

**VISUAL ANALYTICS FOR AVIATION SAFETY:
A COLLABORATIVE APPROACH TO SENSEMAKING**

by

Andrew Wade
B.A., University of British Columbia, 2007

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in the
School of Interactive Arts and Technology
Faculty of Communication, Art and Technology

© Andrew Wade 2011
SIMON FRASER UNIVERSITY
Fall 2011

All rights reserved. However, in accordance with the *Copyright Act of Canada*, this work may be reproduced, without authorization, under the conditions for "Fair Dealing." Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

APPROVAL

Name: Andrew Wade
Degree: Master of Science
Title of Thesis: Visual Analytics for Aviation Safety:
A Collaborative Approach to Sensemaking

Examining Committee:

Chair:

Dr. Marek Hatala
Associate Professor

Dr. Brian Fisher
Senior Supervisor
Associate Professor

Dr. Janet McCracken
Supervisor
Adjunct Professor

Dr. John Dill
External Examiner
Professor Emeritus
School of Interactive Arts and Technology

Date Defended/Approved: August 22, 2011

Partial Copyright Licence



The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the "Institutional Repository" link of the SFU Library website (www.lib.sfu.ca) at <http://summit/sfu.ca> and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Simon Fraser University
Burnaby, BC, Canada

ABSTRACT

Visual analytics, the “science of analytical reasoning facilitated by interactive visual interfaces”, is more than just visualization. Understanding the human reasoning process is essential for designing effective visualization tools and providing correct analyses. This thesis describes the evolution, application and evaluation of a new method for studying analytical reasoning that we have labeled paired analysis. Paired analysis combines subject matter experts (SMEs) and tool experts (TE) in an analytic dyad, here used to investigate aircraft maintenance and safety data. The method was developed and evaluated using interviews, pilot studies and analytic sessions during an internship at the Boeing Company. By enabling a collaborative approach to sensemaking that can be captured by researchers, paired analysis yielded rich data on human analytical reasoning that can be used to support analytic tool development and analyst training.

Keywords: visual analytics, paired analysis, sensemaking, boeing, collaborative analysis

What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.

Herbert Simon (1971)

ACKNOWLEDGEMENTS

I would like to thank my senior supervisor, Brian Fisher, for his support during my trek down the thesis-writing road, a trek that has no doubt been long and winding. This includes helping with funding and conferences as well as advice and encouragement. I would also like to thank Janet McCracken for her assistance, editing and patience. Special thanks to John Dill for his backing and encouragement and for agreeing to provide his insight and expertise at my defense.

I would like to extend my gratitude to The Boeing Company for their ongoing support and collaboration, both for myself and for our school and lab. Boeing's funding, subject matter expertise and data have been crucial to this project and many others at our lab and have helped to advance visual analytics research.

At Boeing, I would like to personally acknowledge and thank Dave Kasik and Roger Nicholson for their time, expertise, guidance and friendship.

I would also like to thank my family and friends for their support and heckling, both of which helped me to finish.

TABLE OF CONTENTS

Approval.....	ii
Abstract.....	iii
Dedication.....	iv
Acknowledgements.....	v
Table of Contents.....	vi
List of Figures.....	viii
List of Tables.....	ix
1: Introduction.....	1
1.1 Research Area.....	1
1.2 Research Objectives.....	4
1.3 Thesis Organization.....	5
2: Literature Review.....	6
2.1 Sensemaking: The Foundation for Analysis.....	6
2.2 Cognitive Limitations and Short-Term Memory.....	11
2.3 Visual Analytics.....	14
2.3.1 Text Analysis.....	18
2.3.2 Numerical Analysis.....	21
2.3.3 Geospatial Analysis.....	22
2.4 Evaluation of Visual Analytics Tools.....	24
2.4.1 Analytical Reasoning.....	27
2.4.2 Participant Observation.....	33
2.4.3 Protocol Analysis.....	37
2.5 Paired Analysis.....	42
2.5.1 Overview.....	42
2.5.2 Bringing Tacit Knowledge to Light.....	46
2.5.3 Coordinating Joint Action.....	47
3: Aeroinfo: Interviews and Pilot Studies.....	49
3.1 Introduction.....	49
3.2 Background: Uncertainty.....	49
3.3 Background: Aircraft Maintenance and Maintenance Planners.....	50
3.4 Pilot Study 1: Interviews at Aeroinfo.....	52
3.4.1 Pilot Study 1: Overview of Findings.....	54
3.4.2 Pilot Study 1 Findings: Aircraft Components.....	54
3.4.3 Pilot Study 1 Findings: Managing People.....	58
3.4.4 Pilot Study 1 Findings: Reliability of Data Sources.....	58
3.4.5 Pilot Study 1: Discussion.....	60
3.5 Pilot Study 2: Paired Analysis at Aeroinfo.....	60
3.5.1 Pilot Study 2: Paired Analysis Sessions.....	61
3.5.2 Pilot Study 2: Findings and Discussion.....	62

4: Paired Analysis for Aviation Safety	64
4.1 Introduction	64
4.2 Exploratory Interviews	65
4.3 Boeing: Paired Analysis Sessions	68
4.3.1 Overview	68
4.3.2 Choosing a Task.....	69
4.3.3 Choosing a Dataset	75
4.3.4 Choosing a Tool	78
4.3.5 Data Cleaning and Preparation.....	79
4.3.6 Initiating Sessions.....	81
4.3.7 Video Recording	82
5: Boeing: Results and Analysis	86
5.1 Overview	86
5.2 Choosing Sessions for Further Analysis	87
5.3 Analysis	90
5.3.1 Coding and Memos	90
5.3.2 Coding: Interactions.....	93
5.3.3 Coding: Sensemaking.....	101
6: Discussion.....	112
6.1 Summary and Contributions	112
Appendices.....	118
Appendix A: Aeroinfo Focus Structure.....	118
Appendix B: Boeing Safety Engineer Focus Structure.....	122
Appendix C: Boeing Internship Task Sheet	126
Appendix D: Collected Data Description.....	130
References.....	140

LIST OF FIGURES

Figure 1: The Sensemaking Loop for Analysts	8
Figure 2: Sensemaking Loop: Later Steps.....	11
Figure 3: Dot array	15
Figure 4: Line Orientation	15
Figure 5: Age of Husband vs. Wife (Scatterplot).....	17
Figure 6: Sensemaking Loop: Early Steps.....	18
Figure 7: IN-SPIRE's themeview	19
Figure 8: Starlight Software	20
Figure 9: Tableau	22
Figure 10: Geotime: Tracking objects through space and time	23
Figure 11: Relation between failure of components and types of maintenance.....	55
Figure 12: Failure rate patterns for individual components	57
Figure 13: Five Stages of Response to Novel Situations	73
Figure 14: Pilot Narrative Description of Incident with Important Words Highlighted	75
Figure 15: Subset of Boeing Safety Databases	77
Figure 16: Steps for Data Preparation	80
Figure 17: Camera Setup	84
Figure 18: Paired Analysis Session Video Screenshot	85
Figure 19: Coding Transcripts with Atlas.ti.....	92
Figure 20: Interaction Code Occurrences	95
Figure 21: Code Family - SME/TE Interactions.....	99
Figure 22: Code Family – TE/Tool Interactions.....	100
Figure 23: Total Sensemaking Codes by Session	105
Figure 24: Bottom Up Process Distribution over Sessions Stages.....	108
Figure 25: Top Down Process Distribution over Session Stages	110

LIST OF TABLES

Table 1:	Age of Husband vs. Wife	16
Table 2:	Paired Analysis Session Details	90

1: INTRODUCTION

1.1 Research Area

Stored digital data continues to increase at an exponential rate. A large portion of this increasing digital data is represented numerically and by strings of text. It can be structured or unstructured, and is often stored in many different locations and formats. As I will further describe, human memory and analysis capacity are severely limited with regard to large amounts of information. The field of Visual Analytics (VA) was born out of the need to combat the increasing amount of information surrounding analysis.

Many industries and government agencies are becoming increasingly overwhelmed by data. As digital data increases, the ability to manage and analyze it becomes a function of time, resulting in valuable resources being diverted away from other aspects of businesses and government. The September 11th attacks in 2001 highlighted the need for better intelligence analysis in the United States (Kean et al., 2004). In 2004 the US Department of Homeland Security helped create the National Visualization and Analytics Center (NVAC) “with the goal of helping to counter future terrorist attacks in the United States and around the globe” (Thomas & Cook, 2005). The center’s agenda was based on research showing the power of visualizations in aiding with analysis. Its goal was to initiate VA research that would ultimately assist with large-scale data analysis.

As the number of increasing VA tools suggest, many organizations and researchers have begun to embrace the rationale that visualizations have the potential to aid analysis. However, evaluating the effectiveness of VA solutions is still in its infancy. VA tools are often complex due to the nature of analytic problems. Because of the cost in terms of money and time invested in learning, users want to know that VA systems are actually an improvement over other types of analysis. To provide this assurance, VA tools need to be evaluated for their efficiency.

In addition to data visualization, VA involves understanding analytic reasoning. VA tools target an audience that is constantly immersed in analysis. Catering to people within this analytic culture requires more than just the creation of good visualizations. In order for tools to assist with the complex nature of many analytic domains they must address and incorporate the analyst's reasoning. This involves looking at how practicing analysts perform and work in the real world. In the past few decades there has been considerable research and advancement in the fields of information visualization and computing. However, there is significant room for increasing our understanding of analytical reasoning. Existing techniques for studying analytic reasoning are largely based on ethnographic methods, many of which draw on methods from participant observation (Becker & Geer, 1957; K. M. DeWalt & DeWalt, 2002; Jorgensen, 1989) and protocol analysis (Ericsson & Simon, 1993). While these methods can elicit very useful data about the analytic process, asking people to talk about their actions in real-time during analysis can sometimes hinder the analytic task

(Schooler, Ohlsson, & Brooks, 1993; Trickett, Fu, Schunn, & Trafton, 2000).

These downsides provide an opportunity for the use of new techniques to study analytical reasoning.

This thesis provides an in-depth, descriptive breakdown of paired analysis (Arias-Hernandez, Kaastra, Green, & Fisher, 2011), a method for studying analytic reasoning as well as evaluating tools. Paired analysis draws on many existing ethnographic methods, but differs in that the verbal protocol is an integral part of the analysis involving collaboration between a visual analytics tool expert and a subject matter expert. The thesis begins by breaking down paired analysis to its roots and examining the reasons for its creation as well as why it was used. The thesis goes on to chart how the method was used across two different domains and with different SMEs, explaining lessons learned and how it evolved along the way. Pirolli and Card's (2005) sensemaking model is used as the main model for analytic reasoning throughout the thesis. Paired analysis is examined within the context of this model, and qualitative data collected from paired analysis sessions are coded using sensemaking steps. By applying the sensemaking model in addition to looking at research done using Joint Activity Theory and paired analysis (Arias-Hernandez et al., 2011), this thesis proposes a combination of these methods for future research. The data presented in this thesis serves as a starting point for paired analysis research by highlighting problems, best-use techniques and analysis of the method's effectiveness as an analytical technique and research tool.

1.2 Research Objectives

Paired analysis is a method that has yet to be thoroughly field-tested, and therefore a major goal of this research was to evaluate its usefulness within the context of “real-world” analysis. Ideally, research on analytical reasoning should take place in the analyst’s natural work setting, when the analyst is behaving normally (Zsombok & Klein, 1997). Therefore, paired analysis sessions were conducted with Boeing safety engineers at Boeing offices. Ultimately, the goal was to study analytical reasoning while a subject matter expert and tool expert used visual analytic tools. This was achieved by video and audio recording paired analysis sessions during the internship.

The main research questions guiding this work are:

- What kind of information can the paired analysis method elicit with regard to analytic reasoning and sensemaking during analysis?
- How does the sensemaking model support collaborative analysis?
- To what effect can the paired analysis method help the SME’s learn new tools?

The first question has been addressed by conducting paired analysis sessions in a “real-world” context at Boeing. The second question seeks to examine the collected data within the context of the sensemaking model (Pirolli & Card, 2005). Any contributions or improvements to the model based on analyzing the transcripts during the paired analysis tasks were noted. The third question is related to using this method for tool evaluation and integration into an organization.

1.3 Thesis Organization

Chapter 2 reviews literature in several areas, including the foundations for visual analytics: sensemaking, human cognitive limitations (and strengths) and short-term memory. It then examines visual analytics itself and methods for evaluating VA tools. Finally, a case is made for the use of paired analysis in evaluating VA tools as well as studying analytic reasoning. Chapter 3 presents the methods and results of pilot studies, questionnaires and interviews with Aeroinfo maintenance planners and how these have helped to shape the paired analysis method. Chapter 4 describes the preparation and use of paired analysis with safety engineers at Boeing. Chapter 5 discusses the results and analysis of transcripts from the Boeing paired analysis sessions. Chapter 6 concludes with discussion, a summary of contributions and future work.

2: LITERATURE REVIEW

2.1 Sensemaking: The Foundation for Analysis

Imagine that you work for a major aircraft manufacturer. Your job is to assess whether the windshields on your company's aircraft are strong enough to protect pilots from bird strikes. Your supervisor gives you a base rate for the acceptable number of windshield cracks from bird strikes per thousand flight hours. You are to determine whether aircraft in the fleet have been above or below this rate for the last five years so that if they are in fact above, reinforcements can be added to make the windshield more resistant.

Where would you begin? What are the steps for determining what counts as a bird strike assuming there is no list of "bird strikes cracking the glass"? (There isn't).

Nearly every type of analysis is a goal-oriented activity requiring the processing and manipulation of information. Analysis often involves collecting a number of data points, sorting through them, forming hypotheses, performing calculations and drawing conclusions. "Sensemaking" can be described as the process of searching for a representation and encoding data in that representation to answer task-specific questions (Russell, Stefik, Pirolli, & Card, 1993). Sensemaking is often an iterative process involving many distinct, yet tightly interwoven cognitive processes and external resources. Sensemaking research has focused on breaking down the analysis process into several steps,

and then charting the cognitive processes and external resources needed at each step. One goal of this research is to identify where the bottlenecks exist. Bottlenecks are steps during analysis that put significant strain on cognitive abilities of the analysis, either wasting valuable time, causing people to miss important information, leading to incorrect assumptions, or all of the above. Ultimately, the goal is to improve overall efficiency by alleviating bottlenecks through the use of new technology and analysis techniques.

When the amount of information an analyst faces begins to increase substantially, attempts to increase efficiency have often focused on information retrieval (Russell et al., 1993). Information retrieval can be described as the search for useful information from large collections of information (Singhal, 2001). However, defining “useful” depends on the task at hand and the specific kind of analysis being performed. When the archiving of recorded information began thousands of years ago, techniques for information retrieval became a necessity. Information retrieval takes up a significant amount of time during any analysis, and its repetitive nature leads to arguments for automating a large portion of the process. However, Russell et al. (1993) argue “speeding retrieval by itself may help very little”. Their argument illustrates that information retrieval is only one step in the entire process.

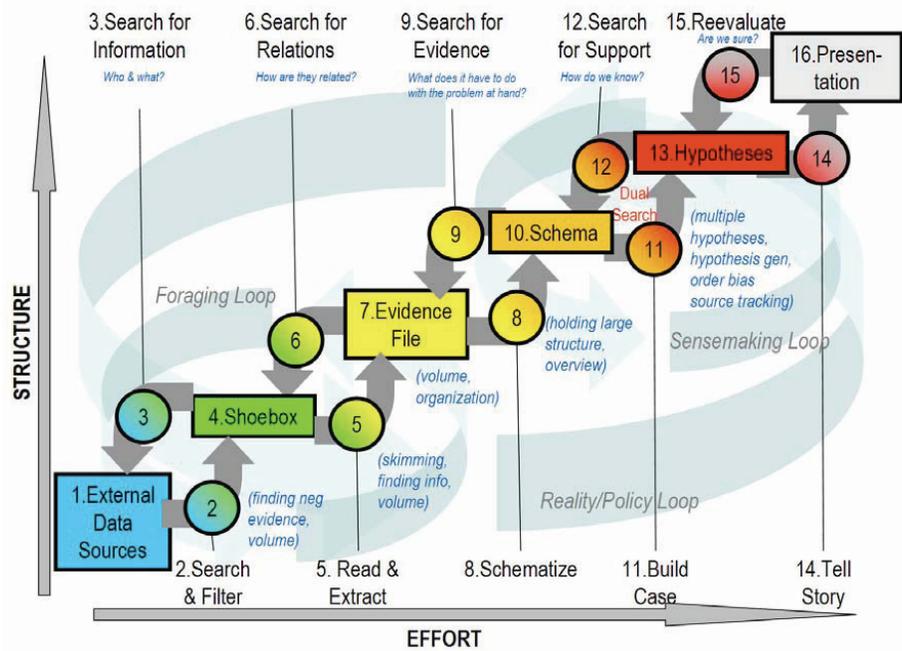


Figure 1: The Sensemaking Loop for Analysts

In Figure 1, Pirolli and Card (2005) have broken down the sensemaking process into a series of loops. The two main loops are the Foraging Loop and Sensemaking Loop. ‘Foraging’ refers to gathering and sorting information, and is often most prevalent during early stages of analysis. These steps are frequently composed of information retrieval tasks. Despite usually having a ‘top-down’, directed goal for analysis, ‘foraging’ steps involve a lot of ‘bottom-up’ cognitive processes (Ruthraff, Remington, & Johnston, 2001). The Sensemaking Loop refers to higher order cognitive functions such as hypothesis generation and matching information to a schema (Pirolli & Card, 2005). Activities within the sensemaking loop occur more frequently during the later stages of analysis, and are often ‘top-down processes’, involving information manipulation to answer goal-specific tasks (Pirolli & Card, 2005).

Within the sensemaking diagram there are numerous arrows pointing both directions. This emphasizes the iterative nature of sensemaking and analysis. There is no direct path from unstructured, raw data to hypotheses and solutions. Instead, there is a constant refining and re-gathering of information.

The loop has helped direct attention to analysis steps that can be improved through the use of technology. For example, intelligent search algorithms such as those present in Google's web search technology have aided steps in the foraging loop by helping to find relevant information quickly (Page, Brin, Motwani, & Winograd, 1998). For collecting and storing information in steps such as the 'shoebox' and 'evidence file', relational databases and SQL queries have proved very useful.

Another significant area that will prove especially useful during sensemaking, particularly at earlier stages, is the field of computational question answering (QA). Answering questions posed in natural language is difficult for a computer due to the many intricacies of human language. However, Ferrucci et al. (2010) succeeded in their grand challenge of designing a computer that could defeat the world's best Jeopardy contestants in real time. This challenge was difficult because of the many steps involved with understanding Jeopardy's complex natural language questions. The computer, "Watson", operates by combining a number of existing QA technologies (parsing, question classification, question decomposition, automatic source acquisition and evaluation, entity and relation detection, logical form generation, and knowledge representation and reasoning) with massive amounts of stored information (millions of documents

from encyclopedias, dictionaries, thesauri, newswire articles, literary works...etc) (Ferrucci et al., 2010).

While most knowledgeable humans only possess a fraction of this information within their brains, Watson's storage is neither the most impressive nor promising advancement. The real improvement in technology comes from understanding the questions posed, a process that involves thousands of algorithms working together simultaneously. While Watson's percentage of correctly answered questions is indeed impressive, the questions Watson answers incorrectly are usually very far from the mark (i.e. "What is Toronto?" when the category was "US cities"). While answering a question incorrectly on Jeopardy results in a monetary penalty, the consequences are relatively small compared to other domains where analysis is carried out. For example, coming to an incorrect military assessment of whether a country has weapons of mass destruction may result in war with said country. Watson's ability to process natural language with incredible accuracy is a sign of things to come. However due to the high-consequence nature of many of the domains where complex analytics are necessary, a human decision maker will still likely make the final call. These types of technology will play a large role in complementing the human decision making process rather than completely replacing it.

Increasing digital representation of information has been both a gift and a curse. More data in a centralized location is beneficial for analytical purposes, but having too much data increases the risk of information overload for the analyst. Information overload can be defined as simply not having enough time or

capability to extract meaningful data from the overwhelming amount of meaningless data. The causes of information overload are rooted in human cognitive limitations such as short-term memory.

2.2 Cognitive Limitations and Short-Term Memory

Due to limitations of cognition and our short-term memory, processing and analyzing complex and vast amounts of new data in a realistic amount of time becomes a challenge.

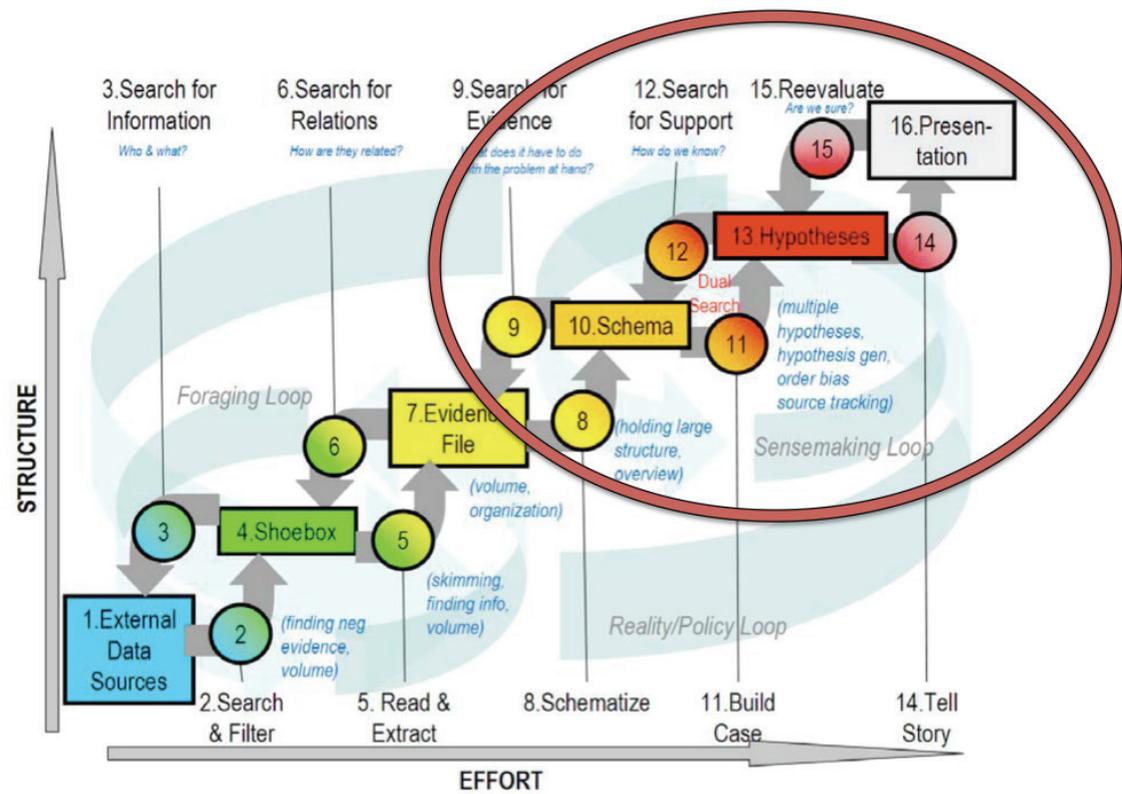


Figure 2: Sensemaking Loop: Later Steps

Later steps in the sensemaking loop (Figure 2) such as creating and evaluating hypotheses rely on recalling information from working memory.

Working memory refers to the part of short-term memory that is concerned with immediate conscious perceptual and linguistic processing. In the mid 1950s, Miller (1956) published a seminal paper in the *Psychological Review* entitled, “The magical number seven, plus or minus two: Some limits on our capacity for processing information”. The paper was concerned with what would later be known as working memory, and highlighted its associated limitations. Miller found that the maximum number of objects an average person could hold in working memory at any given time was seven, plus or minus two (Miller, 1956). However, nearly every “real world” analysis contains many more than just seven objects or data points.

Since Miller’s initial paper there has been a significant amount of research regarding working memory, short-term memory and the capacity of both. Many different models and estimations of capacity have been proposed. Some theories focus on the amount of time that information can be held in memory before decaying, implying that strategies like repetition and chunking must occur to maintain information in working memory (Cavanagh, 1972; Dempster, 1981). Others deal with the limited cognitive resources available to keep the information active and available for processing. Just and Carpenter (2002) state that,

“In addition to its role in storage, working memory can also be viewed as the pool of operational resources that perform the symbolic computations and thereby generate the intermediate and final products.”

From this standpoint, working memory is essential to the later steps in the sensemaking process, acting as a buffer for higher-level analysis and hypothesis

evaluation. Schemas and hypotheses must pass in and out of working memory and therefore are both enabled and limited by it.

The common theme in any theory of working memory is that it is limited. Often, the amount of data needed to perform analysis and evaluate hypotheses is greater than the limit of working memory. In order to compensate for this, we tend to use a combination of external aids and heuristics to keep track of information. By recording and transcribing information onto objects like sticky notes and computer databases, we are in effect extending our working memory to the outside world.

Even when aided by external memory devices such as computers, humans still tend to make errors in judgment when faced with multiple hypotheses and pieces of evidence. Techniques such as Analysis of Competing Hypotheses (ACH) have been developed to aid in this area (Heuer, 1999). ACH is a method designed to help an analyst overcome some of the cognitive biases that occur from making decisions and judgments under uncertainty when there are many possible outcomes (Heuer, 1999; Tversky & Kahneman, 1974). ACH occurs late in the analysis process, typically once an analyst has been able to express two or more mutually exclusive hypotheses. The technique involves listing each possible hypothesis at the top of a column and then listing pieces of evidence in rows. The analyst then looks at each piece of evidence and places a value representing the degree to which that piece of evidence supports or refutes each hypothesis. Once the analyst has gone through each piece of evidence the values in each hypothesis column are added up to give a quantitative measure of

the strength of each hypothesis. This method helps to offset some of the limitations of short-term memory and seeks to overcome biased heuristics by using the external world as an extension of memory, thereby increasing its capacity. Incorporating methods like ACH into visual analytics tools has helped aid with analysis, as in the tool IN-SPIRE (Hetzler & Turner, 2005). IN-SPIRE will be discussed in more detail in the visual analytics section.

Utilizing the strengths of computers is essential for combating information overload and cognitive limitations of human memory. While still underperforming humans in many cognitive areas (vision, dealing with novel situations, conversation) even the most basic computers outperform humans with regard to information storage, computation and processing speed. While these things alone do not yield the answers to complex analytical problems, combining them with approaches such as information visualization has proven to be effective.

2.3 Visual Analytics

Visual analytics is by definition a visual field. Its roots are interdisciplinary, but a large part of what defines the field is its visual component. Humans have evolved a very powerful and efficient visual system capable of detecting patterns and outliers quickly. Recognition of objects within a visual scene is amazingly fast, taking no more than a few hundred milliseconds (Thorpe, Fize, & Marlot, 1996). Our visual system guides many of our everyday tasks, from crossing the street to driving a car. Nearly half of the brain's cortex is involved with the visual system (Hoffman, 2000). Visual intelligence is a form of intelligence that exists separately from rational and emotional intelligence, and often precedes it. It is

often taken for granted, since vision is usually smooth and fluid enough that vision appears effortless.

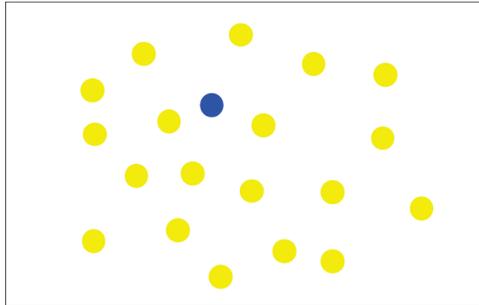


Figure 3: Dot array

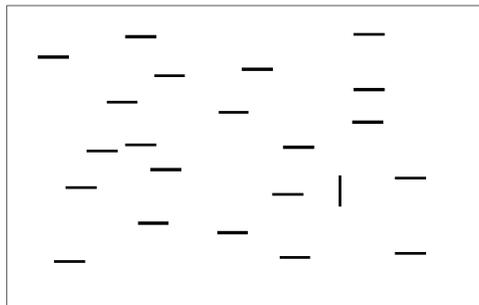


Figure 4: Line Orientation

In Figure 3 there is a series of dots. One of them is different from the others. Our visual system is very good at detecting outliers while grouping similar objects together (Treisman & Gormican, 1988). This principle is evident in the figure. The blue dot immediately stands out among the yellow dots without any conscious effort on our part. The same principle applies to orientation, as in Figure 4. The vertical line stands out among the horizontal. In neither case does the average human have to consciously look at each item one at a time (as a computer would) and decide whether it belonged to the group of similar items or was the outlier. Instead, the grouping occurred automatically and the outlier

“popped out”. This is the case for a number of pre-attentive attributes including color, orientation, spatial location, curvature, size and motion (Julesz, 1984; Rensink & Enns, 1995; Treisman, 1982). These features are processed quickly without the need for explicit attention.

Table 1: Age of Husband vs. Wife

Husband	Wife	Husband	Wife
35	39	56	54
33	36	39	46
59	51	33	33
64	71	78	76
74	68	74	71
71	35	54	55
37	39	39	32
56	52	45	38
34	45	55	60
60	65	43	41
43	41	38	36
34	34	53	55

In Table 1, the ages of a set of husbands and wives are displayed in tabular format. Looking through the data, we see that often the ages differ by only a few years. However, there is one case where there is a large discrepancy in age. In the sixth row on the left, the age of the husband is 71 and the age of the wife is 35. This outlier fails to pop out while in tabular form.

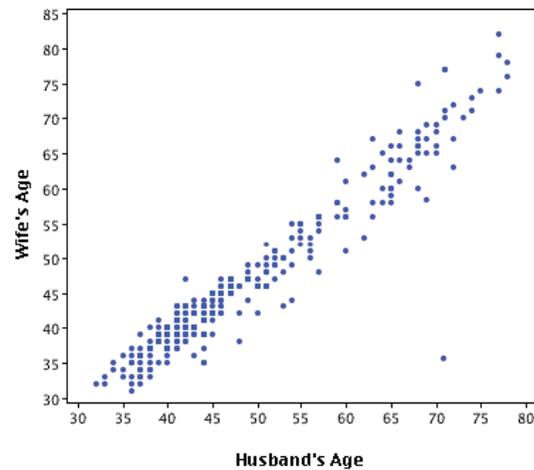


Figure 5 Age of Husband vs. Wife (Scatterplot)

However, the scatterplot in figure 5 displays the same data (plus additional values) regarding age of husbands and wives. In the visual example, the outlier is apparent right away, as there is a single dot far enough away from the main group to be considered a member. These examples illustrate the power of visualizations and exemplify basic principles of visual analytics.

Since the creation of the National Visual Analytics Center (NVAC) in 2004 there have been many visual analytics (VA) tools implemented. A few general categories for VA tools according to the data types they can deal with include:

1. Unstructured and structured text
2. Numerical data
3. Geographic data.

These categories are rarely mutually exclusive. Many times datasets, as well as tools, incorporate and handle multiple types of data. This list does not include all data types, only the most common. The following sections will discuss

some of the issues surrounding different data types and a few of the existing VA tools for handling them. The tools discussed are only a small subsample of tools available.

2.3.1 Text Analysis

In many analyses there is a large corpus of documents to be read and analyzed, leading to a significant portion of analysis time being devoted to understanding the gist of the documents. This tends to occur most early on during sensemaking in the foraging loop (Figure 6), but also happens throughout analysis.

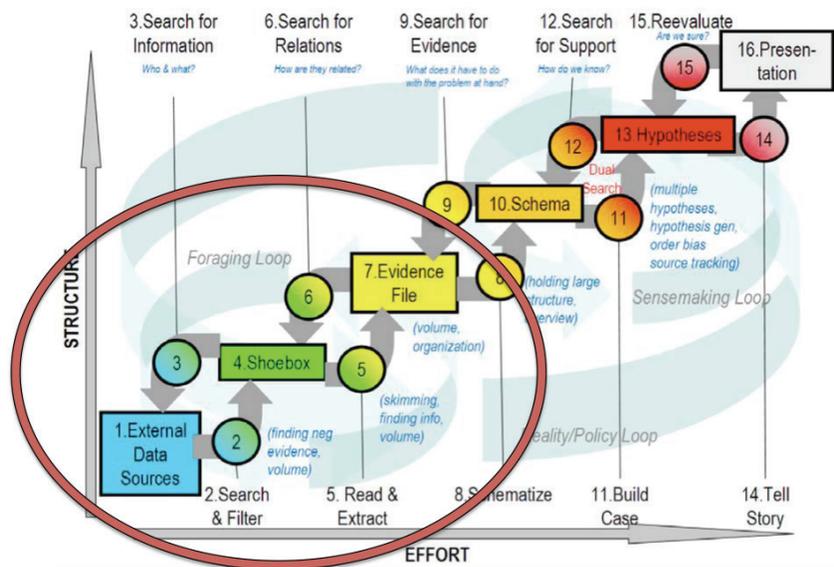


Figure 6: Sensemaking Loop: Early Steps

Due to cognitive limitations (Cavanagh, 1972; Miller, 1956), the amount of raw textual information that can be processed by any analyst for any purpose is severely limited (Wise et al., 1995). Analysts tend to reduce the number of documents to a manageable number, sometimes excluding high value

documents simply due to information overload (Patterson et al., 2001). One of the early tools to emerge from the creation of NVAC was IN-SPIRE, a visual analytics tool born out of the need to understand large bodies of text. The algorithms used for IN-SPIRE examine word frequencies and co-occurrences across a corpus of documents with the goal of extracting important terms (Hetzler & Turner, 2005).

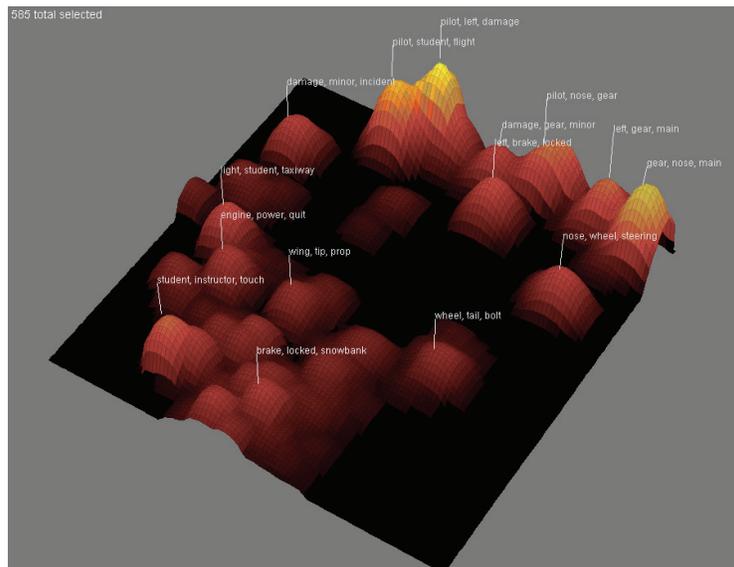


Figure 7: IN-SPIRE's themeview

Figure 7 shows IN-SPIRE's themeview, which groups together documents based on similarity and word occurrence (Hetzler & Turner, 2005). By printing a few of the words that contribute most to each cluster above the respective cluster, the analyst is given a summary of what the documents contain. This type of visualization provides a high-level view of a corpus of documents without reading a single document, and the interactive nature of IN-SPIRE allows the analyst to drill down and filter documents of interest. The rationale behind IN-

SPIRE is to minimize initial reading, and also to support the process of drilling down during analysis (Hetzler & Turner, 2005).

IN-SPIRE also supports later steps in the sensemaking loop through a feature for hypotheses and evidence rating based on the technique of Analysis of Competing Hypotheses (Heuer, 1999). By enabling users to list hypotheses and then rate the reliability of evidence for each, IN-SPIRE helps to counter some of the cognitive biases and memory limitations of the human analyst.

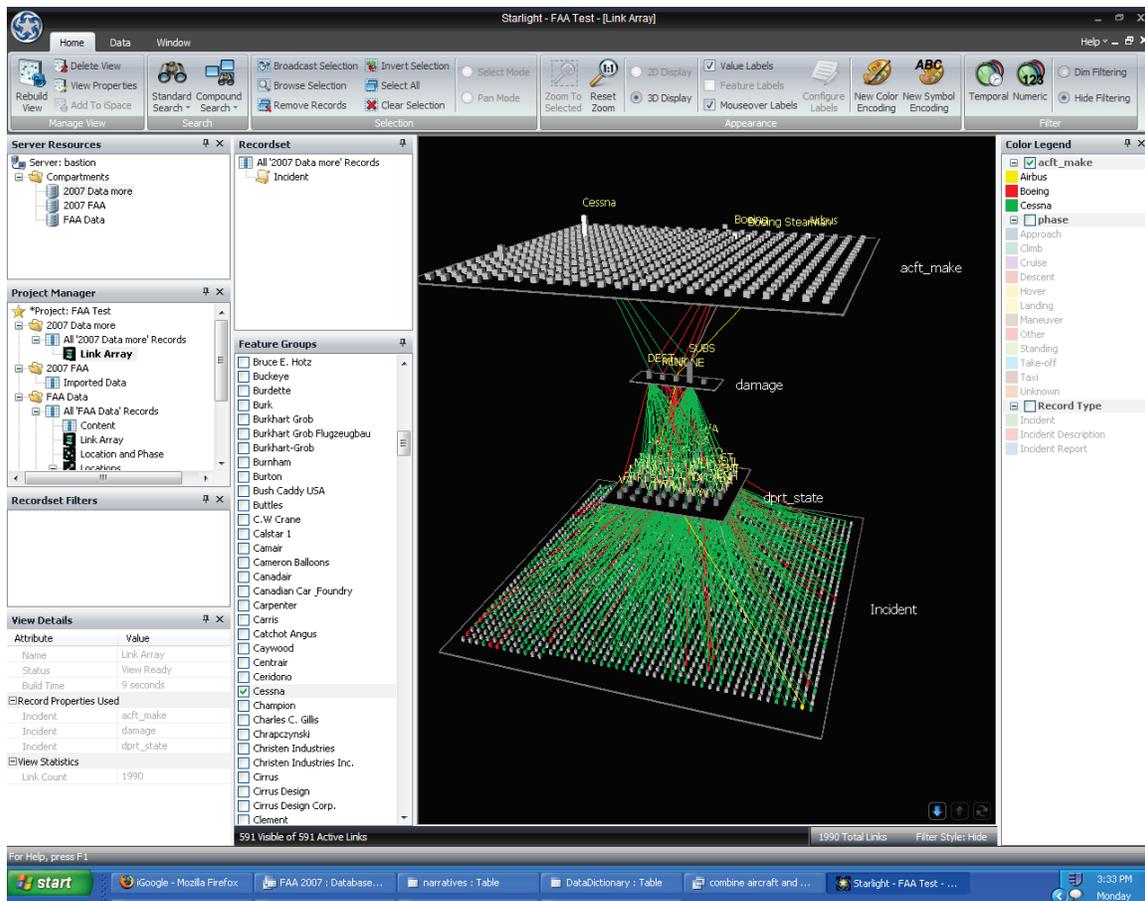


Figure 8: Starlight Software

Figure 8 shows Starlight, another commercial VA tool that can visualize unstructured text (Risch et al., 2002) using a similar method to IN-SPIRE for grouping related documents into clusters and recognizing key words.

Both of these tools employ visualizations and features based on research from cognitive science. However, this is not the case with several other existing tools. In order for existing and future tools to capitalize on human cognitive abilities, additional research into analytical reasoning must continue.

2.3.2 Numerical Analysis

Numbers can represent a vast amount of things, from time and money to measurements from sensors. Current methods for analyzing numbers involve the use of statistics, which are very useful and necessary for performing calculations to learn about the state of a business, when the weather is going to change, or the rate of an epidemic spreading through a population. It is common to present statistical calculations in a data table. However, there are often hundreds of tables representing very detailed information, and depending on the type of analysis, it is often necessary to dig deeper.

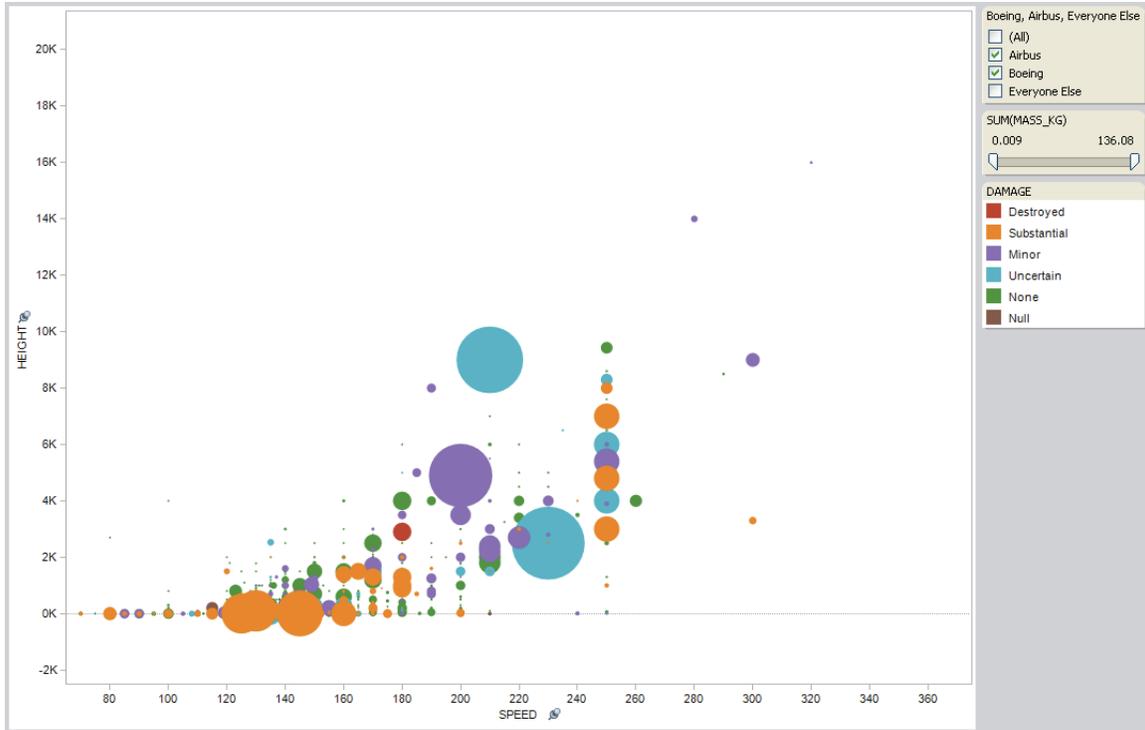


Figure 9: Tableau

There are several visual analytics tools for analyzing numerical and categorical data that are commercially available and many more in development. Some of the most notable include Tableau (Figure 9) (Heer, Mackinlay, Stolte, & Agrawala, 2008), Spotfire (Ahlberg, 1996), Starlight (Risch et al., 2002), and Gapminder (Rosling, 2007). Visualizing numerical and categorical data may be as simple as a histogram or line graph or as complicated as a 3d parallel coordinates plot, such as with Starlight (See Figure 8).

2.3.3 Geospatial Analysis

Every time we look at a map we are looking at a visualization of geospatial data. Geospatial information refers to data with a geographical or terrestrial

component. There is a vast amount of data that falls into this category, including weather data, credit card transactions, crime, and flight patterns.

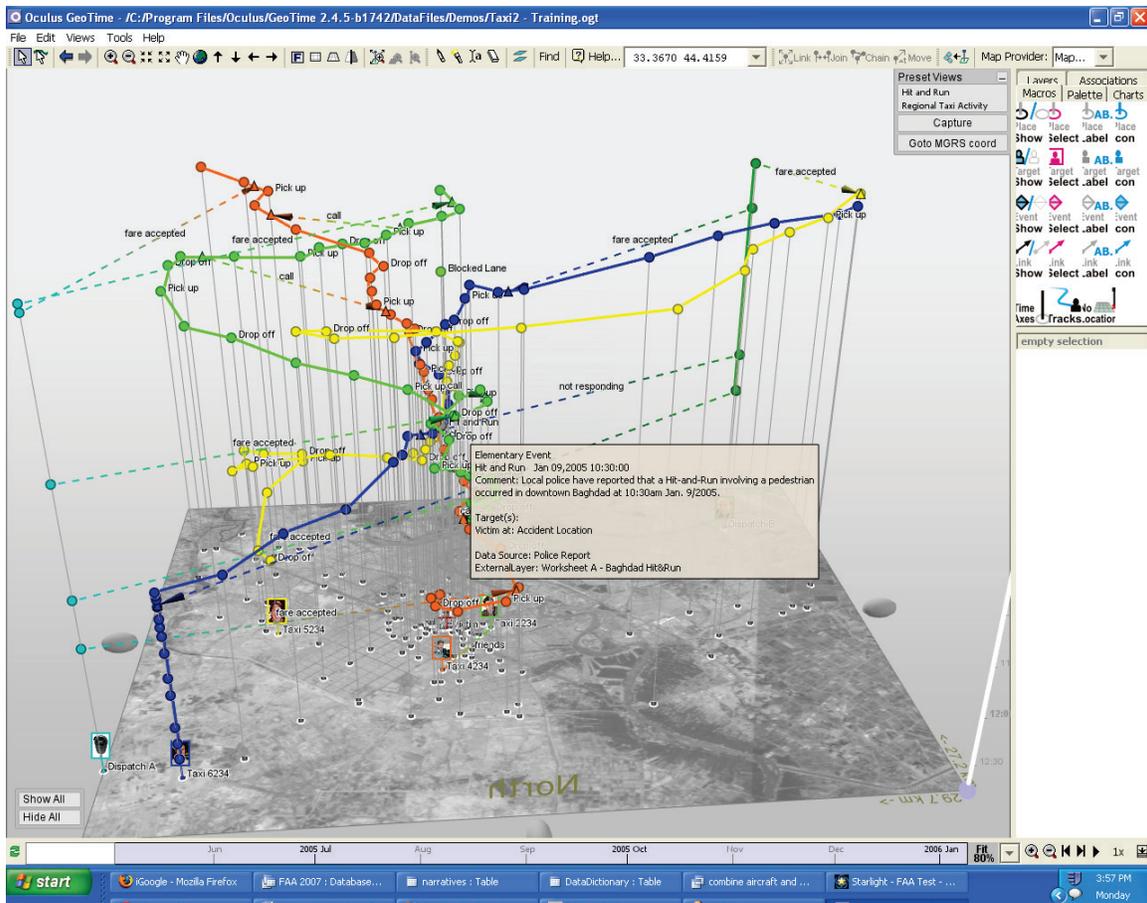


Figure 10: Geotime: Tracking objects through space and time

In addition to the familiar components on a traditional map, visualizing geographic data can also include events and people's movements over time. Tools like Geotime (Kapler & Wright, 2005) allow for a 3d map space, allowing the user to visualize and interact with paths of people, objects and events over time and space. For example, the blue line in Figure 10 represents the path a taxi took throughout a work day. Points closer to the map represent more recent events, while those further from the map happened longer in the past. This

visualization gives the analyst a visual history of where the cab was at certain points in the day. Many other tools have features built in for support of geographic data, including Tableau (Hanrahan, Stolte, & Mackinlay, 2007) and Starlight (Risch et al., 2002).

The visual analytics tools discussed represent only a small subset of the total tools available. These tools have been integrated into many organizations and government agencies and continue to become more widespread.

2.4 Evaluation of Visual Analytics Tools

One of the main goals for visual analytics has been to provide the means to increase insight via the use of visualizations. This statement of intent is present in the visual analytics agenda, *Illuminating the Path*:

“People use visual analytics tools and techniques to synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data”(Thomas & Cook, 2005).

Even before the research agenda for visual analytics was published, insight appeared as one of the desired outcomes of visualizations. Card et al. (1999) state, “The purpose of visualization is insight, not pictures. The main goals of this insight are discovery, decision making and explanation”. But what is insight? And how can we measure it?

“Insight” is often touted as a selling point for VA tools. Dictionaries define insight as “apprehending the true nature of a thing”, or “gaining an understanding that sheds light on or helps to solve a problem” (“Dictionary.com Unabridged,” 2011). Insight can occur at any time during the sensemaking loop. It often

involves connecting new information to existing knowledge in a new way.

Another way to define insight is to say that it is the step when information goes from being chaotic and unorganized to some type of user-defined order.

However, it is often difficult to know when insight has been achieved, making it equally difficult to quantify and measure.

“Aha” moments are the most obvious cases of insight, and represent a conscious recognition of insight. During an “aha” moment, a person will report a subjective feeling that a solution or new understanding has been reached. For these types of moments, there is evidence of three factors:

1. People have usually reached some sort of dilemma in problem solving in which the way to move forward is no longer clear (Dominowski & Dallob, 1995).
2. The person usually cannot report the processing that enables them to overcome the dilemma (Gick & Lockhart, 1995).
3. The solutions they experience are usually sudden and surprising (Davidson, 1995).

The second point is of particular interest in the case of evaluating visual analytic tools. Due to people’s inability to report the processing going on during insight, it has become a sort of “black box” process that occurs within the human brain. However, by creating tasks where insight is expected, cognitive science researchers have been able to measure which parts of the brain are active when insight occurs. By using compound remote associate problems, Bowden and

Jung-Beeman (2003) found that the experience of insight is correlated with solution activation in the brain's right hemisphere. Later, using electroencephalogram (EEG) and functional magnetic resonance imaging (fMRI), they found different brain patterns for different types of problem solving (Bowden, Jung-Beeman, Fleck, & Kounios, 2005; Kounios & Beeman, 2009). Normal problem solving and insight-based problem solving show activity within the frontal and temporal lobes, suggesting that working memory, executive processes, long-term memory and semantic information all play a part in insight (Chang, Ziemkiewicz, Green, & Ribarsky, 2009). However, due to the difference in the type of activity going on in the two hemispheres, Chang et al. (2009) go on to interpret the findings as indicative of a qualitative difference between spontaneous insight and everyday problem-solving. Ultimately, this means that the compound remote association problems and measured insight via EEG and fMRI are a relatively simple version of the type of analysis that occurs in visual analytics problems. These more complex visual analytic problems involve many of the iterative steps in the sensemaking process (Pirolli & Card, 2005) and require solutions that involve more than just word associations.

Chang et al. (2009) argue that insight from visualization is viewed in a fundamentally different manner than for cognitive science, and it is necessary to separate the two. The terminology used to describe insight in cognitive science is often related to an event that happens, where visualization-related insight more often refers to a substance, or unit of knowledge. Developing metrics for spontaneous insight would involve exploratory, pre-query approaches while

evaluating knowledge-building insight would involve measuring the amount of knowledge gained from knowledge-gathering methods (Chang et al., 2009). Both types of insight are viewed as goals for what good visualizations can provide. Experiments carried out by Bowden and Jung-Beeman (Bowden & Jung-Beeman, 2003) analyzed the brain during insightful moments using word problems.

Insight is just one part of the overall sensemaking and analysis process. Defining what should be evaluated is very subjective and up for active debate. The current metrics used to evaluate various steps of the analytic process are often quite complex or simply non-existent. Sholtz (2006) argues that areas such as situation awareness, collaboration, interaction, creativity and utility must also be evaluated. The usability of a tool also needs to be evaluated, as users will resist learning something that is poorly designed. However, to better evaluate these areas, there is a need to better understand the analytic process in the context of real-world analysis.

2.4.1 Analytical Reasoning

As *Illuminating the Path* (Thomas & Cook, 2005) notes, visual analytics is about more than just the visualization: “Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces”. Framed within the context of sensemaking, analytical reasoning refers to the processing and evaluation of information and evidence by making judgments (Thomas & Cook, 2005). In the second chapter of *Illuminating the Path*, there are several strategies

laid out for the advancement of analytical reasoning research. The first recommendation put forth is:

“Build upon theoretical foundations of reasoning, sense-making, cognition, and perception to create visually enabled tools to support collaborative analytic reasoning about complex and dynamic problems.” (Thomas & Cook, 2005).

Thomas & Cook (Thomas & Cook, 2005) further explain the difference between convergent thinking (assembling evidence to find an answer) and divergent thinking (thinking creatively to make sure plausible alternatives have not been overlooked). Both types of thinking are present in analysis, and are often interwoven due to the iterative nature of analysis. The key point is that the visual analytic tools should ultimately support all types of human reasoning about complex problems.

However, research into complex problems faced by analysts and the reasoning process is often met with several difficulties. One obstacle with researching analytical reasoning is that working analysts are very busy people who often deal with proprietary and confidential material. In academic research the goal is most often to publish findings in a journal or conference proceedings. However, this open view of sharing findings with the rest of the world via publications is often in stark contrast to the proprietary and confidential nature of many analysts' work. As a result, published studies of “real-world” analysts are less common than studies of “mock” analysts.

One method of advancing research into complex analytical tasks has been to utilize contests created for advancing visual analytics tools (Grinstein et al.,

2006). The first of these contests specific to visual analytics was the Institute of Electrical and Electronics Engineers (IEEE) Conference on Visual Analytics Science and Technology (VAST) in 2006. The VAST 2006 contest involved a set of approximately 1200 fictional news items from the fictional town of Alderwood and included supplementary material like photos and phone call logs. The contest was based on real-world intelligence analysis, in which an analyst is commonly expected to sort through a large number of text documents and determine what the threat is, who is involved, what locations are of interest and finally a timeframe for the threat. Creation of the contest dataset was aided by the Threat Stream Generator (M. A. Whiting et al., 2006), a tool and methodology from Pacific Northwest National Laboratory (PNNL) utilizing real-world data such as news articles to mask the “true threat”. Participants were encouraged to develop and test their own visual analytics tools by solving the task given by Grinstein et al. (2006).

This contest served multiple purposes. One was to encourage the development of relevant and accurate tools for intelligence analysis. By providing a somewhat realistic “real-world” dataset and challenge, toolmakers had the chance to refine and evaluate their tool’s usefulness by completing the challenge and recording the process. Another was to help establish metrics for evaluating visual analytics tools. By utilizing a standardized dataset and objective, the tools could be assessed both on their own and against one another for a variety of different things, including utility of interface and visualizations, quality of interaction, scalability, versatility...etc (Grinstein et al., 2006).

The final benefit of contest datasets and tasks is their contribution to favorable conditions to conduct a study on analytical reasoning. Though the analysts are often not “professional” due to the nature of the contest, analytical reasoning from non-professional analysts can still be investigated. Most of the subjects tested end up being students at the respective university where the tool was developed. For example, in the second annual VAST contest, Ha et al. (2007) applied cognitive task analysis, grounded theory and activity theory to field notes and interviews of students who attempted the contest. The goal of this study was to illuminate the analytic process used by analysts (students in this case) working through a complex task. Rather than use a single visual analytics tool, Ha et al. (2007) used a “bricolage approach” involving many off-the-shelf and publicly available tools.

Students collected their own field notes about their analytic process including a rough estimate of the amount of time spent on each task. From analysis of interviews and field notes, Ha et al. (2007) found that early stages of analysis were characterized by a much more individualistic approach, whereas later steps of analysis like hypothesis evaluation occurred collaboratively.

This study provided data that can serve to improve design guidelines for visual analytics tools whose goal is capitalizing on what the computer and human do best. Other studies have attempted this by establishing a framework for a human cognition model to help guide the development of visual interfaces, and propose that humans typically outperform computers with regard to adaptation, incorporating new information into existing knowledge schemas, and in

categorizing stimuli almost instantaneously (T. Green, Ribarsky, & Fisher, 2008). Computers, on the other hand, excel in a few areas in which humans are severely limited. Two of these are working memory (they have a lot) and overcoming cognitive biases (they don't have any). Computers can therefore aid in the search for relevant information through computer-aided discovery (T. Green et al., 2008).

Often during analytical reasoning people are forced to make decisions in the face of uncertainty. Humans take varying approaches to dealing with uncertainty, however the most intuitive heuristics people use are often subject to biases that can lead to errors (Tversky & Kahneman, 1974). Strategic decision-making occurs when there is a high amount of novelty, complexity and open-endedness involved in the decision (Mintzberg, Raisinghani, & Theoret, 1976; Schwenk, 1984). This is often the case when dealing with visual analytics tasks. As a result, these problems are very complex, demanding a decision-maker to manage uncertainty. In addition, choices are made using some form of bounded rationality, as perfectly rational decisions are not always feasible due to cognitive limitations of the decision-maker (Simon, 1997).

Humans utilize a limited set of techniques to managing uncertainty. In an influential paper outlining judgment under uncertainty, Tversky and Kahneman (1974) highlight several heuristics that people use when dealing with uncertain data. Within this context the term 'heuristic' refers to a strategy for reducing the complex tasks of assessing probabilities and predicting values. While both useful

and necessary, these heuristics often lead to biases. When the biases lead to errors in judgment it can become a severe or serious problem.

Human short term or working memory is very limited, with the ability to store only a few different things at any given time (Cowan, 2001, 2005; Miller, 1956). However, as discussed, many decisions often involve a large number of factors, each with their own uncertainty. In order to prevent every decision from taking an enormous amount of time, heuristics serve as a shortcut.

During the analytic process there are several occasions when a person may have to deal with uncertainty. Some examples include:

“1) identification of uncertainty, which is frequently treated as a given but usually is actually well hidden and fuzzy, 2) representation of uncertainty, 3) aggregation of uncertainty, and 4) communication of uncertainty” (Thomas & Cook, 2005).

Uncertainty occurs more often in later steps of the sensemaking loop, when hypotheses are being weighed for validity. Very often uncertainty is communicated as a probability or odds ratio as a means to understand and deal with it (Heuer, 1999). It is important to be aware of issues surrounding uncertain data and assumptions and to consider these when developing or evaluating tools. Ultimately, there is still much research to be conducted to understand analytic reasoning and how humans deal with uncertainties.

While methods such as contest data studies with graduate students represent a marked improvement over previous means for understanding complex analytical tasks, monitoring and studying analysts working in natural environments on real tasks represent the gold standard for this type of research.

Because real-world analysts have limited time and accessibility, a researcher needs to be prepared and versed in effective techniques for observation when given the opportunity to carry out such an in-situ or field study.

2.4.2 Participant Observation

Studies of human participants for gathering empirical data have often been carried out using methods from ethnography, and can include a combination of participant observations, interviews and questionnaires. These field studies are useful for collecting data from natural settings and work environments. For our purposes, we are interested in designing and improving analytic tools. The rationale behind observing people in their “natural setting” comes partially from “The Hawthorne Effect”, an idea that people change their behavior in lab or other artificial environments, ultimately tainting the data collected (Adair, 1984; Calder, Phillips, & Tybout, 1981). While the extent of the Hawthorne Effect is debated, there is no doubt one method for reducing this effect is through observation in a natural setting.

Participant observation strives to be as non-intrusive as possible while collecting valuable data about a person in their natural habitat (Becker & Geer, 1957; K. M. DeWalt & DeWalt, 2002; Jorgensen, 1989). In their guide for fieldworkers using participant observation methods, Dewalt & Dewalt (2002) describe participant observation in the following way:

“For anthropologists and social scientists, participant observation is a method in which a researcher takes part in the daily activities, rituals, interactions, and events of a group of people as one of the

means of learning the explicit and tacit aspects of their life routines and their culture” (K. M. DeWalt & DeWalt, 2002).

Observation has the goal of learning both “explicit and tacit aspects” of a person or group of people, both of which apply to a study of analytic reasoning. Explicit knowledge can be articulated more easily and transferred from one person to another. However, as I have discussed in the previous chapter, during analysis the reasoning process cannot be articulated (Gick & Lockhart, 1995; Trickett et al., 2000). This “tacit knowledge” is the black box present throughout many aspects of analytical reasoning. Participant observation methods seek to bring some of this tacit knowledge to the surface.

Participant observation is a qualitative research method with its own defined rules and guidelines (Becker & Geer, 1957; K. M. DeWalt & DeWalt, 2002; Jorgensen, 1989). The methods of recording observations as field notes as well as collecting data from participation are well established in the fields of anthropology, sociology, psychology and education as well as an emergent framework for cognitive science. In order to be an effective observer during participation, the researcher should be documenting the layout of the physical space, the participants involved and both verbal and non-verbal interaction between participants (K. M. DeWalt & DeWalt, 2002). In addition to qualitative data collection, some type of counting can and often should occur for quantitative analysis. Counting certain types of events is a necessary step if there is a desire to look at changes over time. DeWalt & DeWalt (2002) also include “seeing old events with new eyes” as a relevant requirement. Generally, the observer is not a domain expert and will therefore make observations based on his or her own

mental representations and schemas. This is often useful for illuminating the nature of activities and processes that the domain experts take for granted. However, it can also lead to confusion and the need to have particular tasks explained in more detail by the domain expert.

Atkinson and Hammersley (1994) lay out the following features to get a sense of when it is appropriate to use ethnographic methods like participant observation:

- A strong emphasis on exploring the nature of particular social phenomena, rather than setting out to test hypotheses about them
- A tendency to work primarily with “unstructured” data, that is, data that have not been coded at the point of data collection in terms of a closed set of analytic categories
- Investigation of a small number of cases, perhaps just one case, in detail
- Analysis of data that involves explicit interpretation of the meanings and functions of human actions, the product of which mainly takes the form of verbal descriptions and explanations, with quantification and statistical analysis playing a subordinate role at most

Many of these features apply in the case of trying to understand analytical reasoning. As discussed in the previous section on insight, people are not always good at describing the processes they use to solve problems (Amar & Stasko, 2005; Bowden et al., 2005; T. Green et al., 2008). Therefore, interviews are a necessary and important technique for this study supplemented by data collected through observation of the actual task.

Due to people's inability to effectively verbalize their internal processes during analysis (Gick & Lockhart, 1995; Trickett et al., 2000), additional data collection methods are often necessary to help cross-validate findings (Arias-Hernandez et al., 2011). Studies utilizing ethnographic methods to evaluate visualizations have been met with mixed results. Saraiya, North and Duca (2005) used a mixed methods approach to evaluate Bioinformatics visualizations. The study used insight as an evaluation metric, with researchers defining measures for insight across a few different tasks. Participants were split into one of five groups, with each group using a different tool able to visualize microarrays. Using three separate datasets for each group, Saraiya et al. (2005) were able to evaluate participant performance. Using a combination of the users' data questions, insights, visualization usage, and user background, the researchers found that visualizations had a large impact on insight with regard to biological data. By combining observations with questionnaires they were able to increase the depth of their findings.

Shneiderman and Plaisant (2006) describe "Multi-dimensional In-depth Long-term Case studies" (MILCs). These include using a mixed method approach similar to Saraiya et al. (2006) involving "interviews, surveys, as well as automated logging to assess user performance and interface efficacy and utility" (B. Shneiderman & Plaisant, 2006). A combination of methods helps to ensure that more data is collected and fewer things are missed.

Participant observation is a frequently used method in qualitative research. When a rigorous protocol is followed researchers have succeeded in

capturing rich data. Many aspects of past participant observation studies were used for paired analysis and the research presented in this thesis.

2.4.3 Protocol Analysis

Protocol analysis is a research method that has been widely used in psychological research to study thinking. Human subjects give verbal reports of their thoughts and behaviors while performing cognitive tasks, thus providing insight into human cognition (Ericsson & Simon, 1993). This method is based on the assumption that humans are able to verbalize their inner thought processes and the steps they took to move the analysis forward.

The “think aloud” process often used for protocol analysis is based on information processing theories of the brain, which liken cognitive operations to information processing in computers. With information processing theory, behavior can be described as “an interaction between an information-processing system, the problem solver, and a task environment” (Simon, 1978). When approaching the task, the problem solver represents the situation in terms of a problem space, and therefore the framework for problem-solving behavior is composed of the information-processing system, task environment and problem space (Newell & Simon, 1972). Each of these components is further explained by Simon (1978):

- Information-processing system – Due to limited attention, the human system operates serially rather than in parallel. Inputs and outputs are held in short-term memory which has the capacity of only a few chunks, allowing the brain to process them one at a time

- The task environment – The rules and structure of the task at hand make up the task environment.
- The problem space – the manner in which the problem solver represents the task environment in memory. The ease of a task depends on how well the problem solver represents the task environment.

By representing problem solving in this manner, Simon claims that if a think aloud protocol is used the researcher will gain insight into each serial step in the process (Simon, 1978). With this rationale the thought will be verbalized as soon as it enters the subject's consciousness, allowing the researcher to capture it. Based on the assumptions of information processing theory, only information brought into the short-term memory by ongoing cognitive processes can be further processed and verbalized directly (Simon, 1978).

Despite the obvious advantages of protocol analysis when viewed through the lens of information processing theory, Simon (1978) also recognized that verbal reports are often inconsistent with observed behavior. Nisbitt and Wilson (1977) found that when people were instructed to report on their cognitive processes, reports were sometimes based on "a priori, implicit causal theories or judgments about the extent to which a particular stimulus is a plausible cause of a given response". This effectively means that people who were unable to report their processes introspectively either consciously or unconsciously came up with rational explanations for their behavior. While these explanations may sometimes accurately reflect true cognitive processes, many times they do not and therefore cannot be regarded as reliable.

To avoid false introspection on the part of the participant, proponents of protocol analysis employ controlled techniques for both think aloud studies and after-the-task interviews. In the chapter titled “Understanding Users: Qualitative Research”, Cooper et al. (2007) offer several guidelines and pieces of advice regarding interviews and interactions with participants. Several of Cooper’s (2007) guidelines that are most relevant for visual analytics and analytical reasoning research have been summarized below:

- Interviewing Subject Matter Experts: When the SME suggests new features to a tool or method it is important to figure out the cause of the suggestion, often by uncovering underlying problems or issues involved. Finding out information on the industry’s best practices and complex regulations can be particularly helpful.
- Try to formulate most questions in an open-ended way, allowing for responses rich with data for later analysis.
- It is important to possess some domain knowledge to facilitate communication between the researcher and SME
- Understand how products like software fit into the analysis workflow
- Regarding cameras/video: Can't be obtrusive or interfere with the interview
- It is necessary to establish good rapport with the SME

For the master-apprentice model during a one on one interview, four principles should be followed:

1. Context - interview/observe in user's natural habitat
2. Partnership - should take tone of collaborative exploration
3. Interpretation - take all of the facts gathered and analyze to understand implications

4. Focus - subtly direct interview to capture relevant issues

It is important to keep interviews as short as possible while still collecting enough relevant data. To prepare for interviews there are several other guidelines:

- Identifying candidates - need diverse sample of users and user types.
- Develop persona hypotheses - based on likely behavior patterns and factors that differentiate these patterns.
- Create behavioral distinction between technical expertise and domain expertise
- Environmental considerations - interpersonal work space vs. high level of bureaucracy
- Putting plan together – Schedule interviews when it works best for the SME if at all possible

For conducting interviews:

- A team of two may work best - one to drive and take light notes, other to take detailed notes
- Conducting multiple phases of interviews - Early phases contain more broad inquiries and are exploratory. Middle phases are for open-ended clarifying questions. Later phases confirm previous patterns and contain more closed-ended questions
- Interview where the interaction happens
- Avoid a fixed set of questions
- Focus on goals first, questions second
- Avoid making user a designer of the tool
- Avoid discussions of technology
- Encourage storytelling
- Ask for show and tell
- Avoid leading questions

Studies using protocol analysis have been performed looking at visualizations and visual analytics with varying levels of success. For example, in an effort to assess the usage of complex meteorological visualizations, Trafton et al. (2000) used protocol analysis to understand the qualitative mental models of forecasters using those visualizations. They asked participants to talk out loud as they worked through their forecasting tasks and ultimately encoded data using a tool (MacSHAPA) that allows the coder to associate categories, notes and timestamps with events on tape (Trafton et al., 2000). This method proved useful for assessing the qualitative mental models and integration of visualizations for the forecasters. Protocol analysis was also used to assess information foraging of MBA students in a strategic management course (Pirolli & Card, 1999). Pirolli and Card (1999) asked students to use a think aloud protocol when working solo and video recorded them. In a study of the investigative analysis prowess of the visual analytics tool Jigsaw, Kang et al. (2009) conducted interviews with participants after they completed tasks using the software.

Despite the widespread use of protocol analysis, there have been examples of problems with verbal reporting. For example, Schooler et al. (1993) conducted specific experiments designed to assess whether verbalization interferes with insight problem solving. They found that in some cases verbalization disrupted “the non-reportable processes that are critical to achieving insight solutions” (Schooler et al., 1993). Another study found that as participants became more engrossed in the task at hand their instances of verbal reporting began to decrease (Trickett et al., 2000). Simon was aware of this

when talking about human problem solving, stating, "In general, instructions to think aloud do not interfere with subjects' thought processes, although they may slow them down somewhat if the subject is using a non-verbal code in his thinking" (Simon, 1978). Another issue Trickett found was that a scientist using the think aloud protocol "may change her focus to aspects of the data that are more easily verbalized, or may change her representations from visio-spatial representations to more verbal representations" (Trickett et al., 2000).

Verbal reports based on protocol analysis methods provide valuable data towards the sequence of events that occur when a human is solving a problem or performing some other cognitive task. However there are significant limitations and shortcomings of protocol analysis for designing a study to examine analytical reasoning and cognitive processes. Combining other techniques with methods from protocol analysis results in the best way to capture data on analytical reasoning processes.

2.5 Paired Analysis

2.5.1 Overview

Paired analysis is a collaborative method of analysis involving two distinct experts: one whose expertise is in the subject matter of the analysis and a second who is very knowledgeable about a given piece of software used for the analysis. The two collaborate on a common task, but have distinct roles within the task. The software expert is always in control of the software and interaction with data and visualizations. The subject matter expert has explicit and implicit

knowledge in a particular domain, and moves the analysis forward by directing and interacting with the software expert. As an ethnographic method for data collection for visual analytics evaluation, we propose that this protocol has the potential to overcome some of the limitations discussed above with regard to think aloud protocols and participant observation.

The three main advantages of this approach are:

1. Access to the SME's thought process
2. Incorporation of a think aloud protocol into continuous analysis
3. Domain and tool expertise learning for the tool expert (TE) and subject matter expert (SME) respectively

First, the SME/Analyst is forced to verbalize thoughts to advance analysis. By recording and observing the dialogue and actions of the SME, researchers have a real-time view of the analytic process of the SME. As discussed, when an expert is required to speak aloud about what they are thinking or how their analytic process is progressing, the actual task may be hindered (Schooler et al., 1993; Trickett et al., 2000). However, paired analysis involves speech that becomes a defined part of workflow, unlike some of the speak-aloud methods found in protocol analysis studies (Kang et al., 2009; Scholtz, 2006; Trafton et al., 2000). A further benefit is increased interaction, allowing the subject matter expert and software expert to benefit from each other's expertise. The software expert learns the domain and the SME learns what the tool is capable of.

The paired analysis approach has been influenced by existing techniques spanning multiple domains (Arias-Hernandez et al., 2011). The idea of a pair of analysts working together was inspired by a technique from the software development world known as Extreme Programming (Beck, 2000). Also known as pair programming, the technique involves two programmers developing software side by side at a single computer (Beck & Andres, 2004). One of the programmers types the code while the other checks over the code for errors and logs potential future issues. Pair programming was introduced for several reasons:

- To improve software quality by having a second person present to detect bugs and review code.
- Increase the efficiency and responsiveness to changing customer requirements.
- Shorten the software development cycle (Beck & Andres, 2004; Cockburn & Williams, 2001; Williams, 2002; Williams, Kessler, Cunningham, & Jeffries, 2002).

One ambition of pair programming is to allow programmers to work on relevant tasks with a second set of eyes, and not have to face new and difficult situations alone. For customers and managers of software developers, pair programming provides a way to check progress while ensuring that changing the direction of the project is still possible (Beck, 2000). Two of the main principles from pair programming that have influenced paired analysis are:

1. “Its reliance on oral communication, tests, and source code to communicate system structure and intent...”

2. “Its reliance on the close collaboration of programmers with ordinary skills” (Beck & Andres, 2004).

Paired analysis embraces collaboration on a given task in much the same way as pair programming. However the task is no longer developing software, and is instead some type of analysis. The domain and type of analysis is not important, but instead the skill sets and roles of the two people collaborating.

In defining paired analysis for visual analytics, the two participants no longer have similar roles and backgrounds as they did in pair programming. Instead, the first role is that of the tool expert, a person who has had training in visual analytics as well as experience using visual analytics software. The second role is that of the SME, whose extensive domain knowledge provides the expertise necessary to make informed decisions during analysis. The roles begin as distinct, however as the analysis progresses the roles may begin to overlap.

For example, when analyzing paired analysis sessions conducted on airline maintenance data, Arias-Hernandez et al. (2011) found that “higher-order cognitive processes, like hypothesis making, are not only conducted by the SME, and lower-order cognitive processes, like procedural interaction with the GUI, are not only conducted by the VAE (Visual Analytics Expert)”. One possible reason is that the two participants are learning from one another as the task moves forward. The software expert is gaining knowledge about the domain and ultimately begins understanding what types of analysis techniques are used for the given domain. The SME is learning about what the visual analytics tool is capable of. By understanding the different visualizations possible and how to

interact with data, the SME's own analysis schema will begin to accommodate possible uses of the tool, leading to suggestions of how it can be used.

2.5.2 Bringing Tacit Knowledge to Light

Another influence for paired analysis comes from cognitive science and psychology. One of the primary goals of paired analysis is to study the analytic process of the SME. As we have discussed, previous attempts to understand the analytic process of an SME have involved think-aloud protocols and protocol analysis (Ericsson & Simon, 1993). While these attempts have succeeded in uncovering some of the process, sometimes speaking out loud has been disruptive (Schooler et al., 1993; Trickett et al., 2000). A further piece of evidence demonstrating the shortcomings of speaking aloud is that a SME may not know exactly how to verbalize what they are doing while performing a solo analysis. Paired analysis assigns the role of manipulating data and providing visualizations to the software expert, while the SME directs the analysis. This setup not only encourages collaboration through speaking aloud, but in many ways "forces" it in order to be successful. Theoretically, the act of taking on the role as director of analysis should cause the SME's verbalizations to become less disruptive than if he had to both manipulate the data and describe what analysis steps are being taken. The spoken dialogue essentially becomes a necessary part of analysis rather than a separate, additional burden. Because the SME is directing dialogue at the TE rather than an observer or camera, the SME is more likely to describe *why* decisions are being made as opposed to simply *what* decisions are being made.

2.5.3 Coordinating Joint Action

One of the main strengths of paired analysis is its ability to elicit a natural dialogue between the SME and software expert. By incorporating interaction between the participants in a less intrusive manner, the ability to coordinate joint action occurs. According to Clark (1996), language is a form of joint action, which requires coordination of both the content of the activity and the process by which the activity moves forward. From our initial pilot studies using the paired analysis method it is evident that both participants need to be aware of and have at least some familiarity with the other's expertise (Arias-Hernandez et al., 2011). This "common ground" helps minimize problems with communication between the SME and software expert. Without some knowledge of the other's domain, the flow of the paired analysis session would be disrupted with constant clarifications and explanations, resulting in very little analysis actually taking place. Clark (1996) describes the source of our ability to coordinate as the set of knowledge, beliefs and suppositions that both people share. In the case of paired analytics, this common knowledge must be established before the actual analysis session in order to ensure smooth interaction. Common ground makes it possible for a person speaking and a person listening to coordinate on what the speaker means and what the listener understands the speaker to mean.

During the initial pilot studies, paired analysis sessions were both screen captured and video recorded for later analysis (Arias-Hernandez et al., 2011). There are multiple goals and possible ways to learn from these recordings. The first is to analyze the discourse between the SME and software expert with the

goal of exposing and understanding the analytic process of the SME. Second, the sessions help to determine whether the tool was effective for the needs of the SME. Third, the recordings can be used to determine where current visual analytics tools are lacking in areas such as sensemaking, helping to lay a foundation for the design and implementation of new visual analytics tools addressing these concerns.

3: AEROINFO: INTERVIEWS AND PILOT STUDIES

3.1 Introduction

Prior to the implementation of paired analysis with safety engineers we conducted a series of interviews and pilot studies with SMEs who were very knowledgeable about aircraft maintenance and the airline industry. During initial interviews, the research objective was focused on uncovering uncertainty in aircraft maintenance planning. More specifically, we wanted to know how maintenance planners conceptualize uncertainty and how do they deal with it in data and during analysis. This research focus is reflected in the interview questions and responses from participants. Pilot study 1 at Aeroinfo consisted of interviews with maintenance planners. Pilot study 2 involved paired analysis sessions. The paired analysis pilot studies at Aeroinfo helped refine the method for use with Boeing Safety engineers.

3.2 Background: Uncertainty

The average person must manage uncertainty on a daily basis, from determining whether to invest in a particular stock or predicting the outcome of a baseball game. However, in some cases, such as that of aircraft maintenance planners, the consequences can be very serious. When people's lives are at stake, effective tools and strategies to manage and communicate uncertainty are essential. Determining an appropriate visualization of uncertainty is highly dependent on the task or domain as well as knowledge of the cognitive and

perceptual processes surrounding uncertainty. Before being able to visually represent uncertainty effectively there is a need to understand where it fits into the analytic process and current methods of visualization. We interviewed former aircraft maintenance planners to understand the cognitive and procedural aspects of managing uncertainty. Our findings indicate there are a wide variety of conceptualizations of uncertainty within aircraft maintenance planners. Data from these interviews presented in this thesis has helped form a foundation for a cognitive model of uncertainty as well as provide a context for a paired analysis study.

3.3 Background: Aircraft Maintenance and Maintenance Planners

No man-made system is perfect. A commercial aircraft such as the Boeing 747 is made up of millions of individual parts. With such complexity and so many different places for things to go wrong, regular maintenance is essential to keep aircraft functioning safely. Maintenance planners are assigned the task of scheduling both regular maintenance and dealing with unexpected events that require unscheduled maintenance. There are always unexpected events in the real world that were not tested in models. This uncertainty and these breakdowns must be addressed in order to keep a multi-million dollar flying piece of machinery from becoming an extremely large and complex paperweight.

Physical components on a plane have a limited lifespan. When unexpected events raise the need for repair, airplanes must undergo maintenance. Maintenance takes time and money, and while being performed

the aircraft is not generating revenue. The role of the aircraft maintenance planner is to strike a balance between keeping the plane safe through maintenance and minimizing the Aircraft-On-Ground (AOG) time. In order to maximize both profit and safety, planners must take several factors into account, including; man-hours available for maintenance, part availability, hangar space, aircraft base locations and seriousness of the problem. Each of these factors have degrees of uncertainty that are managed by the maintenance planner.

During aircraft maintenance incorrect decisions due to uncertainty can have dire consequences. A plane with inadequate maintenance is more likely to be involved in an accident or incident, putting the passengers' lives at risk. For maintenance planners, this includes creating short and long-term maintenance plans. However there are also many unexpected events that must be dealt with when they arise.

Whether scheduling maintenance for regular checkups or unexpected events, maintenance planners are forced to deal with uncertainty nearly every day. This uncertainty can arise from a number of places. Events such as a bird colliding with a plane or cold temperatures causing a crack in the hull are infrequent and often unpredictable. After such an event occurs, a maintenance planner is required to assess the event in a number of ways. These include determining whether maintenance should be carried out right away, assessing the resources available to fix the problem, and moving the plane to a base or hangar where the maintenance can be completed.

In order to manage uncertainty in these types of tasks and effectively maintain a fleet of aircraft, aircraft maintenance planners develop a conceptualization of uncertainty and strategies of dealing with it. In order to investigate these conceptualizations and management strategies, one-on-one interviews were carried out with former maintenance planners. The interview questions were guided by a focus structure designed to elicit rich responses and anecdotal stories relating to uncertainty (See Appendix A: Aeroinfo Focus Structure).

In addition to uncovering uncertainty, the interviews were designed to give the researchers a detailed background and context of aircraft maintenance in preparation for a paired analysis study. As previously indicated, common ground between the tool expert and subject matter expert is essential before beginning any paired analysis session.

3.4 Pilot Study 1: Interviews at Aeroinfo

The methodology for conducting interviews involved following a focus structure that loosely outlined the topics to be covered (Cooper et al., 2007). Questions were open-ended and allowed for subjects to provide detailed responses. The initial interview subjects were aircraft maintenance planners at Aeroinfo who were interviewed to gain a better understanding of maintenance operations of airlines and aircrafts and daily tasks for improving those operations.

The interviews for pilot study 1 were carried out in two phases. In the first phase, a business development manager (BDM) was interviewed. The BDM is

not a maintenance planner himself, but has worked with several maintenance planners through his previous experience in the airline industry. The goal of the interview with the BDM was to gain some basic knowledge of maintenance planning. The transcript from the BDM interview was used to get an overview of the tasks and activities of the maintenance planners. Questions for subsequent interviews with actual maintenance planners during pilot study 1 were modified based on this initial interview.

The second stage of pilot study 1 consisted of one-on-one interviews with three participants. All three have experience as maintenance planners and extensive backgrounds in the aircraft industry. These interviews lasted approximately one hour each and were audio-recorded. Interviews were performed in the following format: The researcher began with a brief introduction about the goals of the study, letting the planner know that one of the goals was studying uncertainty and what uncertainty meant to the planner. Then, overview questions were asked to learn about each planner's background and experience in addition to the types of tasks and problems they encountered in a regular day. Further probing questions were asked based on responses and whether the interviewer felt there was potential for studying uncertainty. The focus structure is found in Appendix A: Aeroinfo Focus Structure.

Due to the exploratory nature of pilot study 1, we chose an approach of a loosely structured interview, probing deeper when necessary. This yielded a wider range of data on more topics as well as richer qualitative data than would have occurred with a fixed script.

3.4.1 Pilot Study 1: Overview of Findings

Transcripts from interviews during pilot study 1 were analyzed with the goal of understanding where uncertainty fits into the analytic process of aircraft maintenance planning. We found there are two types of maintenance planners; short-range and long-range.

With short-range, maintenance for a fleet of aircraft is planned over the course of weeks and months, while a long-range planner creates maintenance plans for up to five years in the future. Both have to take into account failure rates of individual parts, maintenance personnel, hangar space and unplanned maintenance, also known as 'discrepancies'. These categories are discussed individually with relation to uncertainty in the following sections.

During the interviews, maintenance planners had varying conceptualizations of where uncertainty was most frequent. The three most common areas of uncertainty were:

1. Individual aircraft components and their failure rates
2. Managing maintenance personnel
3. Reliability of data

Each of these areas will be discussed in more detail in the following sections.

3.4.2 Pilot Study 1 Findings: Aircraft Components

At the highest level, there are basically two types of maintenance: preventative and unscheduled maintenance. Preventative maintenance is also

known as scheduled maintenance because it is performed at regular intervals and is designed to prevent deterioration of the system or component (Kinnison, 2004). Unscheduled maintenance occurs when a component fails unexpectedly or some other event occurs that requires attention in order to bring the aircraft back to a safe operating level.

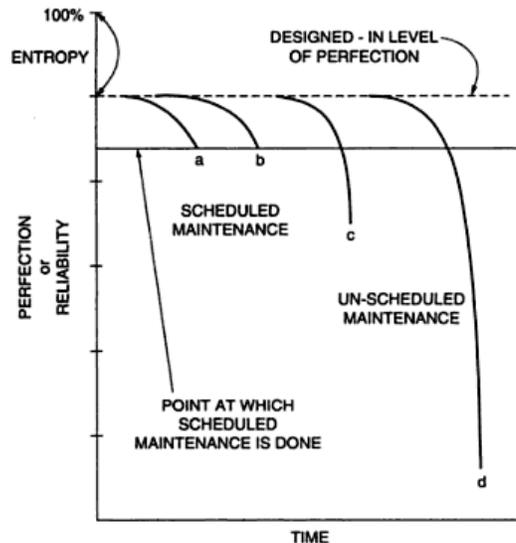


Figure 11: Relation between failure of components and types of maintenance

In Figure 11, the dotted line represents an individual component working perfectly (Kinnison, 2004). Over time the component begins to wear, and once it crosses the threshold of the solid horizontal line maintenance is required. With components 'a' and 'b', the rate of wear is known and occurs accordingly. The maintenance is scheduled at time points 'a' and 'b' respectively, and the component never causes the plane to be unsafe (Their curves never cross the solid line threshold). However with 'c' and 'd', the components wore out at a different rate than expected, crossing the threshold and requiring unscheduled maintenance. This could have been due to a variety of things, ranging from faulty

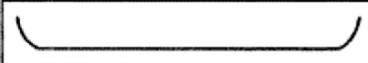
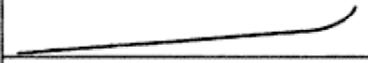
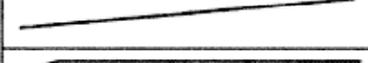
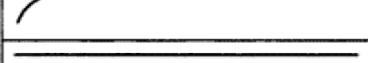
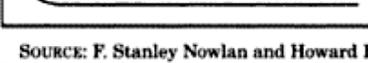
manufacturing to a bird colliding with the plane. In the case of 'd', the component was significantly more unreliable than predicted. It is likely that there will be a full investigation of 'd' as to the cause of the breakdown. During such investigations there may also be a reassessment of the reliability of all 'd' components to determine whether the expected failure rate needs to be modified.

Within the detailed analysis of component 'd' there is significant uncertainty that a maintenance planner manages. For example, the planner initially searches for related component failures, looks for a root cause, and determines whether this is an isolated incident or whether it needs to be addressed in the entire fleet of aircraft. A further job of the maintenance planner is to determine whether the allotted time for discrepancies such as 'd' in regular maintenance checks needs to be changed. Within each scheduled maintenance check for a plane, a percentage of time is allocated for discrepancies. The time allotted to fixing discrepancies can be thought of as a strategy for managing 'certain uncertainties', a conceptualization of uncertainty put forth by the first planner interviewed.

As mentioned before, there are millions of individual parts and components making up each aircraft. Individual components of airplanes have very different reliability measures, and break down at different times. The expected rate for a part to fail is known as the 'failure rate' (Kinnison, 2004). It is important to note that failure rates vary significantly in both length of time and pattern. After performing a study on lifetime failure rates of aircraft components, a major commercial airline found that there are six basic patterns (Kinnison, 2004).

These patterns are illustrated in Figure 12. From this diagram we can see that there can be significant differences in the pattern for wear out and failure of various components.

TABLE 1-1 Failure Rate Patterns

	A. Infant mortality; constant or slightly rising failure rate; definite wear-out period (4%)
	B. No infant mortality; slightly rising failure rate; definite wear-out period (2%)
	C. No infant mortality; slightly rising rate; no definite wear-out period (5%)
	D. Increasing failure rate at outset; constant or slightly rising rate; no definite wear-out period (7%)
	E. No infant mortality; constant failure rate throughout life; no definite wear-out period (14%)
	F. Infant mortality; constant failure rate throughout life; no definite wear-out period (68%)

SOURCE: F. Stanley Nowlan and Howard F. Heap: *Reliability-Centered Maintenance*; U.S. Department of Commerce, National Technical Information Service, Washington, DC, 1978.

Figure 12: Failure rate patterns for individual components

Maintenance planners must take these types of patterns into consideration when scheduling maintenance. From the interviews we found that planners try to minimize AOG time, which means maximizing the maintenance performed each time the plane is taken in for repair. Coordinating maintenance on multiple components at specific intervals involves assessing the failure rate of each component, a process that involves a level of uncertainty. Planners also need to know whether replacement parts will be available when needed, and whether maintenance personnel are available and possess the relevant skills to do the maintenance. There is also the issue of hangar space availability. Finally, aircraft need to be in the proper city at the allotted time for a scheduled maintenance check. Planners manage all of these uncertain factors, and the specific techniques they use warrants further investigation.

3.4.3 Pilot Study 1 Findings: Managing People

One of the planners stressed that the most uncertainty arose in the tasks where he had to deal with other people. Specifically he mentioned that understanding the perspectives of maintenance personnel in relation to what maintenance actions they were taking was very difficult. Through experience he found that factors such as shift time influenced the productivity of maintenance personnel, not only because of physical exhaustion but also due to maintenance personnel changing the amount of work they did based on the relation between hours worked and anticipated time off.

The second planner also noted that there is an instance of uncertainty during shift changes of maintenance engineers. When a shift change occurs during a long maintenance check or repair there is a significant reduction in productivity. This is due to the time it takes for briefing the incoming engineer on what work has been done and how much has been done. During this hand-off there are numerous opportunities for data to be incorrectly entered. The uncertainty inherent with a hand-off can affect later analyses of these reports by both safety analysts and maintenance planners. Through experience the maintenance planner learns to avoid scheduling large and important maintenance procedures when there are personnel changes.

3.4.4 Pilot Study 1 Findings: Reliability of Data Sources

In order to fully investigate a component failure or any other incident causing unscheduled maintenance, planners rely on additional data. One source of data comes from incident reports filed by both pilots and maintenance

engineers. After each maintenance action, whether scheduled or unscheduled, a report has to be filed. This is also true for any incidents the pilots encounter before, during or after flights that may require maintenance.

When the need arises for a maintenance planner to investigate an incident or component failure the first task is often searching through incident and maintenance reports to find similar occurrences. There are two obstacles when using these types of reports: incompleteness and incorrectness. Occasionally the reports are not entirely accurate, whether due to an honest mistake of the person making the report or due to intentional deception. There are currently some systems in place at the point of entry designed to help with data verification, however once filed it is difficult to tell whether these techniques were used. There is sometimes a significant amount of incomplete or missing data as well.

The third planner felt that incomplete and incorrect data were two of the most uncertain factors in his work. When asked what kind of strategies he had for dealing with uncertain data he outlined one main technique: the use of standard deviation charts for incidents and component failures. Effectively, this would be a range of values within an acceptable limit represented by a graph with an upper and lower threshold. Once a value appears that is outside the standard deviation threshold, the planner is alerted. The analyst sketched a typical chart and explained what happens during this kind of analysis. Planners determine the acceptable limit every twelve months based on the previous year's incidents. The visualization of the chart and the methods for adjusting its values will be a focus for further research.

3.4.5 Pilot Study 1: Discussion

Before visualizing uncertainty it is necessary to understand how people conceptualize uncertainty. It is still an ill-defined concept in many domains, with the distinction between related concepts such as data quality, reliability, accuracy, and error remaining ambiguous (MacEachren et al., 2005). As we have found from the aircraft maintenance planners, there are at least three main areas where uncertainty appears: In the aircraft components themselves, managing people and finally with the reliability of reports and data. All three of these areas require vastly different strategies to manage uncertainty.

There is also still work to be done in determining whether and how uncertainty visualizations help in the decision-making process. Recently there has been a movement towards visualizing the “black box” processes going on in analysis, which would include the way that people manage uncertainty (Amar & Stasko, 2005). The human visual system can be a powerful aid in analysis if visualizations are designed to cater to its strengths and avoid its weaknesses. In many domains uncertainty visualizations are still in early stages and not well understood, leaving the potential for significant research to be done.

3.5 Pilot Study 2: Paired Analysis at Aeroinfo

Following the interviews in pilot study 1 the research objective was modified. The focus no longer revolved solely around uncertainty. The research scope was widened to include more of the analytical reasoning process in order to take advantage of the rich data that would be collected from paired analysis.

For pilot study 2, the researchers and SMEs worked towards and ultimately defined a task to explore maintenance data using paired analysis. The task involved looking at the unscheduled aircraft downtime (UAD) present in maintenance data of one particular airline.

3.5.1 Pilot Study 2: Paired Analysis Sessions

The pilot study was conducted at Aeroinfo's offices over four separate paired analysis sessions. In the first session, the Business Development Manager (BDM) served as the SME. As with the exploratory interviews, the BDM provided important initial data to the expert tool user. In this case, the BDM's paired analysis session allowed us to work out issues with the paired analysis setup, including camera placement, audio recording, pacing and length of the session. In each of the following three sessions, a single aircraft maintenance planner served as the SME. The visual analytics tool experts included the author and another graduate student.

Prior to paired analysis, the airline's maintenance data was loaded into the visual analytics tool Tableau. Before actually connecting to the dataset with Tableau, a large amount of data preparation occurred. This consisted largely of connecting and parsing relevant tables and narrowing down the fields in the 90,000 records. Once the data was loaded into the tool the paired analysis sessions could begin.

The four sessions were conducted over a period of 4 weeks, and averaged 2 hours each. Prior to each session we briefed the SMEs about the

structure of the dataset and the nature of the task (Arias-Hernandez et al., 2011). Each session was captured via screen capture software, video cameras and audio recordings.

3.5.2 Pilot Study 2: Findings and Discussion

Following data collection, Arias-Hernandez et al. (2011) transcribed and then coded the recordings from the four paired analysis sessions for further analysis. Using Joint Activity Theory, Arias-Hernandez et al. (2011) found that paired analysis participants (both the SME and TE) coordinated their attention, cognitive workload and navigation of analysis. Please refer to the published article for a more detailed description of findings. Through examination of these three phenomena, researchers provided evidence that the paired analysis method is successful at capturing cognitive processes effectively.

Several important lessons were learned from the paired analysis pilots at Aeroinfo that have helped to refine the method for use at Boeing. These are discussed in the following paragraphs.

First, the selection of the SME is very important to the overall success of the session. As previously discussed, this includes factors like personality type, knowledge of the data and technological capabilities. Establishing common ground between the tool expert and SME is of utmost importance, as well as being able to have dialogue back and forth without constant confusion. Selection of SMEs for future paired analysis sessions should be done according to the results. This may include interviews, case studies and research on the domain of

the SME. The tool expert and SME should be aware of each other's skills and limitations to some degree, and should not meet for the first time on the day of paired analysis.

A second lesson is the importance of capturing participant note taking. During the Aeroinfo sessions, the SME and TE frequently tracked their progress and items of interest. Making sure participant note taking is captured requires the use of the screen-capture software if digital tools are used, or a camera pointing at the desk if paper notes are recorded. An important part of the sensemaking loop for analysts (Pirulli & Card, 2005) involves the shoebox and evidence file where data is collected for later use. Capturing participant note taking during paired analysis will help to understand this aspect of the sensemaking loop.

Finally, the pilot studies provided an idea of how to judge the length of time that is ideal for paired analysis. The ideal length of time is ultimately determined through a balancing act, weighing both the analytic task and the amount of work to analyze the sessions afterward. Due to the labor-intensive nature of transcribing and coding, the length of a session should be kept to a minimum. However, the session must also be long enough to give the SME and TE ample time to approach the task. Therefore, understanding the task and data is again very important.

4: PAIRED ANALYSIS FOR AVIATION SAFETY

4.1 Introduction

Following pilot studies at Aeroinfo, a second domain was examined using paired analysis and visual analytic tools. This domain was aircraft safety engineering at Boeing. Research with aviation safety engineering SMEs was carried out in a similar manner to Aeroinfo pilot studies, taking into consideration lessons learned about the paired analysis method.

Aviation safety research carried out with Boeing occurred in the following phases: exploratory interviews, internship, selection of SME, selection of a task, selection of a dataset, selection of a tool, data cleaning and preparation and finally paired analysis sessions. Each of these parts is described in more detail throughout this chapter. Chapter 5 discusses the results and analysis of the paired analysis sessions.

During paired analysis sessions, the author served as the tool expert (TE). The author's expertise focused on the use of several visual analytics tools, two of which were used during the internship: Tableau and IN-SPIRE. The SME for the paired analysis sessions is a safety engineer with extensive experience in aviation safety.

4.2 Exploratory Interviews

Exploratory interviews of safety engineers were conducted to learn more about the data, domain and tasks performed. Increasing the TE's knowledge of the domain of aircraft safety engineering was thought to help facilitate paired analysis. Interviews were designed to help create a common ground between researcher and safety engineer that would allow for a smoother implementation of paired analysis. The primary objectives of the interviews were:

- Introducing the visual analytics tool expert to the subject matter experts, thereby establishing an ongoing relationship.
- Providing the tool expert with an idea of the daily work procedures and problems of safety engineers.
- Letting the tool expert examine safety engineering data, thereby allowing a preliminary assessment of which visual analytics tools are appropriate.
- Exposing the subject matter experts to a few of the visual analytic tools and setting reasonable expectations for their application to safety engineering data.

Due to time, space and security constraints, it was decided that a group interview format would be used for the exploratory interviews. This involved four researchers from SFU, two of whom were visual analytics tool experts. In five sessions spread over the course of three days, five safety engineers were interviewed. During each session the safety engineers gave a presentation of their work to the researchers.

To guide sessions with safety engineers, a focus structure containing questions of interest was used (See Appendix B: Boeing Safety Engineer Focus

Structure). These questions were organized into several sections, including a general introduction, basic employee and background information, description of a typical day, elaboration of specific tasks, description of current tools used, description of data, collaboration with other analysts and reporting of results from analysts. The focus structure was designed using qualitative research techniques designed to elicit rich responses from the participants (Patton, 2002).

The focus structure was designed more as an outline rather than a verbatim set of questions in which every one was to be asked by the researcher. If the interview session digressed from eliciting useful information, the focus structure was used to get back on track. It was also designed to remind the researchers of which general areas were to be covered for each session.

Due to security policies, the researchers were not able to video or audio record sessions with safety engineers. Instead each researcher took notes. After all of the sessions were completed, the notes from all of the researchers were compiled into a single document summarizing findings. This document was reviewed by the Boeing safety engineers to remove any specific names or proprietary information.

The exploratory interviews with Boeing safety engineers succeeded in several ways. First and foremost, they helped to attain common ground between the visual analytics tool expert and the subject matter expert. The tool expert was exposed heavily to the domain of aircraft safety engineering, and the subject matter expert became more aware of what was possible with their data using visual analytics tools. In addition to establishing common ground, the interviews

helped to establish a level of trust between the subject matter expert and the researchers. As discussed, this is important for the dynamic of working together closely during paired analysis. The interviews provided the tool expert a good estimate of how long it would take to clean and prepare the data for the visual analytics tools and how to factor this into the paired analysis process.

Exploratory interviews also contained discussions of data and tasks to use for future paired analysis sessions. From experience gained during the Aeroinfo paired analysis pilots, it is very important to select good data and tasks for paired analysis. Selection of unfamiliar data or a poor task can result in a disruption to the analytic workflow. Data collected from the Aeroinfo sessions reflected this point. In a few sessions the analysis was stalled due to lack of knowledge regarding the data. With regard to aviation safety, several areas of application for visual analytics were discussed. The potential topics discussed for future paired analysis sessions included “Pitot Probe Icing”, “Bird Strikes”, “Lightning”, “Turbulence” and “Runway Excursions”. These topics were selected based on their relevance to Boeing, the data available and an estimation of time for tasks concerning each one. All of these topics were under investigation by Boeing safety analysts during the time of the interviews and internship.

From interviews with safety engineers we learned that the data was often messy and located in many places. It was also evident that all of these subject areas would require a phase prior to analysis where the tool expert became familiar with the data.

Interviews also aided in the selection of an SME for paired analysis. One SME in particular was very eager to begin analysis using visual analytics tools. During the internship this SME was selected as the participant in the sessions.

4.3 Boeing: Paired Analysis Sessions

4.3.1 Overview

Following the interviews an internship was carried out on site at Boeing. Before paired analysis could begin, there was a need for a common understanding of both the task and the data by the SME and TE. A common understanding helped minimize communication errors between the two participants. It also helped for judging the appropriate length of time for each paired analysis session. The process followed during the internship is outlined below:

1. SME creates a list of tasks that are relevant to his domain and that he feels are possible to attempt using visual analytics.
2. The SME and TE engage in dialogue to determine which tasks are most appropriate given the available tools, data, and timeframe.
3. A single task is chosen.
4. A dataset is selected that can fulfill the needs of the task.
5. The TE determines which tool is appropriate for the task and data.
6. The TE cleans and prepares the dataset for ingest into the selected tool.

7. Once the dataset has been ingested successfully, the TE and SME use paired analysis to approach the task while being recorded.

4.3.2 Choosing a Task

Based on the analysis of the exploratory interviews there were many potential tasks. These included but were not limited to: Pitot probe icing, runway excursions, volcanic ash, lightning, turbulence, kinetic energy of bird strikes, narrow body vs. wide body aircraft strikes and pilot response to bird strikes. In order to assess whether they would be appropriate for paired analysis, the TE was exposed to all of the data repositories at Boeing, and learned where data about each of these safety issues was located. Most of the safety issues had overlapping data sources, spanning both publicly available and proprietary datasets.

Before settling on a task for paired analysis, the SME provided a list of goals for the internship. This contained a wide variety of topics and specific questions for analysis. This list was constantly refined and updated throughout the internship and served as a guiding document. The list can be found in Appendix C: Boeing Internship Task Sheet. This task sheet was referenced throughout the internship and served as a guiding document for

The search for a task was narrowed by both the datasets available and what was particularly relevant to the safety engineers at the time. Because the internship was on-site alongside active safety engineers, the relevant issues were easier to identify. For example, the issue of “runway excursions” had

already been examined extensively by our Boeing challenge team at SFU and was therefore eliminated from contention. “Lightning” and “Turbulence” were also eliminated on the grounds of being less relevant to Boeing at the time and the difficulty of accessing data sources. The two that remained will be discussed in greater detail: Pitot Probe Icing/Unreliable Airspeed, and Birdstrikes.

The SME felt that these were the two most relevant tasks to choose from, as both are ongoing analyses being performed by safety engineers and very relevant to airplane safety. The SME and TE analyzed the pros and cons of both issues.

For unreliable airspeed:

- More of an open-ended approach
- Have to figure out what is useful
- Looking at well known incidents
- Chaotic process
- Iterative
- Broaden search from what has been done in the past

For Birdstrikes:

- Not as exploratory, more focused
- Relatively clean data
- Opportunity to compare to previous analyses using excel
- Easier to see deficiencies in data
- Know what factors play a role (altitude, species, airspeed and damage)

Unreliable airspeed would largely be a task that began from square one, as it had only recently been discovered to be a major issue. There are both pros and cons to a new analysis. A new task that the SME has no previous

experience analyzing would yield interesting data about the analysis process as it initially develops. It seems likely that beginning with a new problem and not as much familiarity with the data would lead to a much more exploratory process. However, this could also lead to roadblocks during analysis and no clear definition of when the task was completed. This is in contrast to the issue of bird strikes, in which the SME already has significant prior knowledge of many of the factors involved with bird strikes. The SME has also completed several bird strike analyses in the past. The SME's prior knowledge and analyses of bird strikes can be viewed as beneficial for evaluating the visual analytics tool, as he would be able to comment on the differences between the traditional method and the new method using visual analytics. There was also a lower chance of the analysis stalling due to direct familiarity with the subject matter by the SME.

In total, there were nine paired analysis sessions recorded based on several of the tasks from the initial list. However, for the purposes of this thesis, two of those sessions have been further analyzed. The two selected sessions involved the same task: *Examining pilot response to bird strikes*. The reasons for selection of this task are summarized below:

- The SME had pursued this line of analysis before using in-house tools. This was an advantage because the SME had familiarity with the data and the types of obstacles involved with performing this type of analysis. Also, the SME would be able to comment on the differences between using paired analysis/visual analytics and previous methods.
- A large portion of the data was publicly available. Because this task was going to be analyzed for academic purposes, minimizing proprietary data sources was key. Much of Boeing and their customer's

data – read “airlines” – is proprietary and sensitive and there are restrictions on publishing both the data and any findings gained from analysis.

- The data contained both structured fields and unstructured text. The data was relatively “clean” as far as large datasets go.
- The task involves looking at a relatively large dataset that, without tools, would take a human a long time to analyze. The TE wanted to choose a task where the data was appropriate for the use of visual analytics tools, and this data fit the bill.
- The task involved “set goals” while also encouraging data exploration. In visual analytics, one of the main goals is to “detect the expected and discover the unexpected” (Thomas & Cook, 2005). While the SME was looking for specific records (detectable strikes), the nature of the unstructured text found in the remarks field of each report allowed for further exploration and learning.
- The task could be broken into manageable chunks. For paired analysis, controlled sessions are needed. Without being able to break the analysis down into smaller parts, the sessions would go on endlessly and there would be far too much video/audio data collected to analyze in a reasonable amount of time. However, sessions also need to be long enough to analyze the SME/TE interactions and interactions with the tool. Finding the happy medium is key, and this task seemed likely to fit this bill.

4.3.2.1 Defining the Task

The SME was interested in several aspects of pilot response. Much of the SME’s analysis was inspired by the human performance models put forth by Rasmussen (1983), who proposed different stages of human response to unfamiliar conditions during routine tasks. The SME developed his own taxonomy based on Rasmussen’s model.

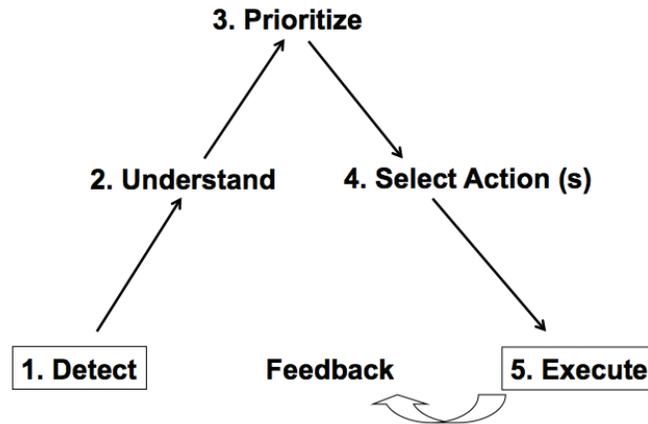


Figure 13: Five Stages of Response to Novel Situations

Figure 13 illustrates the five stages that the SME was interested in examining. They are detection, understanding, prioritization, selection of action and execution. In some cases the SME was also interested in the feedback the flight crew received post-execution of action, and what additional actions this led to. These stages are described in more detail below.

First, the SME wanted to know whether the flight crew detected the bird strike during the flight. Sometimes the strike is not detected when it occurs, and is instead found much later when evidence of the strike is discovered by maintenance crews. For the purposes of analysis, the SME was interested in finding only those reports that were detected by the flight crew, whether in flight or while taxiing.

Next, the SME wanted to know how the detection occurred. Many bird strike detections occur via visual cues, such as seeing the bird strike the windshield. These are the most obvious strikes, as the flight crew has visual evidence of what happened. However, other strikes go unnoticed visually, and

instead are detected via sound. Others are detected by smell, vibration or unusual behavior from the plane. The method of detection can influence whether the flight crew responds in the best way, making it of interest to the SME. Terms relating to detection mechanisms were searched to find the strikes that were detectable in flight.

Once records where the bird strike was detectable have been isolated, the SME was interested in whether the flight crew understood the extent of the damage. Understanding effects of the strike allow the flight crew to make a decision about how to respond, which is the next thing the SME wanted to know.

Selecting an action for response to a bird strike is done based on the best available information to the flight crew, and can vary largely depending on the method of detection and comprehension of the extent of damage. For example, “Did the flight crew choose an action and is this present in the report?”

Finally, once the flight crew had selected an action, the SME was interested in whether it was executed. Merely selecting a response to the strike does not mean it was executed. Other things may have occurred during flight such as distractions on the flight deck, or simply that the crew decided that the action selected was unnecessary. Ultimately, the SME was interested in examining the executed action or inaction for whether it was the best choice given the information.

4.3.3 Choosing a Dataset

In order to assess these criteria it was necessary to examine reports of past incidents with birds. There are several databases within Boeing as well as external sources containing incident reports related to bird strikes. The core part of these reports is the 'remarks' field (or equivalent), where the incident is described using unstructured text.

NAR: SECONDS AFTER ROTATION WE HIT A BIRD. THE BIRD STRUCK THE AIRPLANE WITH AN **AUDIBLE THUD** AND STRUCK THE FO'S WINDOW. HE WAS FLYING. **I MONITORED ENG INDICATIONS** CLOSELY AND DETECTED NO SIGN OF ABNORMALITY. WE CONTINUED THE CLB ON PROFILE. A FEW SECONDS LATER A STRONG SMELL OF SMOKE ENTERED THE COCKPIT AND PERSISTED. FO CONFIRMED THAT ALL ENG INDICATIONS WERE NORMAL AND CONTINUED STANDARD CLB. **SMELL AND SMOKE** WERE STRONG. I MADE A **DECISION TO RETURN** TO ZZZ DUE TO SMOKE AND SMELL. CALLED ZZZ AND DECLARED AN EMER, ANNOUNCED BIRD STRIKE AND REQUESTED TO RETURN TO ZZZ.

Figure 14: Pilot Narrative Description of Incident with Important Words Highlighted

Figure 14 is an example of a 'remarks' field from a single report. The text highlighted in red indicates portions relevant to the SME's analysis using the Rasmussen model. This includes things like how the bird strike was detected (indicating it was actually detected), whether there was damage and what actions were considered and executed. The unstructured text of the remarks field is also rich with other information such as the amount of time that passed between events and decisions. This type of information is valuable during the exploratory parts of analysis and was also recorded in the paired analysis sessions.

In addition to unstructured text in the remarks field, reports contained other information that was necessary to process and analyze. Additional fields include things like airport where the strike occurred, speed of the aircraft, model of plane...etc. There were also many fields containing information irrelevant to our process. The SME's expertise and ongoing discussions with the TE helped identify which fields from the dataset were to be used and which were to be excluded.

There are hundreds of thousands of airplane incident reports within Boeing and external sources such as the FAA and ASRS. Boeing's data includes proprietary in-house databases, customer databases, government data and publicly available data from sources such as the FAA. However, the bird strike incidents within these sources are not always detectable bird strike incidents. In fact, most are not even bird strike incidents at all. Rather than overwhelm the visual analytics tool with both bird strike and non-bird strike incidents all at once, the SME and TE worked together to find a subset of the data containing only bird strikes before the paired analysis sessions began.

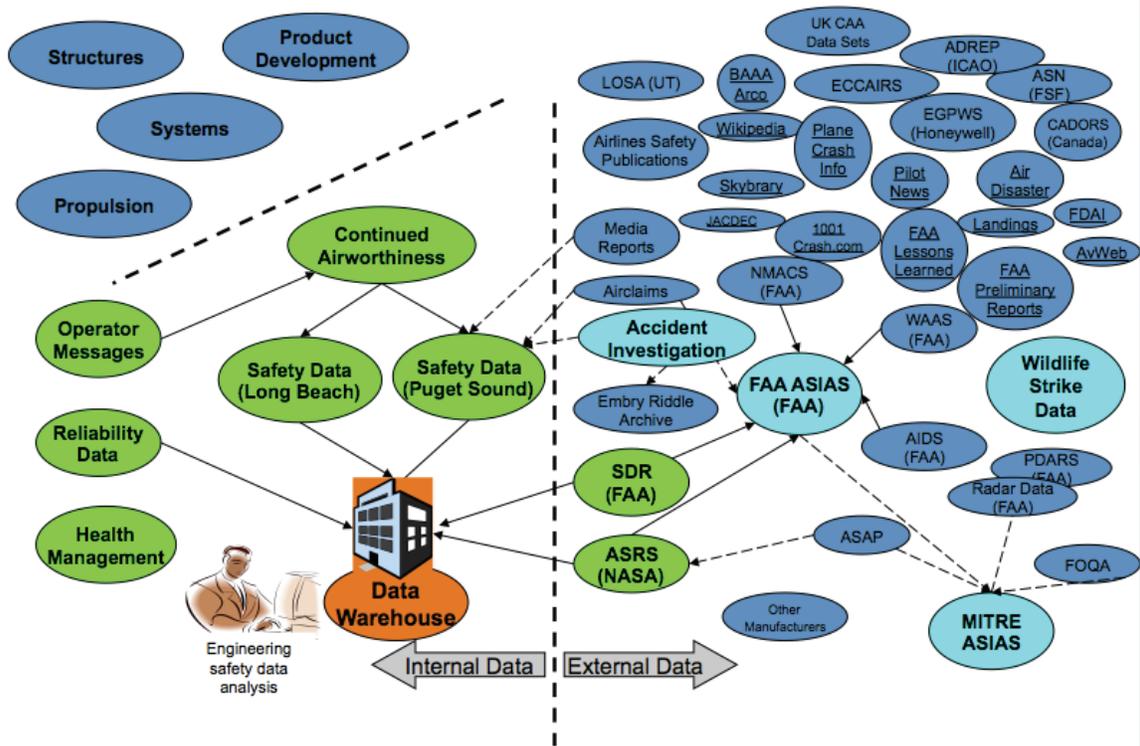


Figure 15: Subset of Boeing Safety Databases

Figure 15 shows only a portion of the existing safety databases examined by Boeing safety engineers. Working with the SME, the TE helped select the data that would be best for the pilot response task. Ultimately two databases were chosen. The first was selected from an in-house database consisting of all different kinds of incident reports. Both bird strike and non-bird strike events were in the original data. However the SME had familiarity with this dataset from previous analyses and had already cleaned it significantly. The result was a subset of 1087 documents involving bird strikes to commercial aircraft. The SME provided this subset to the TE to prepare for ingest into the tool.

The second database was the publicly available FAA National Wildlife Strike Database. This database contained tens of thousands of reported bird and other wildlife strikes to all types of aircraft and is available through the FAA's website.

4.3.4 Choosing a Tool

Each dataset contained both structured and unstructured fields. Due to the nature of the task, the most important information was contained in the unstructured 'remarks' field in each report. However, several other fields within each dataset were also relevant. Data relevant to assessing and analyzing the bird strike hazard to commercial jets fell into a few categories:

- Numerical Data: This included things like the aircraft's altitude and speed as well as an estimated cost of repair.
- Categorical Data: For example, the aircraft type, operator, damage to the plane and species of bird.
- Unstructured Text: For example, the narrative of the incident found in a report by maintenance personnel or the flight crew.

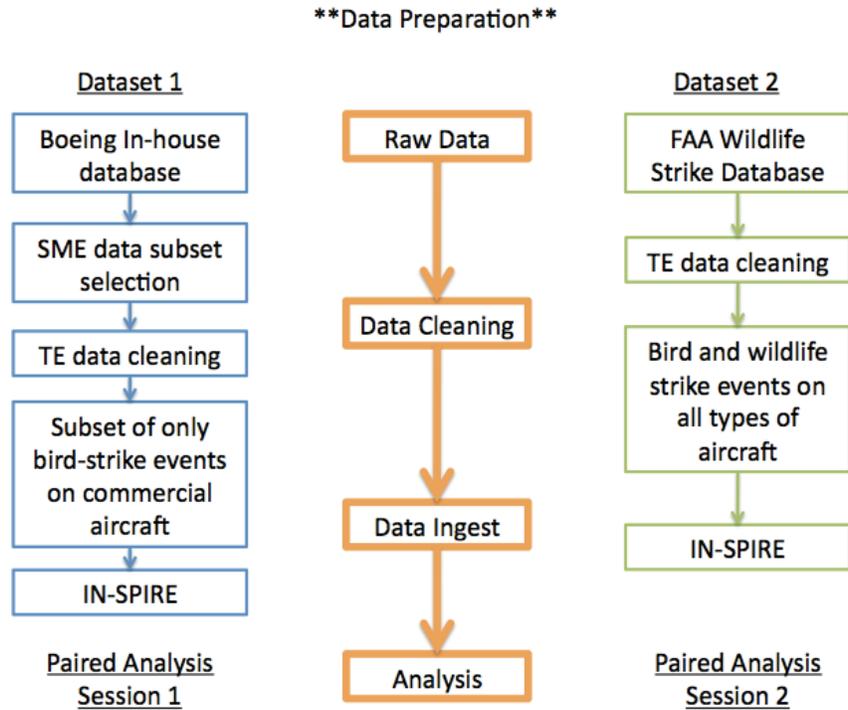
These three types of data were present in multiple databases, including the publicly available National Wildlife Strike Database. There are several visual analytics tools that address some or all of the types of this data. It was therefore important to understand in more detail what some of the tasks were going to be. A task list evolved over the course of the internship based on input from the SME and was very valuable for tool selection. The TE selected a tool based on a few different criteria:

- The most important feature of an eligible tool was the tool's ability to handle unstructured text. From interviews, discussions with the SME and the task sheet, it was clear that analysis usually involved a significant amount of reading. A tool that could process text and help minimize reading would be the most beneficial to the SME and analysis.
- The tool also needed the ability to handle structured fields, such as those of model, aircraft speed and airport. Having the power to group, sort and filter the data by these additional fields would provide additional benefit during analysis.
- The tool had to be able to handle tens of thousands of records at one time. Though the in-house data subset was relatively small (1087 records), the FAA dataset contained tens of thousands of records. A tool that can both process and analyze a dataset of this size without showing performance issues was necessary.

With this knowledge and after discussing the task with the SME, the TE made a decision of which visual analytics tool to use. Because of its ability to handle unstructured text, IN-SPIRE was selected for both sessions and both databases. The TE's previous experience using IN-SPIRE was used as a basis for meeting the criteria outlined above.

4.3.5 Data Cleaning and Preparation

Both datasets required similar steps for data cleanup and preparation for import into IN-SPIRE. In short, this involved conversion of the data from its tabular, spreadsheet formats (the in-house dataset from Excel and the FAA dataset from Access) to a delimited text file.



In the figure above the process is broken down into 4 steps. Beginning with the initial databases, the data was cleaned, then ingested into the tool and finally analyzed using paired analysis. One noticeable difference is that for the in-house database the SME preselected a subset of the data. Because the bird strike analysis was ongoing, the SME and TE decided to use an existing subset of the data. In the case of the FAA data, the entire dataset was used. In both datasets, the data-cleaning phase performed by the TE followed a few different steps:

1. Selection of relevant fields and exclusion of irrelevant fields
2. Conversion from Access/Excel to XML
3. Inspection for correctly parsed fields

4. If the fields were incorrectly parsed, the Universal Parsing Agent (UPA) was used to correct them
5. Moving “remarks” field to the top of each XML document using UPA (The “remarks” field was of the most interest, so its placement at the top of the record allowed for easier access)
6. Import into IN-SPIRE

Step 5 deserves additional elaboration, as this was a very important step born from experience of the TE and knowledge of the SME’s domain. When examining a single document using IN-SPIRE, the order of the fields appears the same as the source documents. However, as previously stated, the most important information for the SME’s task was in the ‘remarks’ field. In the original form of both datasets, the ‘remarks’ field appeared after several of the other fields. In order to enable easier reading for the SME, the TE moved this field to the top of each document.

4.3.6 Initiating Sessions

Common practice of safety engineers is to engage in an assigned analysis until it is complete. For example, if the safety engineer was interested in the effects of volcanic ash on the engines of planes they would begin by locating all records of interest, performing the analysis, writing up the results and finally presenting them to a superior. However, this process can take a wide range of time spanning from a few hours to several days depending on the task. For the

purposes of paired analysis, relatively small chunks of analysis were deemed ideal for recording sessions.

In several instances throughout the internship, the SME and TE engaged in impromptu analyses that were often initiated from something new the SME thought of on the spot. These “desk sessions” were great tests of the visual analytics tools, however they proved troublesome for recording the paired analysis. In many of these instances, the SME and TE would be well into analysis before they realized they should have recorded the session. However, due to the nature of the setting and security concerns, “desk sessions” could not be recorded. These environmental factors played a large role in determining which sessions were actually recorded.

To record the paired analysis sessions, a conference room was often booked for up to 2 hours. The SME and TE then set up the computer and cameras and began working on a predetermined task.

4.3.7 Video Recording

In order to record the paired analysis sessions for later analysis in the lab, a combination of screen, audio and video captures was used. The tools used for capturing the sessions included a single tripod-mounted video camera and the screen-capturing software ‘Camtasia’. The goal was to capture several aspects of the paired analysis sessions:

- Interactions between the SME and TE during analysis. This includes gestures, pointing and where the attention is focused. The

collaborative nature of paired analysis elicits many interactions and to examine sensemaking these interactions must be captured.

- Additional objects being used. For example, “are the TE or SME taking notes or working from documents on the table?”
- The screen where the software is being used. Although there is a simultaneous screen recording being captured, ability to see the screen in the video provides an additional reference point.

The primary goal of the video recording was to capture the SME and TE’s dialogue, mannerisms and gestures while working with the tool during the paired analysis. It was therefore important to position the camera so that it could capture the TE, SME and the screen where the interaction with the tool was taking place. This way, actions like pointing at the screen would be captured in addition to mannerisms and gestures between the SME and TE. The camera was setup on top of a table and captured a side view of the actions of the SME and TE as depicted in Figure 17.

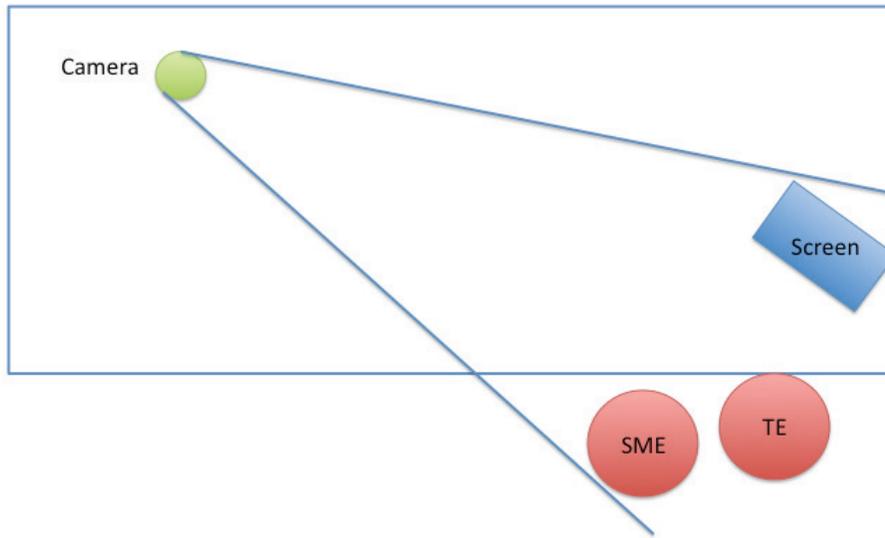


Figure 17: Camera Setup

With this setup the camera had a clear view of the front of the SME and TE as well as activity on the screen. Gestures such as 'pointing' were easily visible and the location on the screen where participants were pointing was also easy to see due to the high resolution (1080p) of the camera (See Figure 18).

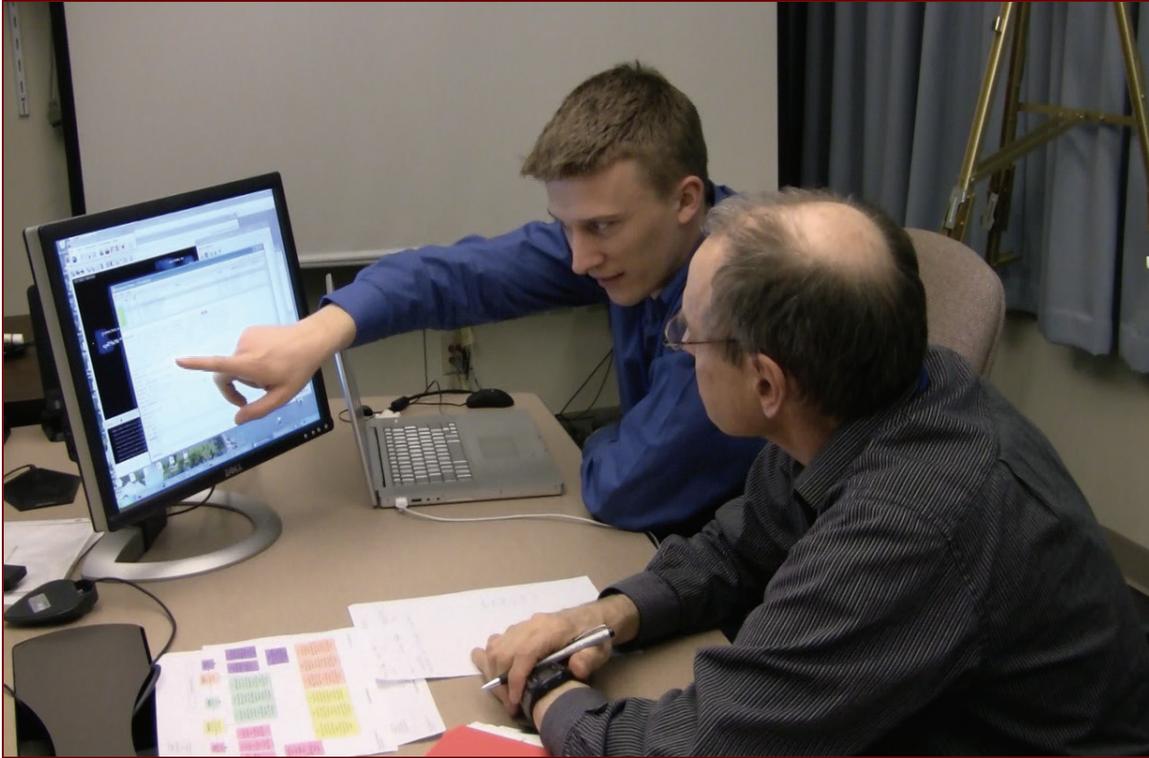


Figure 18: Paired Analysis Session Video Screenshot

The camera captured video and audio of the participants and a sample of what was going on with the tool. However, a detailed screen capture was also being recorded by software on the computer called Camtasia. Camtasia captured the actions on screen as well as a backup audio track of the SME and TE's discussions. The idea behind capturing the screen is to be able to do a detailed analysis of how the tool was being used after the paired analysis session finished. The combination of video, audio and screen recordings provided rich data to analyze the method of paired analysis.

5: BOEING: RESULTS AND ANALYSIS

5.1 Overview

In total, we conducted nine paired analysis sessions. These were captured using video, audio and screen recordings. The nine sessions were distributed as follows:

- 6 using IN-SPIRE
- 1 using Tableau
- 2 using both IN-SPIRE and Tableau

Out of the nine sessions that were recorded, eight were related to bird strike risk assessment. The single, unrelated session used a different SME than the other 8, and was related to funding potential research areas within the company. This session will not be discussed in this thesis. The 9 sessions spanned a period of 37 days, and some were a continuation of a task begun in an earlier session. The average length of time for one session was just shy of 68 minutes. A typical session included time at the beginning for establishing the task as well as time at the end for discussion of what had been accomplished and what would be attempted next time.

For the purposes of this thesis, two sessions were selected for deeper analysis. Many factors went into selecting just two, and those will be discussed in the following section. In order to keep analysis more manageable, the raw video

and audio from these sessions was not analyzed. Instead, each session was transcribed for further analysis.

5.2 Choosing Sessions for Further Analysis

From the nine paired analysis sessions captured, two were transcribed and analyzed further. Selecting the “best” two for transcription was a multi-step process.

One of the research objectives involved looking at the method of paired analysis as a whole and evaluating its effectiveness in capturing the reasoning process. There was also interest in building upon the sensemaking model and incorporating aspects of the collaborative method of paired analysis into the model. Finally, the method was to be examined for how well it facilitated tool assessment and learning on the part of the SME. All of these factors were considered when selecting two sessions for transcription.

Selection of two sessions provided a means for comparison. This was useful when trying to determine whether particular phenomena are isolated events or whether they occur in other sessions. This also enabled an examination of progression and learning between the sessions. As one session would take place after the other, the transcripts can be analyzed for learning on the part of the tool expert and SME. For the tool expert, learning would consist of becoming more familiar with the domain and data. For the SME learning would consist of becoming more familiar with the tool and its capabilities.

Initially, each of the nine sessions was viewed in its entirety while taking notes on the contents and tools used. This information was compiled into a spreadsheet that can be found in Appendix D: Collected Data Description. There were also columns for the length of each session, the file name, notes, and a column for what the initial “gut feeling” was on using the session for transcription. Once each session had been entered into the data table, candidates for transcription began to emerge. Certain sessions could be eliminated immediately due to the flow of analysis being broken up by distractions, or being a session where a very repetitive task (like data cleaning) occurred for the majority of the time. Ideal candidates for further study were sessions where a more complete analytic process occurred, going from a task and raw data to at least some findings and conclusions.

Using this method, the two sessions selected were the third and the sixth recorded session in the series of nine sessions. For the remainder of this thesis, these sessions will be referred to as “Session 3” and “Session 6” respectively. These sessions took place 22 days apart from one another. Both involved the use of IN-SPIRE and both examined pilot response to bird strikes. While the task was the same, the dataset used was different. For Session 3, an “in-house” database was used to examine past bird strikes. For Session 6, a publicly available database from the FAA’s Wildlife Strike Database was used. These two sessions were selected for several reasons:

- Same task and tool, different data - Performing the same task with a different dataset using the same tool was hypothesized to be a useful way to examine learning and progression with the tool. By comparing

dialogue and interactions of the two sessions there was a hope of understanding retention of knowledge during the 22 day break.

- Length of the sessions - The duration of session 3 was approximately 104 minutes and session 6 was about 70 minutes. This provided plenty of time to capture the analytic process while also helping to limit the amount of work required for transcribing, coding and then analyzing the sessions.
- Task during the sessions – The task of examining pilot response to bird strikes showed evidence of many steps of the sensemaking process during the initial viewing, which was desired for further analysis.

The two selected videos with the task of analyzing pilot response to bird strike were first transcribed to text files. Looking ahead to future analysis and to enable error checking, the text transcriptions were hyperlinked to the videos they originated from at periodic intervals. This was achieved using the software “f5”, and the end result was 2 linked transcripts. With linked transcripts it is possible to highlight a portion of text and immediately be taken to that point in the video. Linked transcripts are important for a couple of reasons. First, they enable error checking. Simply reading text of a discussion can sometimes be misleading or confusing. The intention and meaning is not always as clear as when the original audio or video is examined. Therefore, being able to quickly reference where the comment was made in the original video proved useful. Referencing the video occurred on a number of occasions during analysis. A second reason is for future analysis. The video files provide a wealth of additional data and having them linked to the transcripts will save time. Future research may make use of the additional audio and video streams. However, for the purposes of this analysis

the primary data source was the transcripts. The two sessions and their respective statistics are shown in the table:

Table 2: Paired Analysis Session Details

	<u><i>Session 3</i></u>	<u><i>Session 6</i></u>
Date	Feb. 25, 2010	March 19, 2010
Length (Minutes)	105:06	71:05
Primary Tool Used	IN-SPIRE	IN-SPIRE
Primary Task	Examine Pilot Response to Bird Strikes	Examine Pilot Response to Bird Strikes
Dataset	ASIS bird strike subset	FAA’s NWSDB
Data Captured	Video, Audio, Screen	Video, Audio, Screen

5.3 Analysis

5.3.1 Coding and Memos

Analyzing the paired analysis session transcripts was a highly iterative process. It involved several passes of assigning codes to the transcripts from Sessions 3 and 6. “Coding” is a technique developed to help create theory out of data (Strauss & Corbin, 1998). When reading the transcripts, “codes” are added to provide a marker and description of a particular passage or word in the text. The goal of the interplay between the researcher and data is to be as objective as possible, using the techniques of both open and axial coding (Strauss &

Corbin, 1998). Open coding refers to conceptualizing on the first level of abstraction, meaning the text of the transcription is analyzed line by line and concepts that emerge are coded (Strauss & Corbin, 1998). While performing open coding, the researcher strives to minimize the influence of prior conceptualizations when creating new concepts and categories. However, as it is nearly impossible to code with a “blank slate”, it is more important that the researcher understand when and how prior conceptualizations are affecting coding. Open coding generally provides a large number of initial codes that are later grouped together, recategorized and renamed as the analysis proceeds. Axial coding, on the other hand, refers to “the process of relating categories to their subcategories, termed ‘axial’ because coding occurs around the axis of a category, linking categories at the level of properties and dimensions” (Strauss & Corbin, 1998). In Strauss and Corbin’s model of Grounded Theory, axial coding generally occurs after open coding and helps to provide structure to the fragmented data collected during open coding. Both techniques were used during analysis of the transcripts.

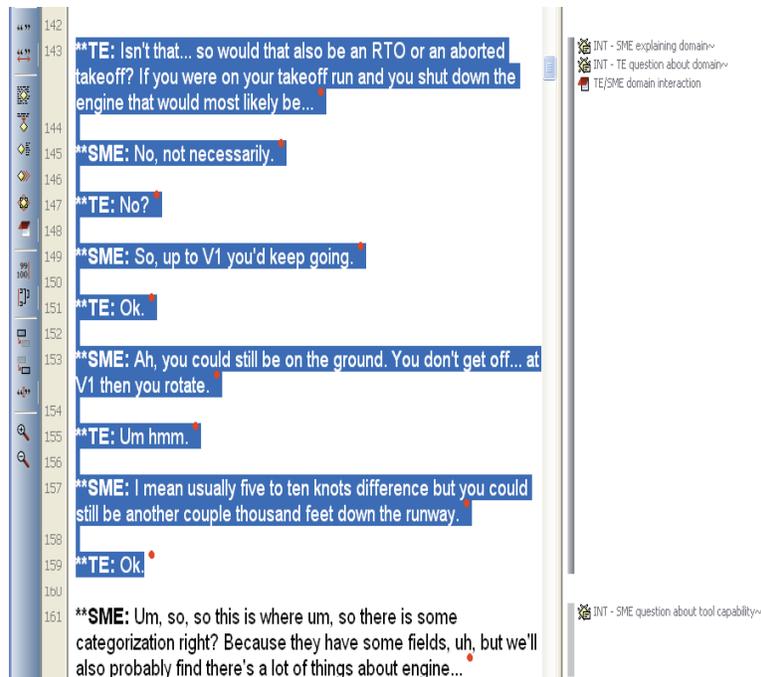


Figure 19: Coding Transcripts with Atlas.ti

Figure 19 shows the software Atlas.ti, which was used for coding the transcripts of sessions 3 and 6. Atlas.ti is a qualitative analysis software platform that allows you to transcribe and code many different file types. It also allows you to build hierarchies of codes and attach memos during analysis. With each additional pass, more codes were added to the transcripts. Eventually, patterns of codes began to emerge and codes were grouped together.

Memos were used when more description was needed. For example, when I noticed there was a unique code or one that was particularly attention grabbing, a new memo was created describing the passage. Memos enabled me to come back to places of interest within the transcript, and ultimately help with identifying unique phenomena while placing them in context. Memos help largely

with theory building and can refer to codes, comments, text passages or all of the above.

5.3.2 Coding: Interactions

By nature, the paired analysis setup is highly collaborative, and the first pass of coding reflected this. During the first pass, codes described interactions that occurred during the session. Here, “interaction” refers to both SME/TE interactions as well as SME/tool and TE/tool interactions. Most of the first pass of codes followed the methodology of open coding. However this process was slightly modified to accommodate the goal of examining the three types of interactions. Therefore, three initial coding categories were identified that were not in fact “open codes” at the onset of coding. However, within these three categories many subcategories emerged using the open coding technique.

Initially, codes that emerged included things like “SME explaining the domain”, “utilizing the visualization” and “Discussing a course of action”. Codes like “SME explaining the domain” and “Discussing a course of action” were interactions between the SME and TE, while a code like “utilizing the visualization” typically described an interaction between the TE and the tool. The interaction codes from sessions 3 and 6 were grouped together based on whether they consisted of an interaction between the two humans, the SME and the tool or the TE and the tool. There were a few instances of overlaps. For example, the TE may have “utilized the visualization” on the screen while the SME also “utilized the visualization” while explaining the domain to the TE. In

cases such as these the code was placed in multiple interaction categories:
TE/tool interaction and SME/TE interaction.

An initial observation of interaction codes indicated that there were very few instances of SME/Tool interaction. This is due to the nature of the paired analysis setup, in which the TE directs the tool. Therefore to be categorized as a SME/tool interaction, the SME would have to directly observe something on the screen and then comment on it, almost as if talking to the tool. Even then, classifying this type of action as SME/tool interaction is a stretch because the SME did not directly influence the tool, making the interaction more like a one-way communication. Within the categories of SME/TE interaction, SME/Tool Interaction and TE/Tool Interaction, many codes emerged from the data during the first pass. The number of interaction code occurrences is displayed in Figure 20 below:

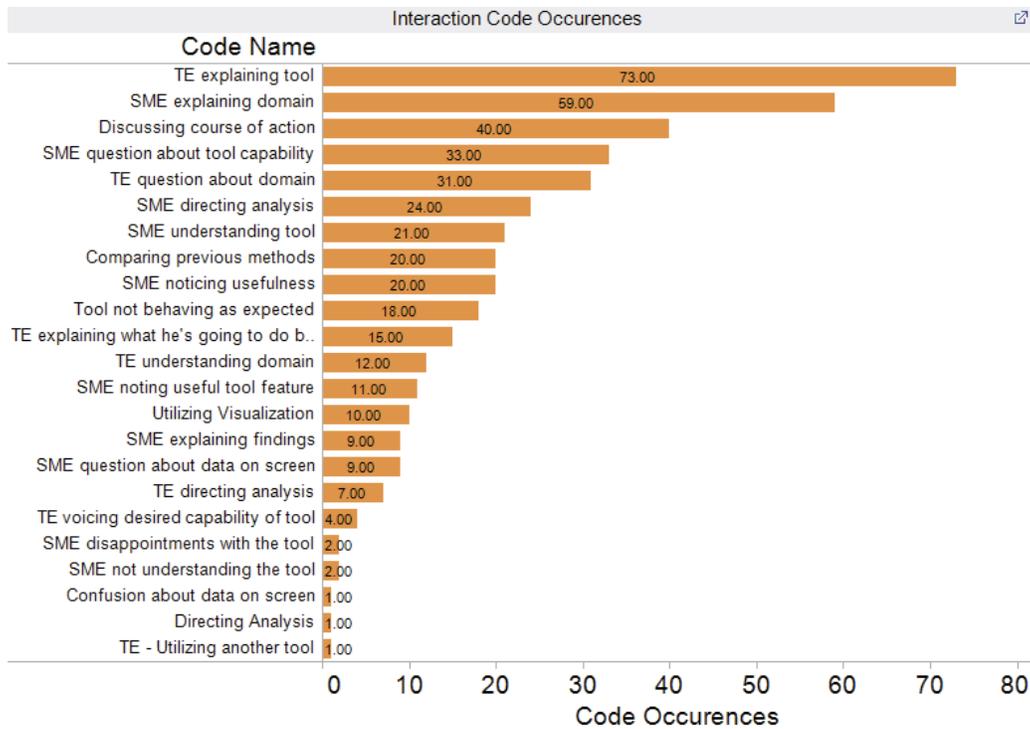


Figure 20: Interaction Code Occurrences

In total, 23 different interaction codes emerged. Total number of codes occurrences gives an idea of what the most frequent types of interactions were during the paired analysis sessions. When ordered by number of instances, the most frequent interactions are “TE explaining tool” (73), “SME explaining domain” (59), “Discussing course of action” (40), “SME question about tool capability” (33) and “TE question about domain” (31). All five of these are human-human interactions, and a few deserve further explanation.

The two most frequent interaction codes were also the two most frequent codes overall. One involves interactions where the tool expert is explaining the tool to the SME. There were 73 occurrences of this across both sessions. This code was applied when the TE described features of the tool to the SME. The following is an excerpt from Session 3 where this code was applied,

****TE: Maybe I should just quickly show you what else is possible, like... ok so there's now 440 detectable ones selected. You could look at those distributed over time. You could look at those distributed by airline. Um, so over time detectable seems to stay pretty steady, there's no really, increase... "*

In this example, the TE wants to make the SME aware of what the tool is capable of. For the purposes of moving analysis forward, this can be seen as beneficial for a few reasons. First, it enables the SME to better understand tool functionality. Ideally, the SME will be able to connect some of his analysis goals to this functionality, helping to smooth the process of learning a new tool. Second, it may provide ideas to the SME about further analyses that are possible beyond the scope of the original goal. Often, the types of analyses people perform are shaped and limited by the tools they have available. When introduced to new capabilities, the types of analysis possible may change and expand in scope or depth.

"TE explaining tool" also covered passages where the TE explained what insight could be gained from the tool:

****TE: K. So you can see, out of the ATBs, 49 of them were RTOs.*

***SME: ah. You can?*

***TE: Yeah, cause the air turnbacks right now are now selected."*

This passage from Session 3 was coded "TE explaining tool" for a different reason than the first example. Here, the TE has manipulated data using

the tool and is explaining to the SME how he has come to the conclusion that 49 of the Air Turn Backs (ATBs) were Rejected Takeoffs (RTOs). The reason has to do with how the tool shows what is selected in the “groups” feature. This type of explanation also involves describing tool features, and was found throughout both sessions.

The second most frequent code that emerged was “SME explaining domain”, with 59 quotations. An example of a passage assigned this code is shown here:

*“**SME: So there's two aspects of the response. Um, one is prior to a bird strike... in other words a prevention strategy.... um, and the other one is after a bird strike. So there's been a known bird strike. And we're gonna look at the latter. So there's been a known bird strike and you know, I'm interested in finding that... what the flight crew response is. Should I give more background?”*

In this passage from Session 6, the SME describes two reasons for examining pilot response in more detail to the TE. This type of interaction serves to educate the TE in the domain of aircraft safety. Through a better understanding of the reasoning behind the SME’s goals, the TE can be more effective and analysis can proceed smoother. These types of interactions were very common in both sessions and also contributed to creating a common ground.

The two main types of interactions occurring in these examples and throughout much of the sessions involve an expert sharing their expertise. This tells us a couple of things about paired analysis:

- Minimizing communication errors was thought to be important to the success of paired analysis sessions. This was shown to be the case during the actual sessions. Establishing “common ground” was not only perceived as important beforehand, but significant time was spent attaining it throughout the sessions.
- Significant learning was occurring on the part of both participants. This provides evidence for using paired analysis as a means of assessing a new tool at an organization.

Both types of interaction occurred throughout analysis, and both involve transferring domain knowledge from one expert to the other. Due to the nature of paired analysis, this should come as no surprise. However, when added together, these codes were attached to 132 passages. Considering there were 433 total codes for any type of interaction, “TE explaining tool” and “SME explaining domain” accounted for just over 30% of the total interactions between the SME and TE.

After a second pass examining each transcript, I was satisfied these interaction codes covered nearly all types of interactions occurring. These codes were then grouped into their respective interaction types. For SME/TE interactions, the group looked like this:

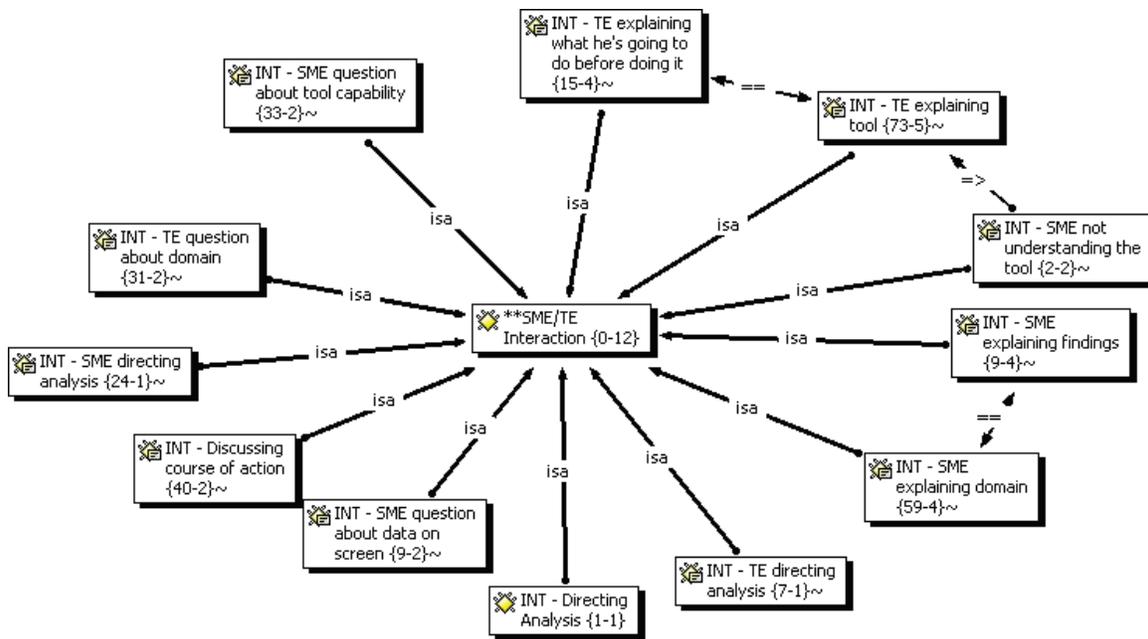


Figure 21: Code Family - SME/TE Interactions

SME/TE Interactions were the most common, and were used for many different types of interactions between the two human participants. In total, 12 different interaction codes fell under the category of SME/TE interaction. These are seen in Figure 21. These codes emphasized the collaborative nature of paired analysis. They ranged from the SME making a statement that he didn't understand the tool to discussions about where the analysis was going. Interactions between the human participants were essential for the analysis to occur, and this is shown through the types of codes present.

There were also a number of interactions with the tool that were essential to paired analysis. The graph below shows the codes that were categorized under TE/Tool interaction (Figure 22).

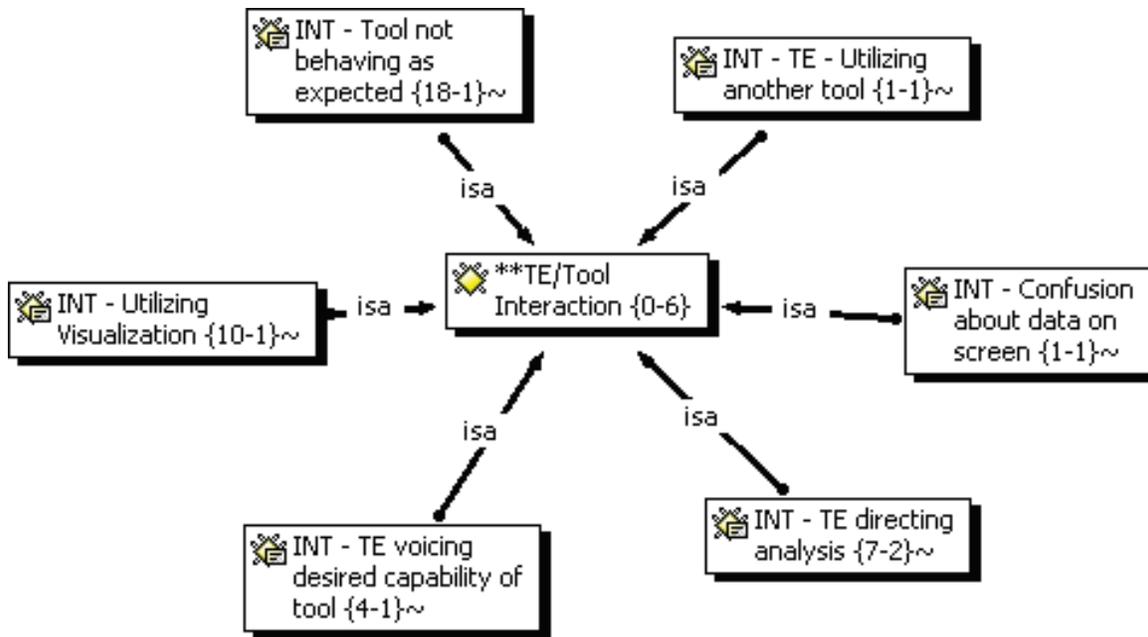


Figure 22: Code Family – TE/Tool Interactions

Due to the nature of paired analysis, any direct interactions between a human and the tool occur on the part of the TE. The SME can direct analysis verbally, but the TE is supposed to be the only one actually manipulating the tool. During the paired analysis sessions at Boeing, the SME did not directly use the tool. In order to capture more codes, TE/Tool interactions also included specific ‘sentiments’ expressed towards the tool on the part of the TE. For example, the codes “Tool not behaving as expected”, “TE voicing desired capability of the tool”, “TE directing analysis” and “Confusion about data on the screen” don’t necessarily imply a physical interaction with the tool. However, for the purpose of including more codes into TE/Tool Interaction, the term “interaction” was interpreted generally. This included adding statements made about how the tool was behaving. The decision to cast a wider net was due to the fact that the TE’s feeling towards the tool’s behavior often influenced future interactions with the

tool. For example, if a particular query was taking a very long time and the TE expressed impatience verbally, the following query was further refined before submission, allowing for quicker processing and decreasing the wait time.

Several types of interactions occur throughout paired analysis and are necessary for it to proceed. It was logical to carry out the first pass of coding using interaction codes due to the desire to examine paired analysis within the context of sensemaking. Using visual analytics tools to approach problems in any domain requires interactions between a human and the tool they are using. However, arguably more important are the interactions of the two people involved (SME and TE). Observing and analyzing these types of interactions helped with achieving a better understanding of analytic sensemaking during paired analysis.

5.3.3 Coding: Sensemaking

The second pass of coding involved the use of codes directly inspired by the sensemaking model proposed by Pirolli and Card (2005). Their model breaks down an analysis into several iterative steps while also separating them into top down and bottom up processes (See Figure 1). The sensemaking model's origins are rooted in intelligence analysis (Pirolli & Card, 2005). Though the task of looking at pilot response to bird strike was not a typical "intelligence analysis", the process of going from raw data to hypotheses and conclusions has many similarities. Coding using the sensemaking model was beneficial for a number of reasons:

- Ability to gauge the amount of time performing top down vs. bottom up analysis

- Ability to gauge the relative time spent in different stages of analysis
- Possibility of illuminating gaps in the model

Each of these points deserves more explanation. For the first point, it is necessary to define what is meant by “top down” and “bottom up”. Here, these terms are referring to a high level view of how an analytic step begins. “Bottom up” refers to processes in which we start with data, fine details and smaller pieces of information and build towards some sort of representation. On the other hand are “top down” processes, referring to instances where we begin with a preconceived notion, concept or idea and work our way down to the finer details. When performing analysis top down, there is typically a defined goal driving the analysis. The analyst goes to the low level data seeking information, confirmation, evidence or all three. Bottom up analysis, on the other hand, often occurs during early stages and when doing exploratory analysis.

To explain the second point, the amount of time spent in each stage of the sensemaking process is valuable knowledge to tool designers. In order to build tools that will accommodate users throughout the process, tool features need to support analytic reasoning steps. It is very unlikely that the time spent in each stage of analysis is equal, thereby making certain types of processes more valuable to focus on when building tools. Understanding the ratios of time spent in each step and the obstacles associated with each should prove to be very useful.

Finally, for the third point, the sensemaking model was originally created based on a single person doing intelligence analysis. In order to incorporate

more diverse types of analysis like collaborative analysis, this model was re-examined in the context of paired analysis.

The codes influenced by the sensemaking model are listed below and are taken directly from steps in the model. A description was created for each of these codes based on the model to help with assigning them during coding. These descriptions were developed as coding went along and experience was gained applying these codes to the transcripts. In the following tables the description is printed after the code name:

Bottom Up (Deductive):

Sensemaking Step Code Name	Description
Search and Filter	Making queries, looking for items of interest, judging relevance of results.
Read and Extract	After finding something of interest, reading it in more detail. If useful, it is noted for later.
Schematize	Re-representing information in some way. This includes creating categories and groups for the data.
Build Case	There is a hypothesis that is either supported or disconfirmed by evidence at hand.
Tell Story	A presentation of findings occurs. This can be the SME presenting findings in the tool by explaining how they fit into context as well.

Top Down (Inductive):

Sensemaking Step Code Name	Description
Re-evaluate	Taking a step back after working to think about where to go next. Can also involve weighing multiple options for how to move forward.
Search for Support	Re-evaluating theories requires looking at lower level schematic organization of facts and evidence.
Search for Evidence	Looking for specific evidence to support a theory or question.
Search for Relations	Evidence that has been found suggests additional patterns or possible relations between other data. Sometimes there is a very specific combination of things the SME will want to look for.
Search for Information	Once the analyst reaches a certain point of the analysis he may dig deeper to explain what is appearing in the data. Often there are inconsistencies in the data.

Here, the sensemaking codes are grouped together based on whether they are “bottom up” or “top down”. It is very common for both types of processes to occur throughout an analysis, and they are often interconnected. The following chart summarizes the number of sensemaking codes occurring in Sessions 3 and 6:

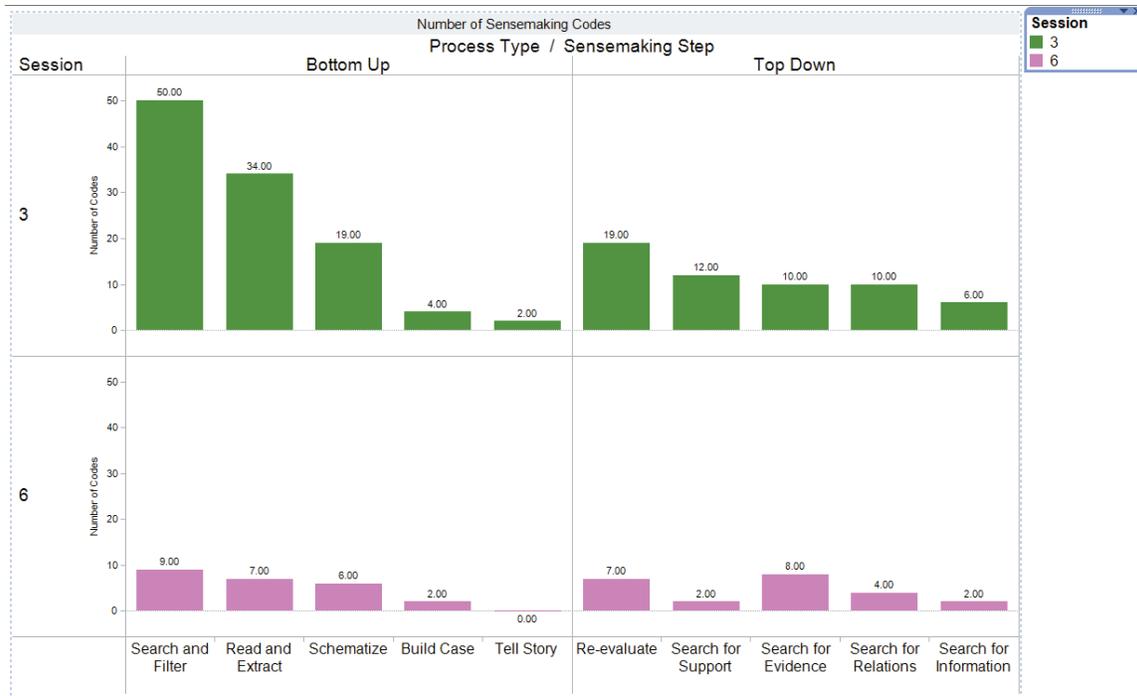


Figure 23: Total Sensemaking Codes by Session

Figure 23 helps give an idea of the proportion of each session that was devoted to a particular stage of sensemaking. Based on the chart it is apparent that Session 3 and 6 differed a fair amount. First, the numbers of identified instances of bottom up and top down processes were proportionally different between the two sessions. In Session 3, there were 109 instances of bottom up processes and 57 instances of top down. This is an approximate ratio of 2:1. In Session 6, there were 26 instances of bottom up processes and 23 instances of top down processes. This is closer to a ratio of 1:1. This signifies a change in the focus of analysis from Session 3 to Session 6.

The change can be explained by the circumstances surrounding the two sessions. Session 3 was the first time the tool IN-SPIRE had been used to examine pilot response to birdstrikes. Therefore, more time was spent in the

exploratory phase of analysis. Session 3 was also a time for the SME and TE to test out the best uses of the tool and learn the best techniques for searching. Session 6 occurred 22 days later and used the same tool on a different dataset for the same task. The learning and experience with the tool on the part of the SME and TE is exemplified by the lower proportion of time spent in the exploratory phase. Faster techniques for searching and wading through raw data had been developed and were applied right away, leading to less time performing bottom up processes.

A further topic of inquiry revolved around the iterative nature of sensemaking and the distribution of analytic steps over time. Though the proportion of bottom up processes was less in Session 6 than in Session 3, the chronological order of when processes occurred was unknown. As mentioned, the task for Sessions 3 and 6 was to examine pilot response to bird strikes. This was a well-defined task that involved a few main steps:

- Identify whether or not a bird strike was detected by the flight crew
- Identify how the flight crew detected the strike
- Identify what actions were taken and the reasons for doing so
- Assess whether the response was appropriate

These steps allowed for the analysis to have defined starting and ending points. Some of the time during each session was spent trying to automate parts of this process (i.e. grouping together reactions). However, in general the analysis followed a path from raw data and questions to hypotheses and conclusions. This led to an inquiry into when certain steps of the process were occurring.

Using Atlas.ti's output functions, the individual sensemaking codes could be examined for the moment during analysis they took place. To aid with explaining this, each session was broken down into five equal parts. The division into five "stages" was based on the number of lines of text rather than length of the sessions. For example, because the transcript for Session 3 has 2447 lines of text, it was divided into five sections of approximately 490 lines ($2447/5 = 489.4$). Session 6 was divided into five sections of approximately 313 lines ($1564/5 = 312.8$). The decision to split the transcripts based on lines of text rather than minutes of audio was due to the fact that for the purposes of this analysis, textual data was the most important. In some ways, the total text transcribed is a better representation of how much and what was accomplished during each session. Because the focus of the analyzing transcripts was on interactions, long periods of silence (and thereby no interaction) are eliminated from affecting the division of stages. The decision to use "five" stages was based on a few reasons. Five is enough to describe how the process changed throughout each session while also being a manageable number to analyze. This number was also deemed appropriate based on the total number of codes, where too many stages would result in too few codes per stage.

Once the sessions had been split into 5 sections, a breakdown of sensemaking codes by stage could occur. Those results are presented here:

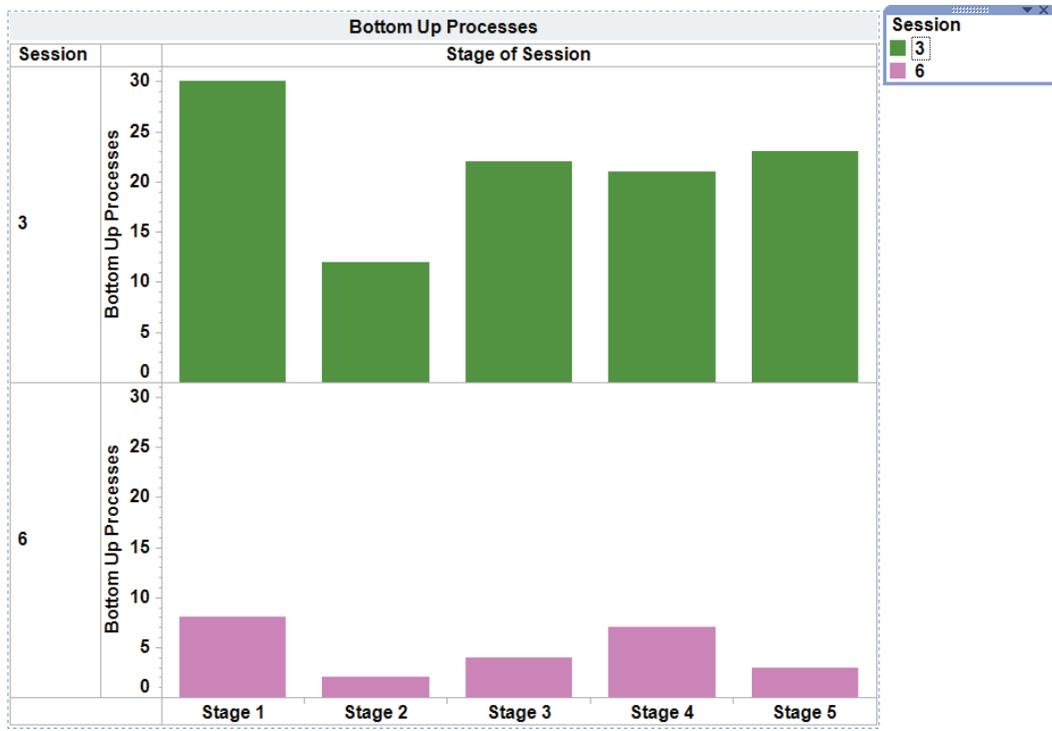


Figure 24: Bottom Up Process Distribution over Sessions Stages

Figure 24 displays the amount of bottom up process codes occurring during the five stages of each session. The green bars represent Session 3 and the pink bars represent Session 6. The x-axis shows the stage of the paired analysis session and the y-axis represents the number of codes present in that stage. For example, there were twelve bottom up processes coded in stage 2 of Session 3, and two for Session 6 during stage 2.

From this visualization a few patterns that immediately jump out. The beginning of each stage was when the highest number of bottom up processes occurred. This makes sense based on what we know about bottom up processes. At this point in each session, the TE and SME were still getting familiar with the tool and dataset and were likely in an exploratory stage of analysis. Then, during the second stage of both sessions there was a steep

decline in the number of bottom up processes. This is followed by a general increase in the last three stages of analysis with a few ups and downs, but never reaching the amount of bottom up processes present in stage 1. This may be due to a couple of reasons. After stage 1's initial exploring and querying, the SME and TE generally have an idea of what they are looking for. This results in the sharp decline in bottom up processes in stage 2.

For the final three stages, the type of bottom up process occurring may explain the increase. Looking at individual sensemaking stages, the amount of codes for "Search and Filter" and "Read and Extract" are highest in stage 1. For later stages (3-5) the codes of "Schematize", "Build Case" and "Tell Story" make up a larger percentage of 'bottom up' codes. After initially wading through raw data in the exploratory stage 1, later stages involve bringing information together to form a coherent story. The prevalence of these types of bottom up codes leads to the increase after the sharp decrease in stage 2. It should be clarified that there are still several "Search and Filter" and "Read and Extract" codes in later stages, however due to the increase in the other bottom up codes the proportion is much lower.

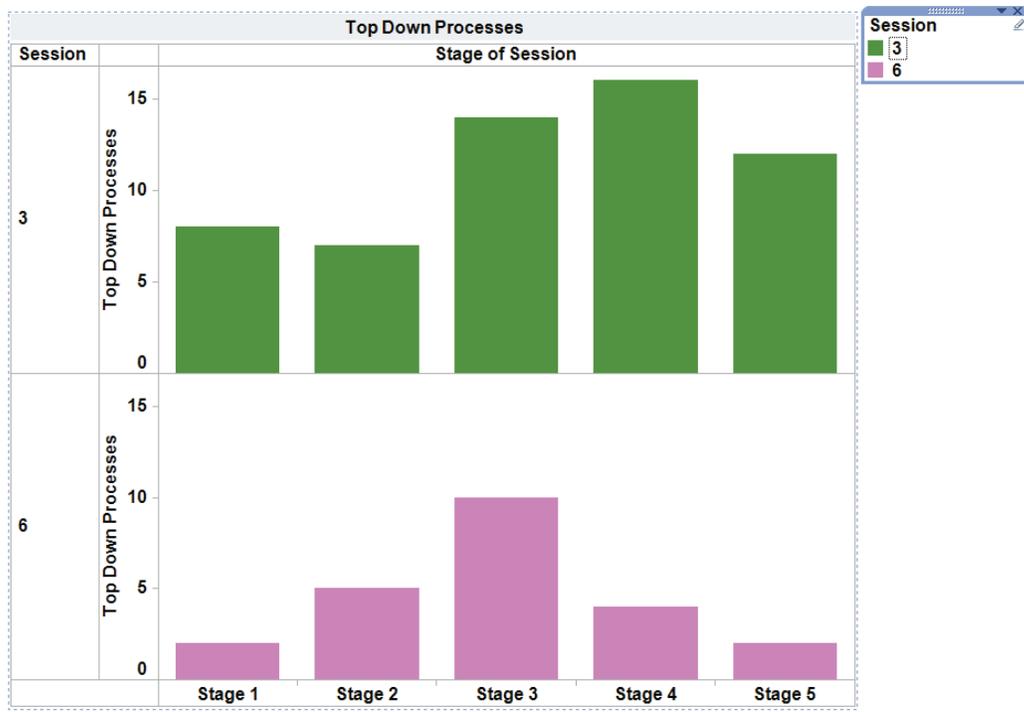


Figure 25: Top Down Process Distribution over Session Stages

Examining the distribution of top down codes in Figure 25 also yields a pattern. For top down codes, there is almost a bell curve. Early stages of analysis in each session have fewer top down processes, followed by an increase and peak during middle stages, and finally a decrease during the final stages of each session. The lower number of top down processes in the first stage of analysis makes sense. Again, this was generally an exploratory stage where the TE and SME were trying to understand the data and tool. Many more bottom up processes occurred during this stage. However, as the analysis goes on the SME and TE begin to ask more direct questions of the data, resulting in more top down processes in middle stages. The decline of top down processes in the final stages may be explained by the fact that fewer new topics are being explored, and instead the existing questions are being examined in greater detail.

Looking at bottom up and top down processes together indicates one very important thing about sensemaking and analytic reasoning: Its iterative nature. Both types of processes occur throughout all 5 stages, and both increase and decrease periodically throughout. This is an important thing to consider when building tools and features to support reasoning. Investigating something visually involves looking at existing information to synthesize new knowledge (Gotz, Zhou, & Aggarwal, 2006). Analysts synthesize knowledge in combination with information exploration, and build on schemas to organize new knowledge. Ideally tools should support the process of defining new knowledge and modifying existing knowledge in response to new discoveries.

6: DISCUSSION

6.1 Summary and Contributions

This thesis presents an in-depth examination of paired analysis, examining its origins and evolution as well as evaluating its application in two separate domains. Paired analysis is a method used for multiple purposes: providing a better understanding of analytical reasoning while also providing a way to integrate and evaluate visual analytics tools within an organization. Paired analysis was inspired by and combines methods and techniques from several disciplines. These include but are not limited to: paired programming, protocol analysis, cognitive task analysis and participant observation.

This thesis is split into six chapters, beginning with an introduction. In addition to introducing the research area, the introduction contains the research questions. Following the introduction, the literature review covers many aspects of both visual analytics and paired analysis, exploring the contributing factors for each. Visual analytics arose as an answer to human short-term memory limitations and other cognitive biases that can hinder analysis of large and/or complex datasets. Visual analytics seeks to utilize the strengths of the human visual system to aid with analysis. The literature review then goes on to examine existing methods for the evaluation of visual analytics tools, highlighting the benefits and shortcomings of each. The literature review also covers methods for investigating analytical reasoning and describes previous research in this area.

This leads to the introduction of the method known as paired analysis, which is described as a combination of existing techniques that can be used to evaluate a tool as well as aid in capturing data about analytical reasoning.

For this thesis the relationship between visual analytics tools and paired analysis was bidirectional. Visual analytics tools were used to help evaluate the paired analysis method, and paired analysis was used to help evaluate visual analytics tools. Paired analysis requires that a subject matter expert work with a tool expert who is using some tool to aid with analysis. The domains where this method was applied – aircraft maintenance planning and aircraft safety engineering – were ideal candidates for the use of visual analytics tools. Both involve significant amounts of data analysis and large, messy datasets. At the same time, VA tools have not been extensively used within either domain in the past, and therefore this research served as a method to evaluate the tools themselves. Therefore, the role of paired analysis was more than just a research methodology for collecting data about analytical reasoning. It was also used as an analysis technique that facilitated tool learning and evaluation on the part of the SME. The literature review ultimately serves to introduce both visual analytics and paired analysis, and explains their relationship and basis for use.

Aeroinfo maintenance planners and their data provided an ideal setting for initial pilot studies using paired analysis and visual analytics tools. At the time of the pilot studies, the research focus was a bit different, and looked at how uncertainty is managed within aircraft maintenance. However, paired analysis was used and many lessons were learned with regard to the technique. Before

paired analysis even occurred, exploratory interviews were carried out to familiarize the tool expert with the domain of the SMEs. Referred to in chapter 3 as “pilot study 1”, these interviews were crucial in facilitating successful paired analysis sessions. In order to conduct a paired analysis session without constant explanation of the domain and data by the SME interfering with analysis, a degree of familiarity with the domain is necessary on the part of the tool expert. Similarly, in order to conduct a paired analysis session without tool explanations by the TE constantly getting in the way of analysis, a degree of familiarity with the tool is necessary on the part of the subject matter expert. The interviews in pilot study 1 helped provide this degree of familiarity with the domain and tool for both tool expert and subject matter expert. A focus structure for these interviews is provided in Appendix A and will aid with future research involving preliminary interviews.

In the second pilot study at Aeroinfo, paired analysis sessions were carried out with a number of SMEs and were video, audio and screen captured. Analysis of these recordings resulted in a published paper (Arias-Hernandez et al., 2011). Pilot study 2 provided the opportunity to examine the paired analysis method in its current form while also evaluating the use of visual analytics tools in the domain of aircraft maintenance planning. Based on analysis of pilot study 2, a few refinements to the paired analysis method were made. These refinements were noted and applied to the paired analysis sessions with the safety engineers at Boeing.

Chapter 4 provides a detailed description of paired analysis research carried out with Boeing safety engineering SMEs. As with Aeroinfo maintenance planners, research began with exploratory interviews. These interviews served multiple purposes: allowing the TE and SMEs to become familiar with the other's domain and tool, exposure to data and possible tasks for paired analysis, and a general face-to-face introduction of the TE to potential SMEs. By familiarizing themselves with the domain and tool before the actual paired analysis sessions, the TE and SME helped facilitate productive and efficient sessions.

Following exploratory interviews, the TE conducted several paired analysis sessions during an internship working closely with Boeing SMEs. Chapter 4 describes the steps to setting up a paired analysis session in detail. These include choosing a task for analysis, choosing a dataset, choosing a tool, data cleaning and preparation and steps taken to record sessions. Each one of the steps played an important role in the success of a given paired analysis session. For future paired analysis research, these steps and their explanations provide guidelines for handling data and the selection potential tasks. They also detail how data was prepared for analysis. Guidelines for these time-consuming steps are very valuable to researchers and any participants of future paired analysis sessions, particularly the tool expert. Other uses for individual steps examined in Chapter 4 include providing suggestions and guidelines for an organization seeking to use the paired analysis method for evaluating and integrating a new tool. Evaluation, learning and adoption of a new tool are very time and money-consuming tasks for most organizations. By providing a detailed

description of steps necessary for paired analysis to be used in evaluating a tool, this research can be used as a framework for organizations seeking to evaluate and integrate new tools.

Chapter 5 provides a detailed description of the analysis done on recorded data from two selected paired analysis sessions. It begins by explaining why two sessions were chosen for deeper analysis and how this selection process occurred. Data collected from paired analysis sessions is inherently rich with information. Because sessions were video, audio and screen recorded, there are a plethora of things a researcher could examine from the data. Some of these include things like gestures, speech patterns, tool use and discourse between the SME and TE. Conducting detailed research on any one of these things could yield a separate thesis. However, for this thesis, transcripts of audio data were the primary source of data. By explaining the factors involved in selecting two sessions and their audio transcripts, this chapter helps contribute to further paired analysis research. A framework for selecting appropriate sessions for analysis is very valuable, as analysis of video and audio data is very time consuming. Chapter 5 explains how sessions were chosen with the goal of finding tasks that would most likely increase researcher knowledge about analytical reasoning.

Following selection of two sessions, Chapter 5 goes on to describe how the sensemaking model was applied during analysis of data. As mentioned, the sensemaking model served as the primary model for examining analytical reasoning during the paired analysis sessions. The sensemaking model provided

inspiration for codes used to analyze transcripts, and acted as a framework for modelling the analytic task of examining pilot response to bird strikes. Chapter 5 describes factors involved with coding the raw transcript data. Coding was one of the primary techniques used to examine the data collected for its possible contributions to analytic reasoning. Future research involving paired analysis and analytic reasoning can benefit from the description of how coding occurred, and how the sensemaking model was used.

The use of paired analysis and visual analytics at Aeroinfo and Boeing has helped advance research in each area. The method of paired analysis has been dissected and analyzed during use in two separate domains. Its ability to elicit useful data about analytic reasoning and sensemaking has been demonstrated through analysis of the two sessions at Boeing. The method has also demonstrated its ability to help organizations evaluate and learn a new tool. These findings provide evidence to be used for justification of future research using real analysts and paired analysis. Chapters 4 and 5 mention video, audio and screen capture data from seven additional paired analysis sessions that were not analyzed in this thesis. This data has been collected and stored, and will provide an opportunity for others to continue paired analysis research on sessions conducted using analysts in the real world. This thesis has provided the ground work for further paired analysis and visual analytics research.

APPENDICES

Appendix A: Aeroinfo Focus Structure

Introduction (to be read first to establish context):

In order to capture the process of managing and communicating uncertainty, Aeroinfo and Boeing analysts and maintenance planners are being interviewed. Our goal is to identify reliability and maintenance tasks that analysts and planners like yourself may have recently performed. Ideally the task or solution involved some type of visual representation.

As the researcher, I will be asking for descriptions of your maintenance and reliability tasks as well as specific incidents with the goal of later analysis. Some of the questions will involve the concept of uncertainty and what uncertainty means to you.

Things to keep in mind during the interview:

- What does the concept of uncertainty mean to you?
- Where in your data is there uncertainty? For example:
 - Sources...etc
 - Missing/incomplete/incorrect data
 - How does this play a part in your analysis or planning?
- How do the tools you currently use aid in overcoming uncertainty, if at all?
- How do you communicate uncertainty in your solutions or to others?

- What terms do you think of uncertainty? E.g., Confidence? Trust? Reliability/quality of data? Their judgment of the data?...etc

- Come back to this as it arises but never forget about it, always try come back to uncertainty where possible.

Questions:

Describe a typical day

- Describe a typical day
 - What are some of the bigger or more important tasks you perform?
 - What are some of the smaller or less important tasks you perform?
(Pick at least one of these tasks from their answers to go into more detail)
- Are there other unique or unusual tasks?
 - How often?
 - Can you describe a few of them?
- What makes a task typical or unique?

For a Specific Typical Task

- Describe the task that you mentioned?
 - What were the steps you took in completing this task?
- Specific steps in collecting data?
 - Can you describe the processes of searching, querying, gathering, converting the data used?
 - If you had to, how did you “clean” the data?
 - How do you know when you’ve searched enough?
 - Threshold?
 - Experience?
- Please describe any problems arising during the completion of the task.
- Please describe specific difficulties encountered during the task.
 - Data search, ingest, messy data...etc
- Why did these difficulties arise?

Tools Used

- What types of tools did you use during the task?
- How was each of these tools used?
- Describe any ways the tools helped or held you back
- What parts of the analysis require a human?
- Where does technology fit in?
- What parts of the knowledge needed to solve the problem are in the head of the analyst?
- How much knowledge is external or in the tools?

Data and Task Analysis

- Which tasks require you to gather and analyze data?
 - Are there other circumstances when you need to look through datasets?
- How is useful data stored?
 - Is it stored for future tasks or queried each time?
 - How is knowledge passed on to new analysts?
 - How do you find an answer?
 - What does an answer look like?
 - How do you know when you're finished with an analysis?
 - How do you keep track of possible findings?
 - How do you track evidence?
 - How do you form hypotheses?
 - What causes you to form a hypothesis?
 - How do you evaluate it?
 - When do you begin forming and testing hypotheses?

Other Analysts

- What role do other analysts serve in solving your problems?
 - How do you collaborate with other analysts?
 - Do you divide tasks into parts?
- How do you communicate with other analysts during a task?
 - How does this help/hinder the task?

- Please comment on the work practices of other analysts...
 - Does each analyst have the same skillset?
 - Describe some of the differences you've seen, e.g.,
 - Differences in data acquisition, processing, schematizing, evaluating hypotheses
 - How are tasks typically distributed across analysts?

Reporting Findings

- What happens when you've completed your task/analysis?
 - How do you report your plans?
 - Do you use a specific format?
 - Is the report task dependent?
 - How do you justify your plans?
 - Do visualizations help in the reporting of your findings?

Specific questions based on previous interview

...

Additional Info

- How long have you been working at Aeroinfo?
- What is your job title?
- What are your main responsibilities?
- What kind of training do you have?
 - Background?
 - Academic degree?
 - Formal training courses?
 - Informal training?
 - Specific classes you've taken for training?
 - At Aeroinfo/Not at Aeroinfo?
- How do you apply your training?

*If you were hiring a new analyst, what would you look for? What skill sets would be most useful?

Appendix B: Boeing Safety Engineer Focus Structure

The following questions-of-interest guided the VA team during the sessions with the Boeing Safety Engineering team. Note that questions were seldom asked in the specific order shown, but an attempt was made to ensure that as many questions as possible were addressed during the sessions.

Introduction

Our goal is to understand the role of the Safety Engineer at Boeing and to identify safety-related tasks that Safety Engineers, like you, have to perform. We are particularly interested in tasks that involve or could involve some type of visual representation in the process or the solution.

Basic Information

- How long have you been working at Boeing?
- What is your current job title?
 - How long have you been doing this job?
- What are your main responsibilities?
- What kind of training do you have?
 - Academic?
 - Background?
 - In Safety Engineering?
 - Formal or informal training?
 - Relevant classes have you taken at Boeing/not at Boeing?

Interview Questions

1. Please describe a typical day

- Please describe a fairly typical or common day...
- In general,
 - What are some of the most frequent tasks you perform?
 - What are some of the more important tasks you perform?
 - What are some of the larger or more complex tasks you perform?
 - What are some of the more mundane tasks you perform?
 - What tasks require expert analysts or groups of analysts?
- Are there other less frequent, unique or unusual tasks that are important?
 - Can you describe a few of them?
 - What makes them unique?
- How is analytical work allocated among individuals and how are machines used, i.e., the division of labour?
- What is the structure of the particular organization that helps to standardize this work?

- How do analysts respond to contingencies and competing demands?
 - What time constraints are they under?

Pick one of these tasks to go into more detail; select a task that is more frequent, more important, larger or complex, or unique.

2. About the Specific Task

- Describe the task that you mentioned?
 - Why did you work on this task?
 - How was the task assigned?
 - Who were the task results for?
- What were the steps you took to complete this task?
 - Where do you get the data needed to work on the task?
 - Can you describe the process of searching for, gathering together, converting if necessary, and querying the data needed?
 - If necessary, how do you “clean” the data?
 - How do you know when you’ve searched enough?
 - A threshold is reached? If so, who sets the threshold?
 - Based on experience with similar tasks?
- Please describe any problems or difficulties that commonly arise during the completion of the task. Why do these difficulties arise?
 - Data search, ingest, messy data...etc

3. Describe Tools Used

- What are the resources (e.g., computer-based tools, standards and regulation documents, other reference documents, other experts in their group, in other Boeing divisions or in other companies (e.g., airline or government, e.g., FAA) used to execute this work?
- What types of tools would you typically use during the task?
 - If more than one, how was each tool used?
- Describe how the tools helped or possibly held you back
- What parts of the analysis definitely require a human? Where could technology fit in?
- What knowledge needed to solve the problem is in the head of the analyst? How much does the analyst’s experience matter?
- How much knowledge is currently in the functionality of the tools?
- How do humans interact with artefacts during analytical work?

4. Describe Data and Task Analysis

- How is useful information stored?
 - Is it saved for future tasks or newly queried each time?
 - How is task knowledge passed on to new analysts?
- How do you know when you have found an answer?

- Is there often a single answer to a problem?
- What does an answer look like?
- How do you know when you're finished with an analysis?
- How do you keep track of interesting findings?
- When do you begin forming and testing hypotheses?
 - How do you go about forming a hypothesis?
 - How do you evaluate it?
- How do you track evidence?
- How can you be certain that you have the solution to the task problem?
 - How do you communicate certainty or uncertainty?
- Is there uncertainty in the data you use? For example,
 - In the data input sources, e.g., the humans?
 - Missing/incomplete/incorrect data?
 - How does data uncertainty play a part in your analysis?
- What terms would you use when you think of uncertainty?
 - E.g., confidence, trust, reliability, quality of the data?
 - Any others?

5. Relationships with Other Analysts

- What are the human interactions, e.g., negotiations, requests, collaborations, etc., that the analyst engages in to perform her/his work?
- How are fragmented activities among different individuals or machines coordinated?
- What are the social procedures for the validation of the analysis?
- What role do other analysts play in solving your problems?
- When and how do you collaborate with other analysts?
 - How are tasks typically distributed across analysts?
 - Do you divide tasks into parts and work on them separately?
- How do you communicate with other analysts during a task?
 - Does the need to communicate help or hinder task completion?
 - Do all analysts have the same skillset?
 - E.g., are there differences in their knowledge or skills related to data acquisition, configuring and processing, or hypotheses generating and evaluating?

6. Reporting Findings

- What happens when you've completed a task/analysis?
 - How do you report the results?
 - Do you use a specific format?
 - How do you report the process?
 - How do you justify your results?
 - Do visualizations, such as graphs, help in the reporting of your findings?
- How is this work evaluated or judged and by whom?
- What constitutes success or failure?

- What is the structure of authority for communicating analysis jobs or results?
- What resources are compiled to make a case for the analysis results?
- What are the audiences and what are the purposes of communicating analyses?
- What are the difficulties that analysts experience when communicating analyses?
- What are the difficulties that audiences (e.g., peers, supervisors, the Boeing Safety Board, FAA, etc.) experience when evaluating analysis results?
- How is a common ground constructed?
- How is trust in information, tools, methods and results developed?

7. If you were hiring a new analyst, what would you look for? What knowledge and skill sets would be most useful?

Appendix C: Boeing Internship Task Sheet

R1: 12/16/09

R2: 01/27/10

BIRD STRIKE HAZARD

Literature Review – read bird strike articles

Assist in development of AERO magazine bird strike article, wildlife hazard guidance for flight crews

Assist in development of Bird Strike Resources Portal

Bird Strike Data Analysis using FAA NWSDB (and/or Boeing ASIS data warehouse): airplane, operational, & environmental requirements.

- Compare Airbus A320 family (A318, 319, 320, 321) and Boeing 737NG family (737-600, 700, 800, 900) 3. I liked the spatial / temporal visualizations; particularly the species and airports (I would like to see species ranking by major airports, transport category aircraft).
- Struck vs. damaged comparisons are helpful; would also like to see single vs. multiple engines; also radome/pressure bulkhead, windshield, nose gear steering, etc. (attention on continued safe flight and landing, effect on flight, flight crew operations – rationale for AERO magazine flight crew guidance); external sensors; flight deck systems & integrated modular avionics; landing gear and associated systems
- Narrow body / single aisle vs. wide body / twin aisle
- Passenger / Cargo vs. time of day (dusk, dawn, day, night) and local clock time.
- Speed and altitude vs. damage / effect on flight & species (Kinetic energy)
- Struck and damaged during missed approach (go around)
- Single and multiple strikes; damage
- Reporting issues; statistical significance and inference from inconsistent & incomplete records. Comparison of data from Boeing ASIS DW and FAA NWSDB.
- Selected locations (SEA, MWH, etc.) / operators: all Boeing, all Airbus, mixed fleet (ASA, JBU, UAL, etc.)
- External data: airport lat/lon; model hours/departures; operator model hours/departures; bird species weights

Bird Strike References

- Bird Strike Damage & Windshield Bird Strike. Final Report by for EASA, http://www.easa.eu.int/ws_prod/r/doc/research/Final%20report%20Bird%20Strike%20Study.pdf
- Dolbeer 2006 JWM Height Distribution of Birds Recorded by Collisions with Civil Aircraft

- WILDLIFE STRIKES TO CIVIL AIRCRAFT IN THE UNITED STATES 1990–2008, <http://wildlife-mitigation.tc.faa.gov/wildlife/downloads/BASH90-08.pdf>

OTHER AVIATION SAFETY APPLICATIONS

- Human performance / human factors: automation, spatial disorientation (given likely SD accidents attempt to extract themes or patterns from published reports; categorize spreadsheets); possible FOQA data (NASA)
- Turbulence Injuries
- Pitot icing / unreliable airspeed
- Stick shaker / stall warning / stall
- Runway excursion / incursion
- FAST topic
- Current or NextGen Air Traffic Management topic: wake vortex, altitude busts, loss of separation

DATA VISUALIZATION

In-spire and Tableau familiarization & training

Paired analysis

Outline of thesis

Proposal

Methods

Data collection

Data analysis

DELIVERABLES

Weekly verbal presentation to Aviation Safety R&D Meeting

Weekly verbal presentation to Safety data Team

Weekly brief summary of work completed, and plan for next week

Briefing / Brown Bag / Webex: Develop presentation, meeting notice, schedule meetings

- Renton, Everett, 787, Long Beach Safety Groups Air Safety Investigation
- Human Factors community
- BR&T Data Concepts
- AIR Board
- CAST
- Safety Special Attention Meeting (SAM)
- Aviation Safety Council
- Bird Strike virtual team

Possible briefings to other Boeing organizations

- Propulsion Safety
- Flight technical

- Systems engineering (requirements)
- FOD prevention
- Structures U
- Systems Concepts Center
- Safety/Airworthiness (SAW) Chiefs
- Reliability
- Flight Test Data
- Aero perf / stability & Control
- Air Traffic Management (ATM)

Write up lessons learned, training packages, recommendations
Add Data Visualization to aviation safety portal

Prepare abstract / outline / paper / presentation, Bird Strike hazard, for a technical conference:

- 2010 Bird Strike Committee USA 21–24 June Salt Lake City
Abstract due: March 1, 2010
http://www.birdstrike.org/meetings/2010_BSCUSA.htm

- 5th International Birdstrike Committee Meeting
21st to 24th September 2010, Cairns, Australia
Abstract due: TBD

http://www.int-birdstrike.org/Meeting_Information.htm

- 63rd annual International Air Safety Seminar (IASS) 2010
November 1–4, 2010 Milan Italy
Abstract due: April 2, 2010
<http://www.flightsafety.org/>
Topic: Addressing bird strike hazards
SAE (S-18 meeting)
AIAA ATIO
ISSC
ISASI

Commercial Aviation Safety Team CAST / JIMDAT Wildlife hazard data analysis briefing <http://www.cast-safety.org/>

FAA/Eurocontrol ATM Seminar

University of Washington Aero & Astro seminar March 12, 2010

Data Union FAA/ASIAS

BOEING ACTIVITIES, VISITS

Cab session

Lab tour & bench session

Factory tour(s): Renton, Everett, Auburn

Bellevue

Flight test tour, possible ground / flight test

Future work

Visit to SEATAC Airport wildlife mitigation

<http://www.portseattle.org/community/environment/wildlife.shtml>

FAA Seattle Aircraft Certification Office, Transport Airplane Directorate, Aviation Safety Data, Alaska Airlines

Appendix D: Collected Data Description

Date (If There)	Session Num	Real Session Num	File Name	Extension	Length (min)	TE	SME	Data Type	VA Tool	Other tool	Description	Notes	Good (gut feeling) (y/n)?
	1,2	1	6	mts	29:50	TE 1	SME 1	Video Recording	IN-SPIRE	Excel	Explanation of where 1087 documents came from. Loading of the actual documents into IN-SPIRE. SME1 helps to determine which fields are valuable on import. Conversion of data to xml.	Interactions where TE asks SME about dataset. TE explains what he is doing. Accidentally stops the screen recording but keeps the video recording going. Lots of spreadsheet editing	
25-Feb-10	3	2	8	mts	37:23	TE 1	SME 1	Video Recording	IN-SPIRE		IN-SPIRE and Bird Strike Response. Have imported all documents in IN-SPIRE. SME1 explains background and where classifications come from and rationale for human response (Rasmussen response). Discussion about how to categorize events, don't want to have to go through events multiple times because it becomes very repetitive. Late in session walk through all sensory terms for the query.	Interactions about 5 categories and why they would be used. Use of document viewer and highlighting. Use of galaxy view and outliers. SME1 asks questions about the tool. More exploratory, trying to figure out what to do. Testing out features (find similar records). Video good for teaching the tool expert about the task. Cuts off at end of video.	y, for early session pre-learning

25-Feb-10	3	2	9	mts	37:12	TE 1	SME 1	Video Recording	IN-SPIRE	Continuation of Session 3. Use of query tool. Creation of custom groups (EICAS models. Some demo of tool (correlation) to show off features. Lots of going through documents keeping or throwing out.	Aha moment when TE realizes EICAS models are all detectable. Good brainstorm back and forth to figure out next steps (combination of both SME and TE leading what to do next). There is some tool checking (using tool to find RTOs and then comparing the number found to the SMEs spreadsheet of the same query in the past. TE asks what phase an ATB is, later sessions he knows...learning.	y, for early session pre-learning
25-Feb-10	3	2	10	mts	28:31	TE 1	SME 1	Video Recording	IN-SPIRE	Continuation of session 3. Trying to determine what's useful (looking at detectable strikes and then the response). Looking for specific things (detectable and continued flight). Discover that the groups tool is a good way to clean up groupings	This is later in the same session (3) and contains more specific searches using the group tool. Discussion of when and how to code. Discussion to finish task and then pick up the next day. Decided to finish making groups for possible responses (ATB and DIV). Decide to do the IFSD the next day. Makes comment that he was faster at doing the counts his way.	y, for early session pre-learning
25-Feb-10	3 (Post-session Discussion)		11	mts	3:50	TE 1	SME 1	Video Recording		Discussion of session 3. Talk about what the next part would be (identifying two more crew actions).	Interaction between SME and TE about what should come next.	
26-Feb-10	4 (Pre-session discussion)		12	mts	2:20	TE 1	SME 1	Video Recording		Discussion of what we did yesterday. Talk about what we want to do in the session today. Want to look at responses at V1		

26-Feb-10	5	3	14	mts	37:31	TE 1	SME 1	Video Recording	IN-SPIRE	Excel	<p>Begins with talk about finding the detectable strikes. Use SME's spreadsheet to help with creating search terms. Walking through records categorizing them one by one. Lots of individual report analyzing, spend significant amount of time on some reports. Fair bit of checking with SME1's spreadsheet where he's already done calculations and analyses.</p>	<p>Interruption a little ways in when a hard drive begins making noise. Began labeling responses as incorrect and correct (used group tool). This comes from SME's expertise. Occasionally there are unique records that puzzle the SME on how to score. SME lets TE know approx how many records he expects, this influences the way the TE analyzes it and groups it.</p>
26-Feb-10	5	3	15	mts	36:52	TE 1	SME 1	Video Recording	IN-SPIRE	Excel	<p>Continuation of session 5. More comparison of number of results in IN-SPIRE and Excel. Determined we've seen a lot of miscoding of flight phase. Finding some records where the pilot reaction was ambiguous.</p>	<p>SME leaves to go to washroom, TE keeps working. SME allowing TE to do a lot, keeping track of the categories is becoming a bit confusing.</p>
26-Feb-10	6	4	16	mts	37:29	TE 1	SME 1	Video Recording	IN-SPIRE	Excel	<p>Picking up with responses to bird strikes. SME sees potential for affect feature of tool (cockpit voice recorder). Noticed "amp" as a word that popped out, but this is a product of the tool (amperstand). SME and TE work through this to discover what it is.</p>	<p>TE suggests at one point that another VA tool may be useful for a particular type of comparison the SME wants to do. SME points at the tool to ask for clarification. Both TE and SME are trying to figure out what viewpoints do (learning about tool). SME asks specifically about whether there is a methodology printed for how to use tool (from PNNL). Are there strategies and possible case studies. SME states that he feels we are unstructured, undisciplined. Exploratory stage.</p>

26-Feb-10	6	4	17	mts	10:53	TE 1	SME 1	Video Recording	IN-SPIRE	Excel	continuation of session 4. summing up what we did. Went through list to look for detections. Looking to understand intermediate steps. Progressed from looking at detection and execution, SME says he is interested in exploring and finding evidence where something has happened and then what the pilot does.	SME states he's still not quite sure what the tool's capable of. A couple minutes in it switches to discussion of what to do next. SME makes point that visualization doesn't really have seemed to help us too much.
	official session 4	5	18	mts	37:22	TE 1	SME 1	Video Recording	Tableau		Begins with description of what we want to do. Airbus vs. boeing for number of strikes in tableau, wide body vs. narrow body, look at differences in rate. Have a set rate from some report, SME wants to investigate that rate. Looking at rate differences between Boeing and Airbus	had to recreate a little bit of what we did already. Problem with the rates, checking the rates against manual calculations. Having trouble making rates be what they should. *SME asks "how would you compare...? Using visualization to confirm SME's hypotheses. Working through problems with linking the tables through Tableau, it's losing records when combing rates tables. Spend a significant amount of time assessing this problem (trying to figure out tool).

	official session 4 (continued)	5 (continued)	19	mts	20:42	TE 1	SME 1	Video Recording	Tableau	Access	continuation of Tableau Session. Begins with still continuing to figure out what went wrong combining the tables. Checking rates against SMEs. Most of this second part is trying to figure out what went wrong with combining the tables.	Evidence of learning the data (figuring out exactly how rates are calculated and where they come from). Discover that TE made wrong grouping for 720 family hours.	no, too much trying to figure out data and sideline discussion (don't stay focused on analysis) yes
19-Mar-10	6	6	29	mts	37:23	TE 1	SME 1	Video Recording	IN-SPIRE	Tableau	Pilot response to bird strikes using the FAA data. Loaded FAA data with goal of comparing what we did with ASIS. 39,000 FAA records of Boeing and Airbus. Want to look at pilot response. However, the groups weren't automatically created by IN-SPIRE. Utilizes Tableau to find the flight phases because it is quicker and easier.	*See "Preliminary findings" in 2010-04-26-Wade-IAT885-Project-Paired-Analysis.docx writeup for detailed description of this session.	
19-Mar-10	6	6	30	mts	33:42	TE 1	SME 1	Video Recording	IN-SPIRE		continuation of session 6. Looking at takeoff run. Looking at groupings and intersections.	SME asks specifically how to drill down to specific records using tool. TE makes several comments about how long the computation is taking. Amperstand popping out in the themes. Realization that group made in access will be very useful for searching documents. SME making comments about what he thinks is useful in the tool (intersections and grouping ability).	

	7	7	31	mts	5:08	TE 1	SME 2	Video Recording		Excel	Dataset is different business units saying they need things from other areas, usually funding for research. Goal is to determine what the needs are that reflect a need for visualization and interactive techniques.	phone call interrupts this session.	
	7	7	32	mts	37:47	TE 1	SME 2	Video Recording	IN- SPIRE, Tableau	Excel	continuation of session 7. Looking at different processes of the company where business units need funding. How to fund in a more efficient manner.	Able to find duplicates in IN-SPIRE by seeing dots stacked on each other. Combination of IN-SPIRE and Tableau with data. Query tool in IN-SPIRE.	
	7	7	33	mts	25:02	TE 1	SME 2	Video Recording	IN-SPIRE	Excel	continuation of session 7. Using query by example tool to find similar documents. Want to have as many needs addressed as possible with as few projects as possible. This means understanding the similarity of needs is essential.	Interruption by meeting someone SME 2 works with, but then continues. Noticing different ways to spell things while reading through documents and adding new spellings to the original query.	
30-Mar-10	8	8	40	mts	10:04	TE 1	SME 1	Video Recording			This is the session that was observed by SFU students.	Briefing the observers on the pilot response task.	
30-Mar-10	8	8	41	mts	37:42	TE 1	SME 1	Video Recording	IN-SPIRE		Beginning of session 8 with SME after briefing. Briefing observers and going over search terms again.	SME1 goes to images of flight decks to explain where the pdfs, eicas, and readouts are for warnings. Becomes a team effort.	no, not paired analysis, group analysis.
30-Mar-10	8	8	42	mts	25:24	TE 1	SME 1	Video Recording	IN-SPIRE		Continuation of session 8. More group analysis and explaining analysis.		no, not paired analysis, group analysis.

31-Mar-10	9	9	43	mts	27:07	TE 1	SME 1	Video Recording	IN-SPIRE	Engine fuel leak as requested by someone in everett. Specific task using SME1's ASIS documents and IN-SPIRE.	Not exploratory, looking for a very specific type of incident (different from others). Lots of refining queries to capture specific problem (fuel leak). Interrupted by phone call and pause session	yes, but it is a different kind of task completely than before
31-Mar-10	9	9	44	mts	37:59	TE 1	SME 1	Video Recording	IN-SPIRE	Continuation of session 9 after phone call. Continuing to look at overlaps and queries to refine it best to capture the type of event we want.	Creating categories that describe exact events as well as those that are "close" or that can't quite be determined based on the narratives. Spend some time clarifying categories.	yes, but it is a different kind of task completely than before
31-Mar-10	9	9	45	mts	13:44	TE 1	SME 1	Video Recording	IN-SPIRE	Continuation of session 9. Lots of reading individual articles and determining their categories. At end we have discussion of where to go, then continue categorizing records for eventual report to group that posed initial query.	point where SME shows good knowledge of capabilities of tool.	
	1	1	cap re-1	camrec		TE 1	SME 1	Screen Recording				
	1	1	cap re-2	camrec		TE 1	SME 1	Screen Recording				
	3	2	cap re-3	camrec		TE 1	SME 1	Screen Recording				
	5	2	cap re-4	camrec		TE 1	SME 1	Screen Recording				
	3	3	cap re-5	camrec		TE 1	SME 1	Screen Recording				
	4	4	cap re-6	camrec		TE 1	SME 1	Screen Recording	Tableau			

19-Mar-10	6	6	capture-7	camrec			Screen Recording	
	7	7	capture-8	camrec	TE 1	SME 2	Screen Recording	
	7	7	capture-9	camrec	TE 1	SME 2	Screen Recording	
30-Mar-10	8	8	capture-10	camrec	TE 1	SME 1	Screen Recording	UBC student observation
31-Mar-10	9	9	capture-11	camrec	TE 1	SME 1	Screen Recording	engine fuel leak
31-Mar-10	9	9	capture-12	camrec	TE 1	SME 1	Screen Recording	
			Boeing Log - Week 1	doc			Journal	Description of week's activities, research, progress, notes and observations
			Boeing Log - Week 2	doc			Journal	Description of week's activities, research, progress, notes and observations
			Boeing Log - Week 3	doc			Journal	Description of week's activities, research, progress, notes and observations
			Boeing Log - Week 4	doc			Journal	Description of week's activities, research, progress, notes and observations

Boeing Log - Week 5	doc	Journal	Description of week's activities, research, progress, notes and observations
Boeing Log - Week 6	doc	Journal	Description of week's activities, research, progress, notes and observations
Boeing Log - Week 7	doc	Journal	Description of week's activities, research, progress, notes and observations
Boeing Log - Week 8	doc	Journal	Description of week's activities, research, progress, notes and observations
Boeing Log - Week 9	doc	Journal	Description of week's activities, research, progress, notes and observations
Boeing Log - Week 10	doc	Journal	Description of week's activities, research, progress, notes and observations

Boeing Log - Week 11	doc	Journal	Description of week's activities, research, progress, notes and observations
Boeing Log - Week 12	doc	Journal	Description of week's activities, research, progress, notes and observations
Tasks for TE1 Wade	doc	Description of analysis goals	Compiled by SME1 and later refined by TE1 and SME1. This document details many of the safety investigations that we were attempting
Areas Influenced	rtf	analysis results	Written up by SME1. Specific areas and planes that were affected by VA results.

REFERENCES

- Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology, 69*(2), 334-345. doi:10.1037/0021-9010.69.2.334
- Ahlberg, C. (1996). Spotfire: an information exploration environment. *ACM SIGMOD Record, 25*(4), 25–29.
- Amar, R., & Stasko, J. (2005). A knowledge task-based framework for design and evaluation of information visualizations. *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on* (pp. 143–150).
- Arias-Hernandez, R., Kaastra, L., Green, T., & Fisher, B. (2011). Pair Analytics: Capturing Reasoning Processes in Collaborative Visual Analytics. *HICSS-44 Conference Proceedings*. Presented at the HICSS-44, Kauai, HI.
- Atkinson, P., & Hammersley, M. (1994). Ethnography and participant observation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 248-261).
- Beck, K. (2000). *Extreme programming explained: Embrace change*. (Vol. 190). NJ: Addison-Wesley.
- Beck, K., & Andres, C. (2004). *Extreme Programming Explained: Embrace Change (2nd Edition)*. Addison-Wesley Professional.
- Becker, H. S., & Geer, B. (1957). Participant observation and interviewing: A comparison. *Human Organization, 16*(3), 28–32.
- Bowden, E. M., & Jung-Beeman, M. (2003). Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychonomic Bulletin & Review, 10*(3), 730.
- Bowden, E. M., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences, 9*(7), 322–328.
- Calder, B. J., Phillips, L. W., & Tybout, A. M. (1981). Designing research for application. *Journal of Consumer Research, 8*(2), 197–207.

- Card, Stuart K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann.
- Cavanagh, J. P. (1972). Relation between the immediate memory span and the memory search rate. *Psychological Review*, 79(6), 525–530.
- Chang, R., Ziemkiewicz, C., Green, T. M., & Ribarsky, W. (2009). Defining insight for visual analytics. *IEEE Computer Graphics and Applications*, 29(2), 14–17.
- Clark, H. H. (1996). Using language. *Computational Linguistics*, 23(4).
- Cockburn, A., & Williams, L. (2001). The costs and benefits of pair programming. *Extreme programming examined*, 223–248.
- Cooper, A., Reimann, R., & Cronin, D. (2007). *About face 3: the essentials of interaction design*. Wiley-India.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, 24(01), 87–114.
- Cowan, N. (2005). *Working memory capacity*. Psychology Press.
- Davidson, J. E. (1995). The Suddenness of Insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight* (pp. 125-156). Cambridge, MA: MIT Press.
- Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin*, 89(1), 63–100.
- DeWalt, K. M., & DeWalt, B. R. (2002). *Participant observation: a guide for fieldworkers*. Rowman Altamira.
- Dictionary.com Unabridged. (2011). Retrieved from <http://dictionary.reference.com/browse/insight>
- Dominowski, R. L., & Dallob, P. (1995). Insight and problem solving. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight* (pp. 33–62). Cambridge, MA: MIT Press.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.
- Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., Lally, A., et al. (2010). Building Watson: An Overview of the DeepQA Project. *AI Magazine*, 31(3), 59–79.

- Gick, M. L., & Lockhart, R. S. (1995). Cognitive and affective components of insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight* (pp. 197-228). Cambridge, MA: MIT Press.
- Gotz, D., Zhou, M. X., & Aggarwal, V. (2006). Interactive visual synthesis of analytic knowledge. *2006 IEEE Symposium On Visual Analytics Science And Technology* (pp. 51–58).
- Green, T., Ribarsky, W., & Fisher, B. (2008). Visual analytics for complex concepts using a human cognition model. *IEEE Symposium on Visual Analytics Science and Technology: VAST* (Vol. 8, pp. 21–23).
- Grinstein, G., O'Connell, T., Laskowski, S., Plaisant, C., Scholtz, J., & Whiting, M. (2006). Vast 2006 contest-a tale of alderwood. *Visual Analytics Science And Technology, 2006 IEEE Symposium On* (pp. 215–216).
- Ha, D., Kim, M., Wade, A., Chao, W. O., Ho, K., Kaastra, L., Fisher, B., et al. (2007). From Tasks to Tools: A Field Study in Collaborative Visual Analytics. *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on* (pp. 223–224).
- Hanrahan, P., Stolte, C., & Mackinlay, J. (2007). *Visual Analysis for Everyone: Understanding Data Exploration and Visualization*. Tableau White Paper, Tableau Software Inc., Online: http://www.tableausoftware.com/docs/Tableau_Whitepaper.pdf.
- Heer, J., Mackinlay, J., Stolte, C., & Agrawala, M. (2008). Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 1189–1196.
- Hetzler, E., & Turner, A. (2005). Analysis experiences using information visualization. *IEEE Computer Graphics and Applications*, 24(5), 22–26.
- Heuer, R. J. (1999). *Psychology of intelligence analysis*. United States Govt Printing Office.
- Hoffman, D. D. (2000). *Visual intelligence: How we create what we see*. WW Norton & Company.
- Jorgensen, D. L. (1989). *Participant observation: A methodology for human studies*. Sage Publications, Inc.
- Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neurosciences*, 7(2), 41–45.
- Just, M. A. (2002). A Capacity Theory of Comprehension: Individual Differences in Working Memory Marcel A. Just and Patricia A. Carpenter. *Cognitive modeling*, 131.

- Kang, Y., Gorg, C., & Stasko, J. (2009). Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study. *IEEE Symposium on Visual Analytics Science and Technology, 2009* (pp. 139–146).
- Kapler, T., & Wright, W. (2005). Geotime information visualization. *Information Visualization, 4*(2), 136–146.
- Kean, T. H., Hamilton, L. H., Ben-Veniste, R., Kerrey, B., Fielding, F. F., Lehman, J. F., Gorelick, J. S., et al. (2004). *The 9/11 Commission Report*. Storming Media.
- Kinnison, H. A. (2004). *Aviation maintenance management*. McGraw-Hill Professional.
- Kounios, J., & Beeman, M. (2009). The Aha! moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science, 18*(4), 210–216.
- MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., & Hetzler, E. (2005). Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science, 32*(3), 139–160.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *V, 63*, 81–97.
- Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of “unstructured” decision processes. *Administrative science quarterly, 24*–275.
- Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104). Englewood Cliffs, NJ: Prentice-Hall.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The PageRank Citation Ranking: Bringing Order to the Web.
- Patterson, E. S., Woods, D. D., Tinapple, D., Roth, E. M., Finley, J. M., & Kuperman, G. G. (2001). Aiding the intelligence analyst in situations of data overload: From problem definition to design concept exploration. *Institute for Ergonomics/Cognitive Systems Engineering Laboratory Report, ERGO-CSEL*.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Sage Publications.

- Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Proceedings of International Conference on Intelligence Analysis* (Vol. 2005, pp. 2–4).
- Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review - New York*, *106*, 643–675.
- Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems Management Cybernetics*, *13*(3), 257–266.
- Rensink, R. A., & Enns, J. T. (1995). Preemption effects in visual search: Evidence for low-level grouping. *Psychological Review*, *102*(1), 101–129.
- Risch, J. S., Rex, D. B., Dowson, S. T., Walters, T. B., May, R. A., & Moon, B. D. (2002). The STARLIGHT information visualization system. *Information Visualization, 1997. Proceedings., 1997 IEEE Conference on* (pp. 42–49).
- Rosling, H. (2007). Gapminder. *About Gapminder*. Retrieved from <http://www.gapminder.org/about-gapminder/history/>
- Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993). The cost structure of sensemaking. *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 269–276).
- Ruthraff, E., Remington, R. W., & Johnston, J. C. (2001). Switching between simple cognitive tasks: The interaction of top-down and bottom-up factors. *Journal of Experimental Psychology*, *27*(6), 1404–1419.
- Saraiya, P., North, C., & Duca, K. (2005). An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 443–456.
- Saraiya, P., North, C., Lam, V., & Duca, K. A. (2006). An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 1511–1522.
- Scholtz, J. (2006). Beyond usability: Evaluation aspects of visual analytic environments. *IEEE Symposium on Visual Analytics Science and Technology, 2006* (pp. 145–150).
- Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology General*, *122*, 166–166.
- Schwenk, C. R. (1984). Cognitive simplification processes in strategic decision-making. *Strategic Management Journal*, *5*(2), 111–128.

- Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. *Proceedings of the 2006 AVI workshop on Beyond time and errors: novel evaluation methods for information visualization*, 1–7.
- Simon, H. A. (1978). Information-processing theory of human problem solving. *Handbook of learning and cognitive processes*, 5, 271–295.
- Simon, H. A. (1997). Administrative behavior: A study of decision-making processes in administrative organizations. Free Pr.
- Singhal, A. (2001). Modern information retrieval: A brief overview. *IEEE Data Engineering Bulletin*, 24(4), 35–43.
- Strauss, A. L., & Corbin, J. M. (1998). Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage Publications, Inc.
- Thomas, J. J., & Cook, K. A. (Eds.). (2005). *Illuminating the path: The research and development agenda for visual analytics*. Phoenix: IEEE Computer Society Press.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *nature*, 381(6582), 520–522.
- Trafton, J. G., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: extracting and generating information from complex visualizations. *International Journal of Human-Computer Studies*, 53(5), 827–850.
- Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology*, 8(2), 194–214.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95(1), 15–48.
- Trickett, S. B., Fu, W. T., Schunn, C. D., & Trafton, J. G. (2000). From dippy-doodles to streaming motions: Changes in representation in the analysis of visual scientific data. *Proceedings of the twenty second annual conference of the cognitive science society* (pp. 959–964).
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185, 1124–1131.
- Whiting, M. A., Cowley, W., Haack, J., Love, D., Tratz, S., Varley, C., & Wiessner, K. (2006). Threat stream data generator: creating the known unknowns for test and evaluation of visual analytics tools. *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation*

- methods for information visualization*, BELIV '06 (pp. 1–3). New York, NY, USA: ACM. doi:10.1145/1168149.1168166
- Williams, L. (2002). Integrating pair programming into a software development process. *In Proceedings of 14th Conference on Software Engineering Education and Training, 2001.* (pp. 27–36).
- Williams, L., Kessler, R. R., Cunningham, W., & Jeffries, R. (2002). Strengthening the case for pair programming. *IEEE Software*, 17(4), 19–25.
- Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., & Crow, V. (1995). Visualizing the non-visual: spatial analysis and interaction with information from text documents. *Information Visualization, 1995. Proceedings.* (pp. 51-58). Presented at the Information Visualization, 1995. Proceedings.
- Zsombok, C. E., & Klein, G. A. (1997). *Naturalistic decision making*. Routledge.