outlined in yellow when the directory is **moused** over.

1. On the menu bar click on the **Options** button  . When the **Options** dialog box pops up, click the **Sort** tab and then click on the **Size** radio button to enable sorting by size; click on the **Increase** radio button to present the larger root level directories on the right-hand side of the presentation. Now click **Apply.**



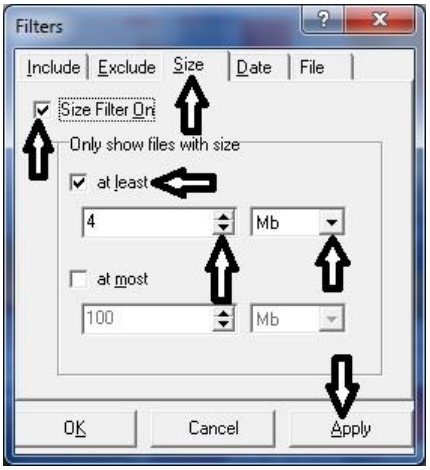


Mouse over the tiles until locating the largest root level directory on the right. Notice the yellow outline.

You have completed **Task 4.** Please complete the questionnaire.

Task 5: Identify the largest file
Perform the **SequoiaView Launch Sequence** above as needed. **Task 5** identifies files by size. Enable the **Filter Options**, and the presentation will identify file size. The graphic display presents tiles proportionate to their file size **Mousing** over the tiles presents the file directory, name, and file size.

1. On the menu bar select the **down arrow** on the filter options  . When the filter dialog box pops up, select the **Size** tab then click the **Size filter on** checkbox to enable the filter, next click the **at least** checkbox**.** Specify the specified in bytes and megabytes by selecting the appropriate option from the right-most comboboxes and enter '**4'** in the edit boxes on the left**,** filtering for files that are at least **4 MB** in size. Click **Apply.**



Once again, locate the largest file and **Mouse** over the tile to display the directory, file name, and size. Can you locate the file '**Kittie.jpg**'? Again, **right clicking** on the tiles displays the file properties; traversing up and down directories is also possible.

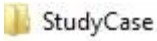
You have completed **Task 5,** the graphic display portion of the study**.** Please complete the questionnaire.

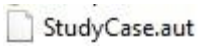
SleuthKit (Autopsy)

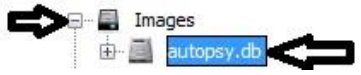
SleuthKit is an open-source traditional digital forensics investigation tool that runs on multiple platforms. The tool is used to analyze disk images and perform in-depth analysis of file systems. Examiners and analysts can use the Autopsy graphical interface to conduct an investigation and to interface with SleuthKit rather than using the command line. For this study you will use the Autopsy to interact with the SleuthKit tool and perform the tasks necessary to conduct the study.

Autopsy Launch Sequence

1. Open Autopsy by clicking the **Start** button. Select **All Programs;** in the list of program results, click **Autopsy** to launch the application.
2. When the **Autopsy Welcome** window pops up, select **Open Existing Case** 

. When the **Open** dialog pops up, select the folder **StudyCase** 

. Next, select the case file **StudyCase.aut** 

and click **open**.
3. Once the Autopsy case opens, expand the **Images** and then **click** the **autopsy.db** 

.

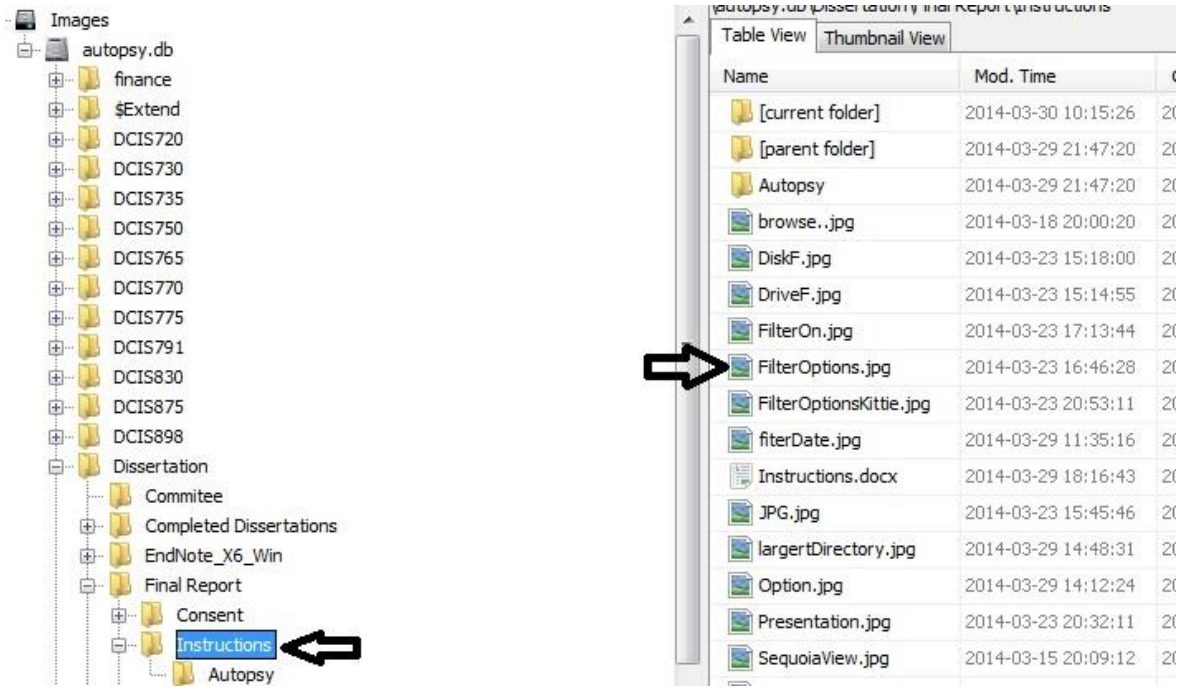
The expanded **autopsy.db** image is a traditional Windows hierarchical folder structure known as the **Data Explorer** or **Directory Tree,** which can be traversed and expanded. The Directory and File attributes are presented in the **Results Viewer** in the upper right of the interface under the **Directory Listing** tab. The file and folder attributes are presented within the **Table View** tab, and thumbnails of graphic files can be viewed within the **Thumbnails** tab.

Note: "**Images**" here relates to copies of a drive or directory as a file structure image.

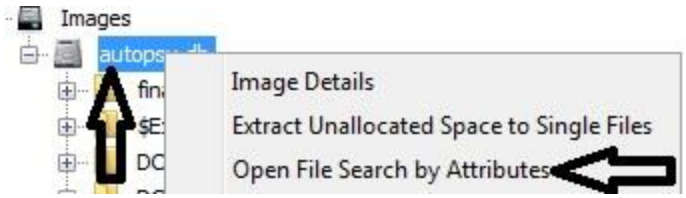
TASK INSTRUCTIONS:

Task 1: Locate files with the **.jpg** extension.
There are two options for identifying the file type '**.jpg**': 1) using the **Data Explorer** presentation 2) using **File Search By Attributes** filter option. You will perform both options.

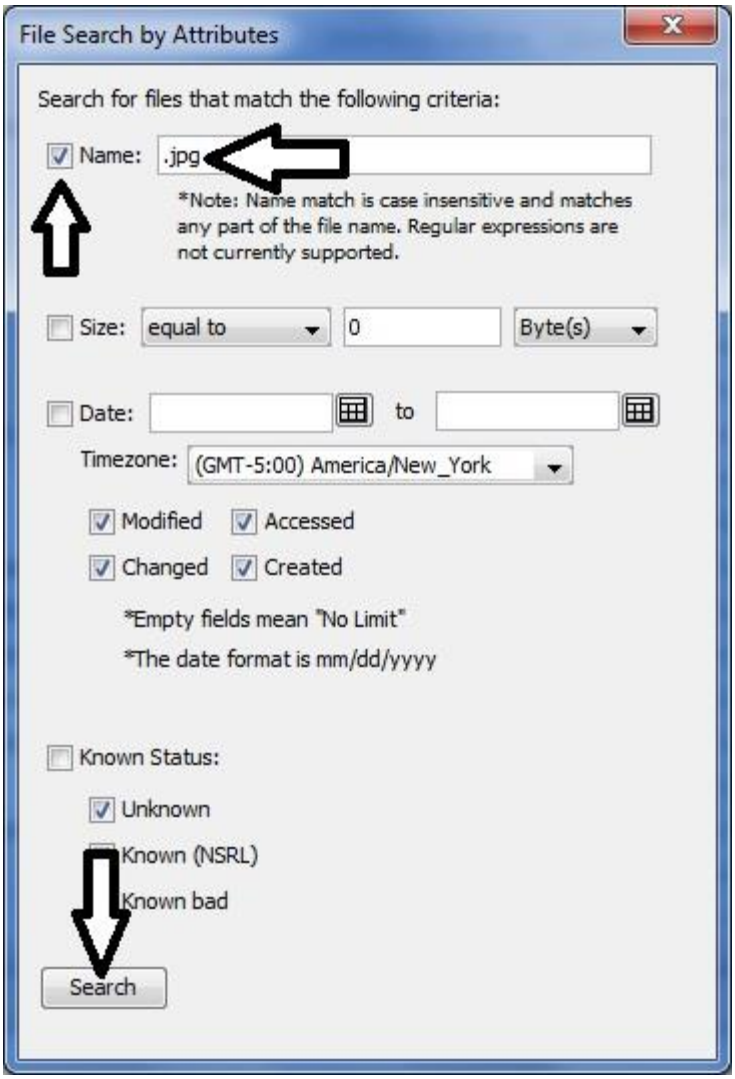
1. Identify the files with '.jpg' extension by locating those files in the **autopsy.db** image, expanding the hierarchical **Data Explorer** folder in the **Data Tree** and viewing their attributes in the **Table View** tab on the right of the interface.



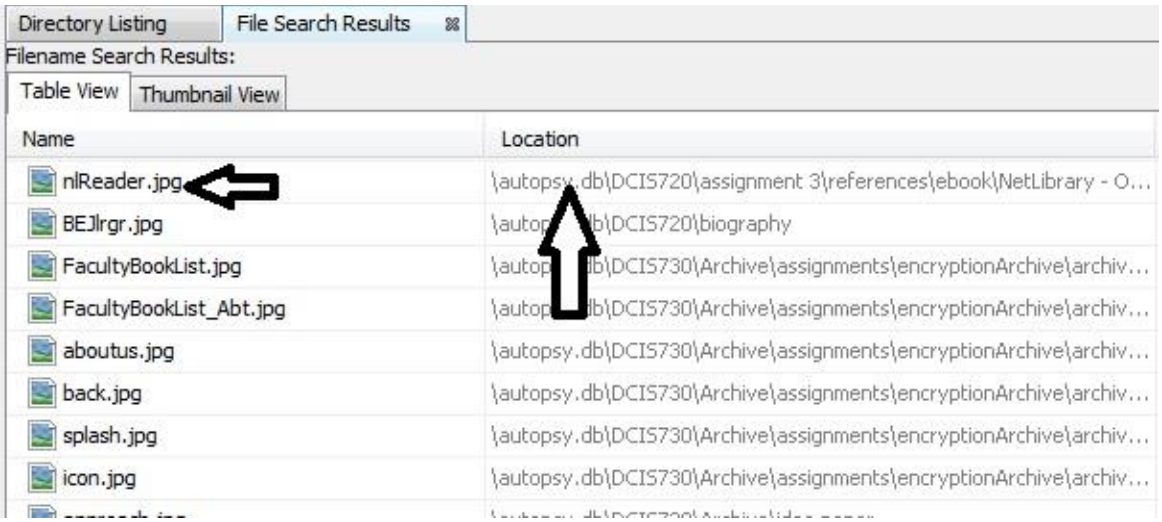
2. In the **Data Explorer** right click the **autopsy.db** image and, when the dialog box pops up, select the **File Search by Attributes** option.



In the **File Search by Attribute** dialog box, click the **Name** checkbox and type '**.jpg'** in the text box. Now click **Search**.



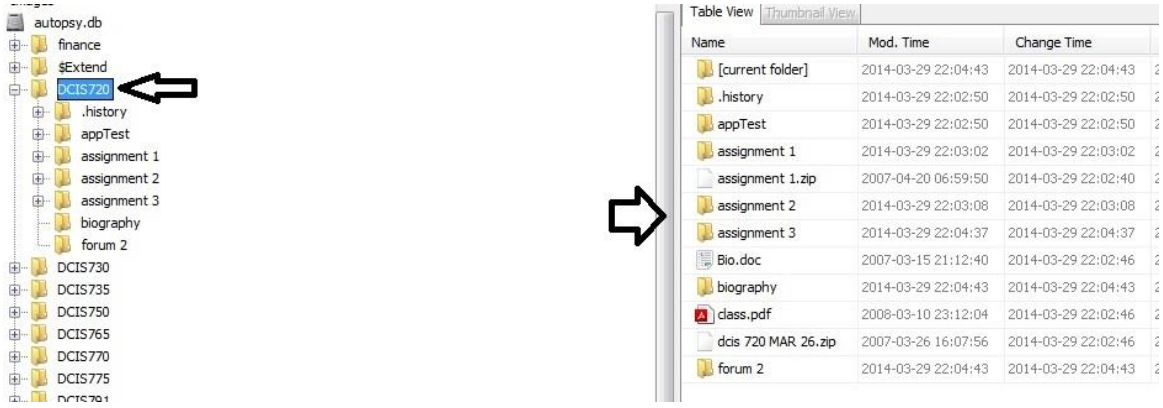
The search results appear in the **Directory Listing** under the **Table View** tab. The **Name** of the file and the respective directory are presented.



You have completed **Task 1.** Please complete the questionnaire.

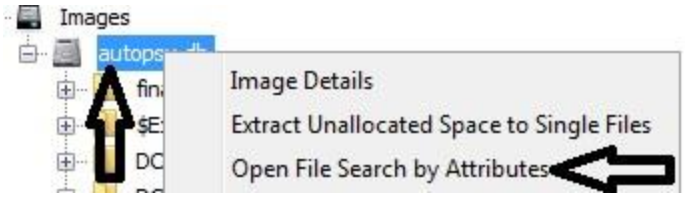
Task 2: Locate the file named **Kittie.jpg**
Perform the **Autopsy Launch Sequence** above as needed. Just as with Task 1, there are two options to completing **Task 2**: 1) use the **Data Explorer** to expand each file folder and search the files listed in the **table view** for the file **Kittie.jpg** 2) use the **File Search by Attribute** option. You will perform both options.

1. Locate the file Identified as '**Kittie.jpg**' by searching the **Directory Tree**.

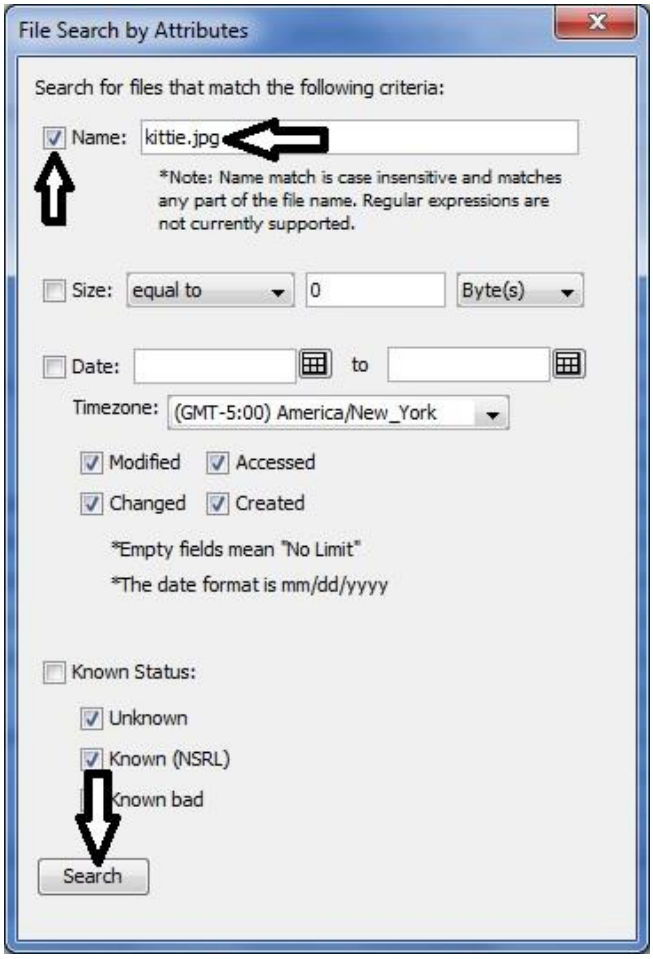


By right clicking the file folder you have the option of saving the directory folder or to collapsing the folder. By right clicking on the file you will receive a dialog with several options for opening, extracting, or bookmarking the file.

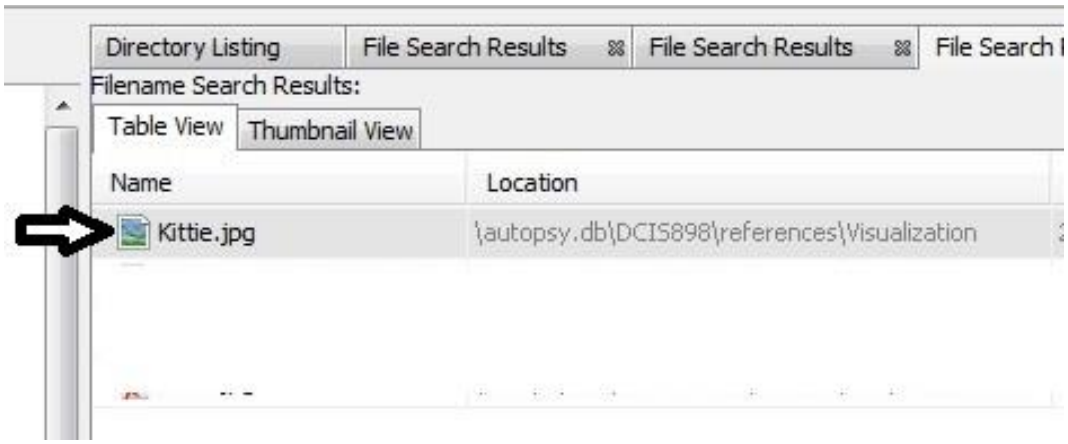
2. In the **Data Explorer,** right click **autopsy.db** image and, when the dialog box pops up, select the **File Search by Attributes** option.



In the **File Search by Attribute** dialog box, click the **Name** checkbox and type "**kittie.jpg**" in the text box. Now click **Search**.



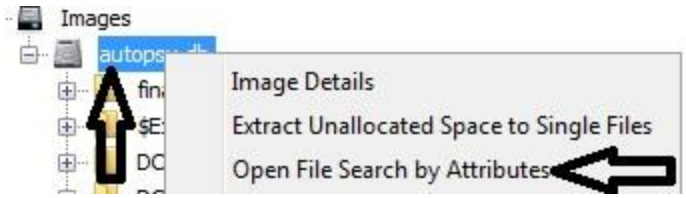
The search results are presented in the **Directory Listings** results viewer within the **Table View** tab.



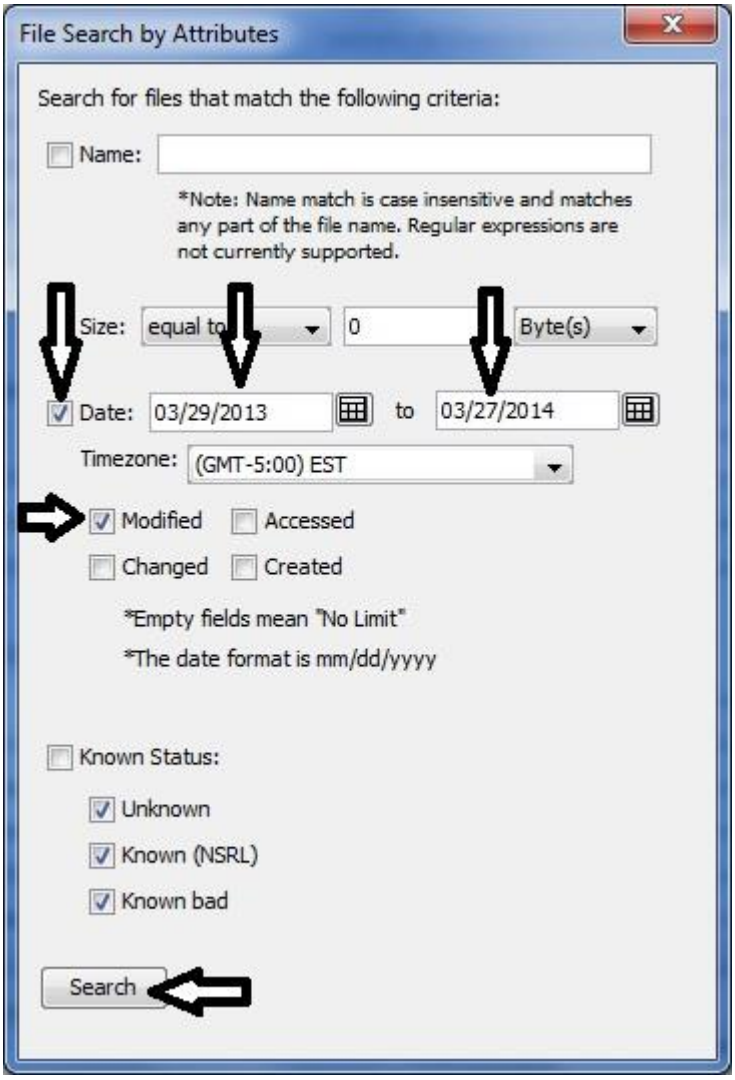
You have completed **Task 2.** Please complete the questionnaire.

Task 3: Date range to establish timeline
Perform the **Autopsy Launch Sequence** above as needed. **Task 3** establishes a timeline by applying the **File Search by Attribute** option and locating suspected files by utilizing the results pane **Table View**. Utilizing the **File Search by Attribute** filter eliminates any files presented outside the selected range.

1. In the **Data Explorer,** right click **autopsy.db** image and, when the dialog box pops up, select the **File Search by Attributes** option.



In the **File Search by Attribute** dialog box, click the **Date** checkbox and select the **from to dates**—enter the dates as seen here (03/29/2013 to 03/27/2014). Click the checkbox **Modified** and **uncheck** the check boxes **accessed**, **created,** and **changed**. Now click **Search**.

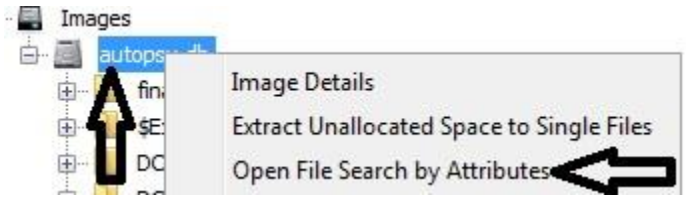


Again, by right clicking the file folder you have the option of saving the directory folder or collapsing the folder. By right clicking on the file you will receive a dialog with several options for **opening**, **extracting**, or **bookmarking** the file.

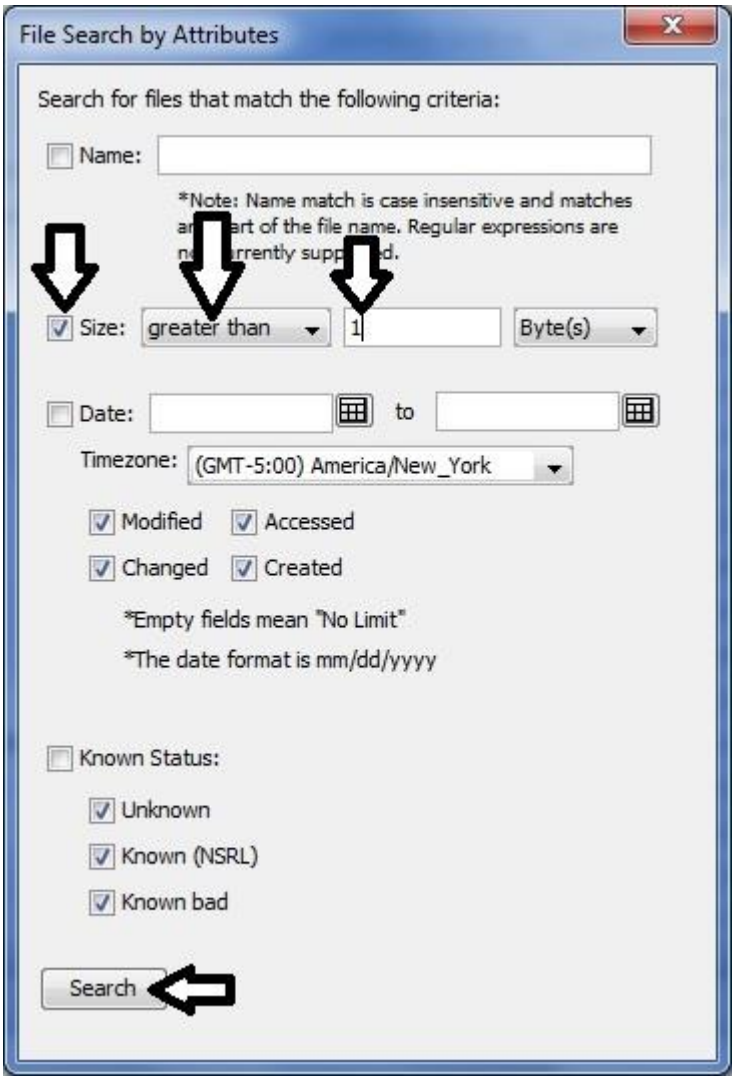
You have completed **Task 3.** Please complete the questionnaire.

Task 4: Directory structure that takes up the most space
Perform the **Autopsy Launch Sequence** above as needed. **Task 4** establishes the largest file directories. Enable **File Search by Attribute** to prompt the **Table View** to present/ identify files/directories by size.

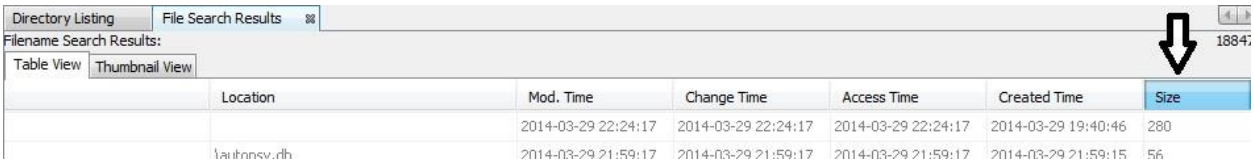
1. In the **Data Explorer,** right click **autopsy.db** image and, when the dialog box pops up, select the **File Search by Attributes** option.



In the **File Search by Attribute** dialog box, click the **Size** checkbox, change the pattern to **greater than**, and select the size to be greater than **1 byte**. Now click **Search**.



Once the results are presented in the **Table View,** click **Size** twice TO ORDER the files/directories largest to smallest. This may take a minute.

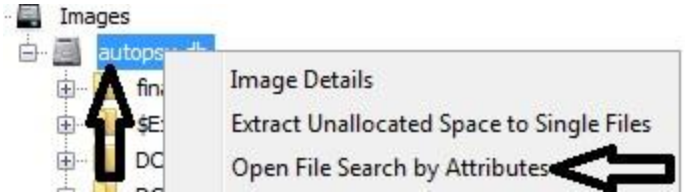


Now scroll to find the largest directory in the **Table View**.

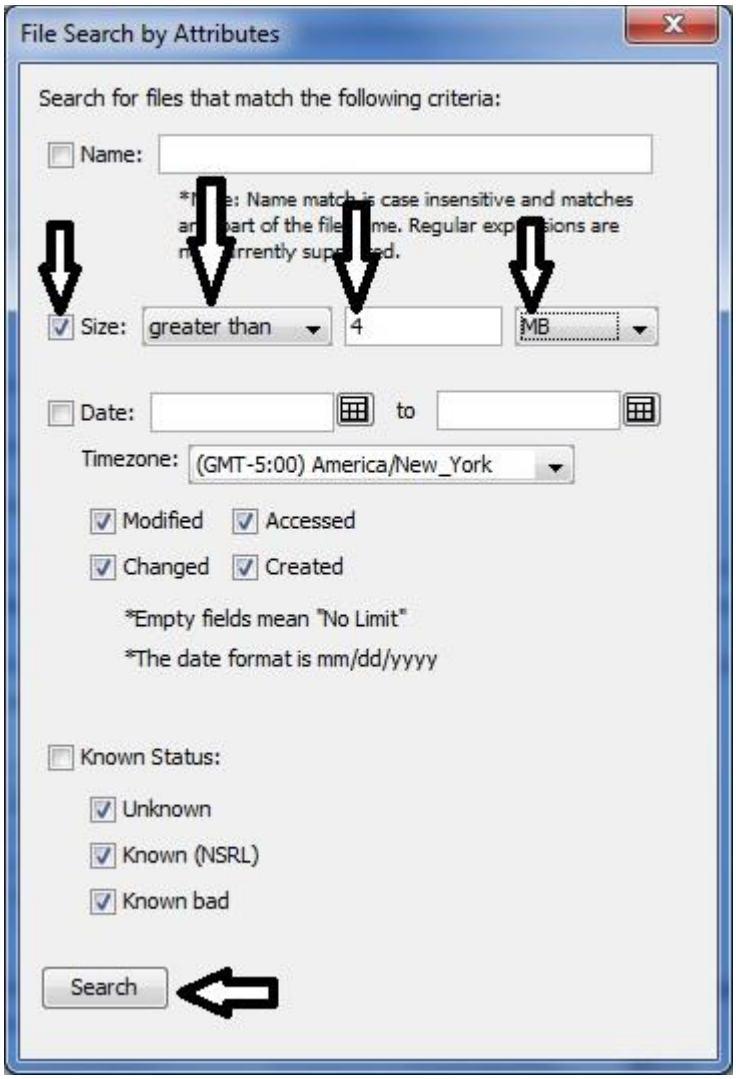
You have completed **Task 4.** Please complete the questionnaire.

Task 5: Identify the largest file
Perform the **Autopsy Launch Sequence** above as needed. **Task 5** identifies files by size. Enabling the **File Search by Attribute** to prompt the **Table View** to present/identify files/directories by size.

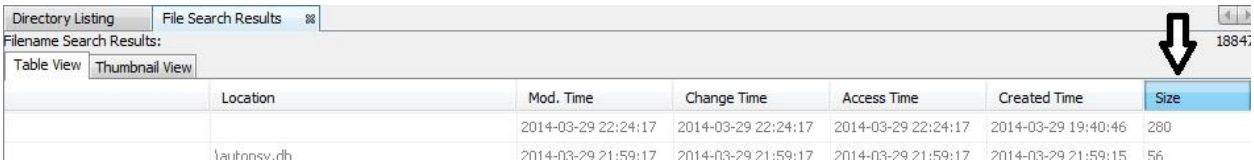
1. In the **Data Explorer** right click **autopsy.db** image and, when the dialog box pops up, select the **File Search by Attributes** option.



In the **File Search by Attribute** dialog box, click the **Size** checkbox, change the pattern to **greater than,** and select the size to be greater than **4 MB**. Now click **Search**.



Once the results are presented in the **Table View,** click **Size** twice to order files/directories largest to smallest. This may take a minute.



Now scroll to find the largest file in the **Table View**. Can you locate **Kittie.jpg**?

You have completed **Task 5:** the **Directory Tree** display portion of the study**.** Please complete the questionnaire.

References

AccessData's Forensic Toolkit (FTK). (2013). Retrieved July 28, 2013, from http://www.accessdata.com/products/digital-forensics/ftk/

Ahn, J., Plaisant, C., & Shneiderman, B. (2011). A task taxonomy of network evolution analysis. Retrieved May 18, 2011, from http://hcil.cs.umd.edu/trs/2011-09/2011-09.pdf

Al-Zaidy, R., Fung, B., & Youssef, A. (2011). Towards discovering criminal communities from textual data. *Proceedings of the 2011 ACM Symposium on Applied Computing*, 172-177. New York, NY: ACM. doi:10.1145/1982185.1982225

Araujo, P., & Frøyland, L. (2007) Statistical power and analytical quantification. *Chromatography, 847*(2), 305-308. doi:10.1016/j.jchromb.2006.10.002

Atkinson, R., & Shiffrin, R. (1971). The control processes of short-term memory. Retrieved May 18, 2011, from http://suppescorpus.stanford.edu/techreports/IMSSS_173.pdf

Ayers, D. (2009). A second generation computer forensic analysis system. *Digital Investigation, 6*(1), S34-S42. Montreal, Canada. doi:10.1016/j.diin.2009.06.013

Barbara, J. (2013). Starting a career in digital forensics: Part 1. In *Digital Forensic Investigator*. Retrieved April 1, 2013, from http://www.dfinews.com/print/7703

Beebe, N., Clark, J., Dietrich, G., Ko, M., & Ko, D. (2011). Post-retrieval search hit clustering to improve information retrieval effectiveness: Two digital forensics case studies. *Decision Support Systems, 51*(4), 732-744. doi:10.1016/j.dss.2011.01.009

Bhat, V., Rao, P., Abhilash, R., Shenoy, P., Venugopal, K., & Patnaik, L. (2010). A novel data generation approach for digital forensic application in data mining. *Machine Learning and Computing (ICMLC), 2010 Second International Conference*, 86-90. Bangalore, India. doi:10.1109/ICMLC.2010.24

Çakir, A. (1997). International ergonomic HCI standards. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of human-computer interaction, 2*(407-420). Amsterdam. doi:10.1016/B978-044481862-1.50084-4

Card, S., & Mackinlay, J. (1997). The structure of the information visualization design space. *Information Visualization, IEEE Symposium*, 92-99. Oct. IEEE; Phoenix, AZ. doi:10.1109/INFVIS.1997.636792

Card, S., Mackinlay, J., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think.* 1-35. San Francisco, CA; Morgan Kaufmann Publishers, Inc.

Carrier, B. (2003). Defining digital forensic examination and analysis tools using abstraction layers. *International Journal of Digital Evidence, 1*(4), 1-12. Retrieved March 2, 2013, from http://www.utica.edu/academic/institutes/ecii/publications/articles/A04C3F91-AFBB-FC13-4A2E0F13203BA980.pdf

Carroll, J. (2003). *HCI Models, theories, and frameworks: Toward a multidisciplinary science.* San Francisco, CA; Morgan Kaufmann Publishers, Inc.

Chi, E. (2000). A taxonomy of visualization techniques using the data state reference model. *Proceedings of the IEEE Symposium on Information Visualization 2000 (INFOVIS '00)*, 69-76. Retrieved October 15, 2011, from the IEEE Computer Society Digital Library. Washington, DC.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York, NY: Psychology Press.

Comer, D., Gries, D., Mulder, M., Tucker, A., Turner, A., & Young, P. (1989). Computing as a discipline. *Communication of the ACM* 32, 9-23. New York, NY: ACM. doi:10.1145/63238.63239

Conti, G. (2007). *Security data visualization: Graphical techniques for network analysis.* San Francisco, CA: No Scratch Press, Inc.

Cooper, P., Finley, G., & Kaskenpalo, P. (2010). Towards standards in digital forensics education. *Proceedings of the 2010 ITiCSE working group reports (ITiCSE-WGR '10)*, 87-95. New York, NY: ACM. doi:10.1145/1971681.1971688

Creswell, J. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Los Angeles, CA: Sage.

Digital Corpora. (2011). Retrieved February 11, 2012, from http://digitalcorpora.org/corpora/files

Digital Forensic Research Workshop (DFRWS) Technical Report. (2001). A road map for digital forensic research report. First Digital Forensic Research Workshop (DFRWS). Utica, NY: DFRWS. Retrieved February 11, 2012, from http://www.dfrws.org/2001/dfrws-rm-final.pdf

Ebert, A., Gershon, N., & van der Veer, G. (2012). *Human-Computer Interaction. KI-Künstliche Intelligenz, 26*(2), 121-126. Springer-Verlag. doi:10.1007/s13218-012-0174-7

Eng, J. (2003). Sample size estimation: How many individuals should be studied? *Radiology*, *227*(2): 309-313. doi:10.1148/radiol.2272012051

EnCase®. (2011). Retrieved February 11, 2012, from http://www.guidancesoftware.com/forensic.htm

Endicott-Popovsky, B., & Frincke, D. (2007). Embedding Hercule Poirot in networks: Addressing inefficiencies in digital forensic investigations. *Proceedings of the 3rd International Conference on Foundations of Augmented Cognition (FAC'07)*, 364-372. Retrieved August 14, 2011, from SpringerLink, Berlin, Heidelberg.

Fetterman, D. M. (2010). *Ethnography: Step-by-step* (3rd ed.). Thousand Oaks, CA: Sage.

Ferster, B. (2013). *Interactive visualization.* Cambridge, MA: The MIT Press.

Fu, X., Hong, S., Nikolov, N., Shen, X., Wu, Y., & Xu, K. (2007). Visualization and analysis of email networks. *Proceedings of the 6th International Asia-Pacific Symposium on Visualization*, 1-8. Sydney, Australia. doi:10.1109/APVIS.2007.329302

Garfinkel, S. (2010). Digital forensics research: The next 10 years. *Digital Investigation, 7*(S64-S73). Portland, OR. doi:10.1016/j.diin.2010.05.009

Glatz, E. (2010). Visualizing host traffic through graphs. *International Symposium on Visualization for Cyber Security (VizSec '10)*, 58-63. ACM, New York, NY. doi:10.1145/1850795.1850802

Hansen, D., Shneiderman, B., & Smith, M. (2010). *Analyzing social media networks with NodeXL.* Burlington, MA: Morgan Kaufmann.

Hargreaves, C., & Patterson, J. (2012). An automated timeline reconstruction approach for digital forensic investigations. *Digital Investigation, 9* (Supplement), S69-S79. doi:10.1016/j.diin.2012.05.006

Heer, J., Card, S., & Landay, J. (2005). Prefuse: A toolkit for interactive information visualization. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, 421-430. ACM: New York, NY. doi:10.1145/1054972.1055031

Hoelz, B., & Ralha, C. (2013). A framework for semantic annotation of digital evidence. *Proceedings of the 28th Annual ACM Symposium on Applied Computing (SAC '13)*, 1966-1971. ACM, New York, NY. doi:10.1145/2480362.2480729

Huang, M., & Eades, P. (1998). A fully animated interactive system for clustering and navigating huge graphs. *Graph Drawing*, 374-383. Berlin, Heidelberg: Springer-Verlag. doi:10.1007/3-540-37623-2_29

Huang, X., Eades, P., & Lai, W. (2005). A framework of filtering, clustering and dynamic layout graphs for visualization. *Australasian Conference on Computer Science (ACSC '05)*, 38, 87-96. Retrieved October 11, 2011. Darlinghurst, Australia: ACM.

Huang, W., Eades, P., & Hong, S. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization, 8*(3), 139-152. doi:10.1057/ivs.2009.10

Huang, M., Zhang, J., Nguyen, Q., & Wang, J. (2011). Visual clustering of spam emails for DDoS analysis. *Proceedings of the 15th International Conference on Information Visualization*, 65-72. doi:10.1109/IV.2011.41

Hughes, J., King, V., Rodden, T., & Andersen, H. (1995). The role of ethnography in interactive systems design. *Magazine Interactions, 2*(2), 56-65. doi:10.1145/205350.205358

Ieong, R., & Leung, H. (2007). Deriving case-specific live forensics investigation procedures from FORZA. New York, NY: ACM. doi:10.1145/1244002.1244049

Inselberg, A. & Dimsdale, B. (1990). Parallel coordinates: A tool for visualizing multi-dimensional geometry. *Proceedings of the First IEEE Conference on Visualization*, 361-378. San Francisco, CA. doi: 10.1109/VISUAL.1990.146402

John, J. (2012). Digital forensics and preservation. DPC Technology Watch Report 12-3 November 2012. Retrieved September 14, 2013, from http://www.dpconline.org/component/docman/doc_download/810-dpctw12-03pdf-dpctw12-03pdf

Jankun-Kelly, T., Franck, J., Wilson, D., Carver, J., Dampier, D., & Swan, J. (2008). Show me how you see: Lessons from studying computer forensics experts for visualization. *Visualization for Computer Security Lecture Notes in Computer Science*, 5210, 80-86. doi:10.1007/978-3-540-85933-8_8

Jones, K. (2008). Visual computer forensic analysis. Presented at the Black Hat 2008 Conference. Las Vegas, NV. Retrieved February 2, 2013, from http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202428248638&Visual_Computer_Forensic_Analysis&slreturn=20130321120229

Keim, D. (2000). Designing pixel-oriented visualization techniques: Theory and applications. *Visualization and computer graphics, IEEE Transactions on visualization and computer graphics, 6*(1), 59-78. doi: 10.1109/2945.841121

Keim, D. (2002). Information visualization and visual data mining. *Visualization and computer graphics, IEEE Transactions on visualization and computer graphics, 8*(1), 1-8. doi:10.1109/2945.981847

Kessler, G., & Ramsay, J. (2014) A Proposed Curriculum in Cybersecurity Education Targeting Homeland Security Students. *47th Hawaii International Conference on System Sciences (HICSS, '14)*, 4932-4937. Waikoloa Village, HI. doi: 10.1109/HICSS.2014.605

Krasser, S., Conti, G., Grizzard, J., Gribschaw, J., & Owen, H. (2005). Real-time and forensic network data analysis using animated and coordinated visualization. Information Assurance Workshop, IAW '05. *Proceedings from the annual IEEE SMC*, 42- 49. West Point, NY. doi:10.1109/IAW.2005.1495932

Kruse, W., & Heiser, J. (2002). *Computer forensics incident response essentials.* Indianapolis, IN: Addison-Wesley, Pearson Education.

Mislan, R., Casey, E., & Kessler, G. (2010). The growing need for on-scene triage of mobile devices. *Digital Investigation, 6*(3–4), 112-124. doi: 10.1016/j.diin.2010.03.001

Mohay, G., Anderson, A., Collie, B., McKemmish, R., & de Vel, O. (2003). *Computer and Intrusion Forensics.* Norwood, MA: Artech House, Inc.

National Institute of Justice Special Report. (2004). Forensic examination of digital evidence: A guide for law enforcement. Retrieved September 26, 2012, from http://www.nij.gov/nij/pubs-sum/199408.htm

Neufeld, D. (2010). Understanding cybercrime. *IEEE Hawaii International Conference on System Sciences*, Koloa, Kauai, HI, 1-10. doi:10.1109/HICSS.2010.417

Olsson, J., & Boldt, M. (2009). Computer forensic timeline visualization tool. *Digital Investigation, 6*, S78-S87. doi:10.1016/j.diin.2009.06.008

Ormerod, T., Mariani, J., Morley, J., Rodden, T., Crabtree, A., Mathrick, J., . . . Lewis, K. (2005). Mixing research methods in HCI: Ethnography meets experimentation in image browser design. Engineering Human Computer Interaction and Interactive Systems Lecture Notes in *Computer Science* 3425, 112-128. Springer Berlin Heidelberg. doi:10.1007/11431879_7

Osborne, G., & Turnbull, B. (2009). Enhancing computer forensics investigation through visualization and data exploitation. IEEE International Conference: *Availability, Reliability and Security*, Fukuoka, Japan. doi:10.1109/ARES.2009.120

Osborne, G., Turnbull, B., & Slay, J. (2010). The 'Explore, Investigate and Correlate' (EIC) conceptual framework for digital forensics information visualization. *Proceedings of the 2010 International Conference on Availability, Reliability and Security*, 629-634. doi:10.1109/ARES.2010.74

Osborne, G., Turnbull, B., & Slay, J. (2012). Development of InfoVis software for digital forensics. *Computer Software and Applications Conference Workshops (COMPSACW)*, 213-217. Izmir, Turkey. doi:10.1109/COMPSACW.2012.47

Paas, F. (1992) Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429-434. doi:10.1037/0022-0663.84.4.429

Paas, F., & van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive load approach. *Journal of Educational Psychology*, 86, 122-133.

Paas, F., Tuovinen, J., Tabbers, H., & van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71. doi:10.1207/S15326985EP3801_8

Palomo, E., North, J., Elizondo, D., Luque, R., & Watson, T. (2012). Application of growing hierarchical SOM for visualization of network forensics traffic data. *Neural Networks.* doi:10.1016/j.neunet.2012.02.021

Peisert, S., Bishop, M., Karin, S., & Marzullo, K. (2007). Toward models for forensic analysis. *Systematic Approaches to Digital Forensic Engineering, SADFE 2007*, 3-15. Bell Harbor, WA. doi:10.1109/SADFE.2007.23

Peisert, S., Bishop, M., & Marzullo, K. (2008). Computer forensics in forensis. *SIGOPS Operating Systems, (42)*3, 112-122. doi:10.1145/1368506.1368521

Pendergast, G. (2010). Getting started in digital forensics: Do you have what it takes? Retrieved June 1, 2013, from http://computer-forensics.sans.org/blog/2010/08/20/getting-started-digital-forensics-what-takes/

Prajapati, B., Dunne, M., & Armstrong, R. (2010). Sample size estimation and statistical power analyses. *Optometry Today*, *16*(07). Retrieved November 22, 2013, from http://dv.fosjc.unesp.br/ivan/downloads/Aulas%20em%20PDF*GPower_by_Prajapati-et-al.--2010-.pdf

Ray, D., & Bradford, P. (2007). Models of models: Digital forensics and domain-specific languages. Retrieved February 21, 2013, from http://www.ioc.ornl.gov/csiirw/07/abstracts/Bradford-Abstract.pdf

Regional Computer Forensics Laboratory (RCFL). (2013). The RCFL program's annual report for fiscal year 2012. Retrieved August 6, 2014, from http://www.rcfl.gov/downloads/documents/RCFL_Nat_Annual12.pdf

Peterson, G., Raines, R., & Baldwin, R. (2007). Graduate Digital Forensics Education at the Air Force Institute of Technology. *Proceedings of the 40th Annual Hawaii International Conference on System International Conference on System Sciences (HICSS ' 07)*, 264-271. Waikoloa, HI: IEEE. doi: 10.1109/HICSS.2007.240

Richard, G., & Roussev, V. (2006). Next-generation digital forensics. *Communication of the ACM*, *49*(2), 76-80. New York, NY: ACM. doi:10.1145/1113034.1113074

Rose, A., Shneiderman, B., & Plaisant, C. (1995). An applied ethnographic method for redesigning user interfaces. *Processes, Practices, Methods & Techniques (DIS '95)*, 115-122. New York, NY: ACM. doi:10.1145/225434.225447

Saltzer, J., & Schroeder, M. (1975). The protection of information in computer systems. *Operating System Principles*, 1278–1308. doi:10.1109/PROC.1975.9939

Saraiya, P., North, C., & Duca, K. (2010). Comparing benchmark task and insight evaluation methods on time series graph visualizations. *Proceedings of the 2010 Workshop: Beyond time and errors: Novel evaluation methods for Information Visualization (BELIV'10)*, 55-62. New York, NY: ACM. doi:10.1145/2110192.2110201

Schrenk, G., & Poisel, R. (2011). A discussion of visualization techniques for the analysis of digital evidence. *Proceedings from the International Conference on Availability, Reliability and Security (ARES)*, 758-763. Vienna, Austria. doi:10.1109/ARES.2011.119

Selamat, S., Sahib, S., Hafeizah, N., Yusof, R., & Abdollah, M. (2013) A forensic traceability index in digital forensic investigation. *Journal of Information Security*, *4*(1), 19-32. doi:10.4236/jis.2013.41004

SequoiaView. (2014). Retrieved September 6, 2014, from http://sequoiaview.en.softonic.com/

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Visual Languages, IEEE Symposium*, 336-343. Boulder, CO. doi:10.1109/VL.1996.545307

Shneiderman, B. (2008). Extreme visualization: Squeezing a billion records into a million pixels. *Proceedings of the 2008 ACM International Conference on Management of Data (SIGMOD '08)*, 3-12. New York, NY: ACM. doi:10.1145/1376616.1376618

Shneiderman, B., Dunne, C., Sharma, P., & Wang, P. (2012). Innovation trajectories for information visualizations: Comparing treemaps, cone trees, and hyperbolic trees. *Information Visualization, 11*(2), 87-105. doi:10.1177/1473871611424815

Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. *Novel Evaluation Methods for Information Visualization (BELIV '06)*, 1-7. New York, NY: ACM. doi:10.1145/1168149.1168158

Shneiderman, B., Plaisant, C., Cohen, M., & Jacobs, S. (2010). *Designing the user interface: Strategies for effective human-computer interaction* (5th ed.). USA: Addison-Wesley Publishing Company.

Sleuth Kit. (2012). Retrieved September 6, 2014, from http://www.Sleuth Kit.org/

Stanovich, K., & West, R. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences 23*(5), 645–726.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285. Retrieved February 25, 2012, from http://csjarchive.cogsci.rpi.edu/1988v12/i02/p0257p0285/MAIN.PDF

Toker, D., Conati, C., Steichen, B., & Carenini, G. (2013). Individual user characteristics and information visualization: Connecting the dots through eye tracking. *Human Factors in Computing Systems*. Paris, France.

Trček, D., Abie, H., Skomedal, A., & Starc, I. (2010). Advanced framework for digital forensic technologies and procedures. *Journal of Forensic Sciences, 55*(6), 1471–1480. doi:10.1111/j.1556-4029.2010.01528.x

Tufte, E. (1990). *Envisioning information.* Cheshire, CT: Graphics Press.

Tufte, E. (2000). *Visual explanations.* Cheshire, CT: Graphics Press.

U.S. Bureau of Labor Statistics. (2012). *Occupational Outlook Handbook.* Retrieved September 9, 2012, from http://www.bls.gov/ooh/Protective-Service/Private-detectives-and-investigators.htm

U.S. Department of Defense. (2012). Information Assurance Workforce Improvement Program (DoD 8570.01-M, Change 3). Retrieved September 30, 2012, from http://www.dtic.mil/whs/directives/corres/pdf/857001m.pdf

Urdan, T. (2010). *Statistics in plain English* (3rd ed.). New York, NY: Routledge.

Vidmar, G. (2007). Statistically Sound Distribution Plots in Excel. Retrieved September 13, 2012, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.127.1969

Wang, S. (2007). Measures of retaining digital evidence to prosecute computer-based cyber-crimes. *Computer Standards & Interfaces, 29*(2), 216-223. doi:10.1016/j.csi.2006.03.008