A Study Of Factors Contributing To Self-reported Anomalies In Civil Aviation

Chris Andrzejczak
University of Central Florida

A STUDY OF FACTORS CONTRIBUTING TO SELF-REPORTED ANOMALIES IN CIVIL AVIATION

by

CHRIS ANDRZEJCZAK
B.A. Scientific and Philosophical Studies of Mind, Franklin & Marshall College, 2002
M.S. Human Factors and Systems, Embry-Riddle Aeronautical University, 2004

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2010

Major Professors: Waldemar Karwowski
Mansooreh Mollaghasemi
Gene Lee
Piotr Mikusinski

ABSTRACT

A study investigating what factors are present leading to pilots submitting voluntary anomaly reports regarding their flight performance was conducted. The study employed statistical methods, text mining, clustering, and dimensional reduction techniques in an effort to determine relationships between factors and anomalies. A review of the literature was conducted to determine what factors are contributing to these anomalous incidents, as well as what research exists on human error, its causes, and its management.

Data from the NASA Aviation Safety Reporting System (ASRS) was analyzed using traditional statistical methods such as frequencies and multinomial logistic regression. Recently formalized approaches in text mining such as Knowledge Based Discovery (KBD) and Literature Based Discovery (LBD) were employed to create associations between factors and anomalies. These methods were also used to generate predictive models. Finally, advances in dimensional reduction techniques identified concepts or keywords within records, thus creating a framework for an unsupervised document classification system.

Findings from this study reinforced established views on contributing factors to civil aviation anomalies. New associations between previously unrelated factors and conditions were also found. Dimensionality reduction also demonstrated the possibility of identifying salient factors from unstructured text records, and was able to classify these records using these identified features.

To all the dreamers, whose efforts created travel across the skies, and to those that brave them

ACKNOWLEDGMENTS

The author would like to acknowledge the Dr. Waldemar Karwowski for "selling his best ideas" which lead to this work. In addition Dr. Karwowski has a knack for always "asking the right questions" for which I am very grateful. The spontaneity and flexibility of the entire dissertation committee allowed this work to pass key milestones in an efficient and timely manner.

My sister, Olivia Victoria, whose tireless efforts and inspiring efforts sometimes make me feel like the younger sibling is also responsible for the progression and polish of this work. To my Parents, Andrej and Barbara, who taught me the value of love and dedication, and who made sure this work was on track and making progress on a daily basis.

The UCF Library team delivered documents, advice, and occasional therapy, they are truly underappreciated and the unsung enablers of many a project. The IEMS office staff deserves mention and my gratitude for their constant support and role in enabling students to pursue their academic dreams, whatever they may be.

And finally to my dearest friends who listened to me whine and offered their advice and their cheerleader pom poms, thank you so much!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER ONE: INTRODUCTION

The demand for worldwide air travel continues to increase. In addition, the business model of commercial aviation is continually evolving, with more direct flights between city-pairs on smaller aircraft replacing more traditional "spoke and hub" operations where large aircraft connect major cities, thus requiring de-boarding and connecting flights on smaller aircraft. The changing operational environment comprises longer flight times between city-pairs, increased aircraft capability and complexity, and more airplane traffic, all of which, taken together, create many opportunities for anomalous events.

Accidents in day-to-day aviation operations are rare. However, aircraft anomalies occur much more frequently. These anomalies mimic conditions that lead to accidents. Understanding what factors are present and how they are associated with anomalies can lead to methods aimed at reducing or otherwise managing factors before they lead to serious incidents.

Nagel (1988) reports that 90 percent of aviation mishaps are labeled as and attributed to human error. Studies conducted by Lautman and Gallimore (1987) report that about 70 percent of accidents in commercial jet transport can be attributed to crew error. This percentage is consistent over any time period under review, and has not changed in recent times despite implementations of new technologies and findings from human factors and related safety disciplines. An understanding of what factors are present when anomalous events occur will strongly aid in managing or preventing future anomalies.

The Federal Aviation Administration (FAA) and National Transportation Safety Board (NSTB) keep detailed accident reports of commercial aircraft incidents, in an effort to use knowledge gleaned from such incidents to prevent future problems. Analysis of this data using methods from the fields of traditional statistical analysis, human factors studies, clustering, and dimensionality reduction may yield new information on causes and provide insights into what conditions are present for similar error types. This information may be used to influence design, training, or operations that have the potential to further reduce error.

Today's high performance computers and the vast storage capabilities of these computers constitute unprecedented opportunities for data creation and archiving. Many accident and incident databases exist, yet pertinent information may be overlooked in all of these data. Methods are required to reduce the dimensionality and "noise" in all of these data while leaving relevant structural information intact.

Maintaining vast stores of information is only useful if this information is organized and retrievable when needed. Classifying the data using human operators is tedious and possibly inaccurate, as two individuals may classify the same record differently. Semi-supervised and unsupervised accurate, reliable classification algorithms and applications would greatly increase the value of maintaining vast data stores of incident or anomaly data.

Problem Statement

Through advancements in training, hardware, and regulations, aviation accident rates have declined drastically since data began being formally collected in the 1950s and 1960s. However, rates of decreases in aviation accidents have slowed noticeably in the past 30 years. Human and mechanical contributions to accidents are routinely studied and classified.

Accidents are seemingly rare when viewed in light of the many safe flights completed every day. Aircraft anomalous events however, are more numerous. The causal factors contributing to these anomalies are not widely understood, though the conditions surrounding anomalies are often identical to those surrounding accidents. Surveying the more plentiful anomaly data will yield valuable insights as to what factors contribute to the occurrence of aircraft anomalous events. Understanding these factors can lead to improved management of said factors, and in turn, provide additional insight into how to further decrease the rate of accidents.

CHAPTER TWO: REVIEW OF THE LITERATURE

Accident Databases

The FAA maintains databases on multiple aspects of the National Airspace System (NAS) (Hendricks, 1987). The databases are employed to identify potential areas of concern, and to drive recommendations for improvement. A significant element of this system is the human performance and human factors portion. The FAA databases concerned with human error are the near midair collisions (NMACs), pilot deviations, and operational errors.

The NMAC database tracks near collisions and categorizes them according to severity, with the highest speed, smallest distance between aircraft scoring the highest level of severity. These classifications, however, can be highly subjective, and reporting of incidents is voluntary. Pilots may not report collisions for fear of penalty. There also appears to be a lack of causal factor information. An interesting development stemming from NMA data is that the number of reports has steadily increased – likely due to increasing traffic – though the level of actual collisions has remained consistent for the study period (1977-1986).

The pilot deviation (PD) database tracks those incidents where a Federal Air Regulation (FAR) was violated or some bounding area or interdiction zone was crossed. The PD database is an information store, and little causal information exists.

The Operational Error (OE) database tracks occurrences related to ATC shortcomings where appropriate separation between aircraft is violated.

A fourth database, maintained by NASA, is called the Aviation Safety Reporting System (ASRS). This is a voluntary self-report system to which pilots who experience aviation incidents, accidents, mishaps, or close calls can submit reports of the incident. The data is then complied and analyzed for the purpose of improving the national aviation system. The purpose of the collection of the incident reports is to decrease the likelihood of future aviation accidents through a "lessons-learned" approach.

One of the factors differentiating the ASRS system is its voluntary report nature. Reports are completely anonymous; the original report data is destroyed after entry into the electronic database, thereby further protecting the identity of the report-maker. Reports are not compulsory to file even in the event of an incident. In addition, the FAA has agreed to not use ASRS information in its own enforcement activities and investigations, and even to waive fines and penalties (with certain exceptions) for unintentional violations of regulations reported to ASRS. Another defining characteristic of the ASRS database is its narrative section, which includes both a section written by the individual making the report as well as a synopsis written by an analyst.

Human Error Models

One of the most popular approaches to human error analysis and modeling is the cognitive perspective. This approach likens the human mind to an information processing system. Inputs from the senses are integrated and transformed into behavior. Many types of cognitive models exist, though some models are better suited for certain elements of cognition than others. The models attempt to break down human behavior into discrete processing steps and identify at which step an error occurred, allowing for classification and management of the error. A contrasting perspective, to illustrate the concept, is that of the behavioral perspective. The behavioral perspective views human performance in terms of rewards and the avoidance of unpleasant situations or admonitions from superiors (Skinner, 1974). Neither perspective can explain all of human behavior, but they model and shed light on various aspects.

The fields of psychology, cognitive science, human factors, and safety science all have different methods of classifying and managing various errors. Errors are important because of their possible drastic effects on human reliability and performance, which directly affect manned system reliability and performance. Rasmussen (1983) clearly states the need for predicting human performance in automated systems, which is due to the high risk involved in the potential for accidents given the amount of responsibility now attributed to the automation. The automation cannot account for all of the unforeseen states and conditions it may experience. This is why there is still a human "supervisor" presiding over the automation, monitoring and intervening as necessary. However, like the automated system, the human's actions are defined by a set of constraints between the environment, actions, and effects. Human actions related to system operation are based on internal representations that are based on these constraints. Thus, by defining and representing the constraints employed one can categorize and relate various human behaviors (Rasmussen, 1983).

When human behavior fails to satisfy a goal, the cause is called an error. Errors are viewed as those situations where an intended or planned set of actions does not meet expected outcomes, and no chance event could be blamed for the failure of these outcomes (Reason, 1990). Reason further defines other types of errors, such as slips and mistakes. Slips occur when an execution fails; the planning element is irrelevant to a slip. A mistake, by contrast, does not consider the actual execution of desired action; it is the result of improper or incorrect planning. The decisions made regarding how to achieve a goal are involved when a mistake occurs, not the execution stage itself. Slips are more apparent to the observer, while mistakes are much more subtle and difficult to detect.

Due to their more complex nature, mistakes can further be classified into two categories: failures of expertise, where knowledge is incorrectly applied, and lack of expertise, where necessary knowledge does not exist. Rasmussen (1983) divides mistake types into rule-bused and knowledge-based forms. These error types and their primary areas of occurrence are shown in Table 1.

Table 1: Classification of error types, reproduced from Reason (1990)

Cognitive Stage	Primary error type
Planning	Mistakes
Storage	Lapses
Execution	Slips

Rasmussen (1986) has grouped these error types into respective levels of information processing. These levels are skill-based, rule-based, and knowledge-based processing levels and therefore also error types. Slips occur at the skill-based level, where a rehearsed and learned action goes awry. Mistakes, on the other hand, are more complex, occurring at either the rule- or knowledge-based processing stage.

These information processing steps can occur in hierarchies or loops for complex processes. Different levels of processing can "stack up," either worsening or improving a situation. For example, a healthy adult human walking is employing skill-based processing, as walking is well-rehearsed at this point in the adult's development and is practically automatic from a cognitive workload point of view. The adult does not have to actively think and process the actions that involve walking. However, should a slip occur and balance become unstable, it can be argued that the adult will employ knowledge-based processing, employing sensory feedback information to command novel movements to recover balance. Had the adult been studying martial arts or dance movements, perhaps these newly learned "rules" of movement could be applied to restore balance, as martial arts and dancing moves result in the formation of new "rules" for balance in for new students. The preceding example illustrates how skill-, rule-, and knowledge-based processing relate.

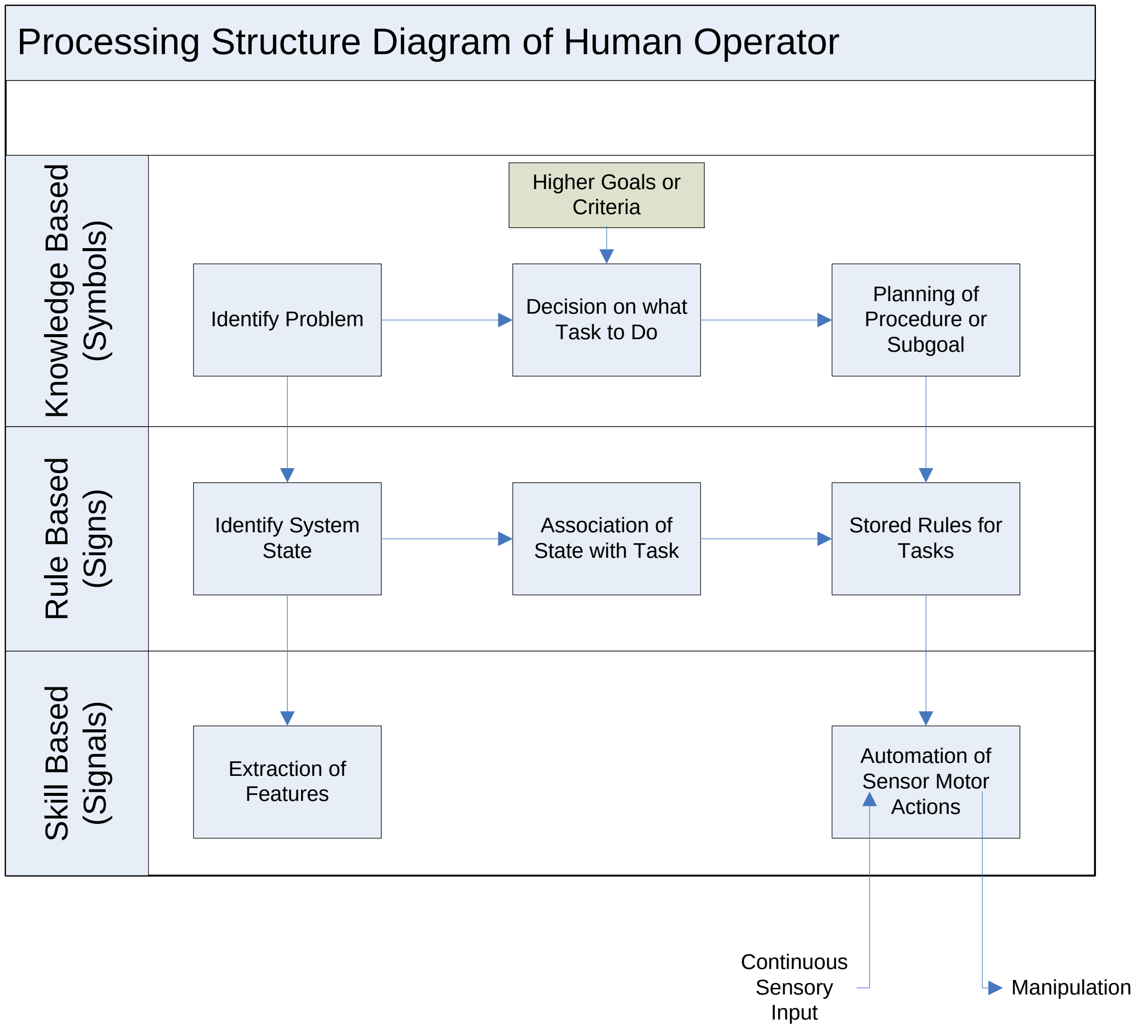


Figure 1: Relationships between types of errors

Studies of errors require differing means of investigation. Investigative methods include self-reporting activities, observation comprising both synchronous and asynchronous methods, and various evaluations comprising multiple levels of abstraction and fidelity.

Sources of Error

Many sources of error exist within the aviation domain. Weiner (1987) discusses these sources, the first being design-induced error. These errors occur when design of procedures, controls, or other interactive elements induce error. For example, handles that activate completely different functions yet are placed near each other and are similar in shape can lead to confusion and accidental triggering of unwanted aircraft functions.

Errors occur when outdated or arbitrary naming conventions are given to navigation points. Phonetic spellings, unconventional pronunciations, shortened naming conventions and frequent "double-checking" of orders and communications between pilots and air traffic controllers all increase workload and invite opportunities for error. Pilots unfamiliar with these discrepancies would have to check charts or communicate unnecessarily with controllers; both are high-workload tasks.

Errors can happen when meanings are misconstrued, or when intended meanings are not interpreted correctly. Weiner (1987) expounds on this by describing an incident where a pilot commanded a flight engineer for “take-off power,” which was interpreted as “take off the power,” and resulted in the loss of an Air Force C-124.

Finally, an often-cited contributor to error is “loss of situational awareness (SA).” The term situational awareness has been used as a buzzword in the industry, with its formal definition often forgotten or obscured. The classic definition of situational awareness has three levels (Endsley, 1995).

1. Level 1 – Perception of appropriate aspects of environment
2. Level 2 – Understanding how these aspects relate to goals
3. Level 3 – Ability to predict future states of the environment

A good pilot always strives to maintain all three levels of SA. Flight training focuses on teaching pilots how to scan their environment and instruments for important information. Experience, knowledge, and practice hone observational skills to predict performance and constantly achieve goals or missions.

Error Management Methods

The aviation domain has a rich history of attempting to manage error. The field of Human Factors owes its origins to WWII military flight operations. This is largely due to the fact that aircraft operation is a very demanding task with many opportunities for error. Flight activities are prone to error and high workload because aircraft capabilities far exceed human capabilities. Pilots often employ the phrase, "staying ahead of the airplane," meaning that the pilot must plan and be ready to respond to changes in aircraft operation. For general aviation, during seemingly uneventful cruise flight, the unofficial recommendation is for pilots to stay five minutes "ahead" of the airplane. This means that the pilot should know where the airplane will be in five minutes, and what its state will be. This practice of predicting future states from current relevant information is often called Situational Awareness (SA) (Endsley, 1995). Any unexpected changes call for changes to this plan, and will likely shorten the amount of time the pilot can stay "ahead" of the airplane. The opportunity for accidents greatly increases when a pilot can no longer predict and manage airplane states, any states such as position, altitude, or even physical configuration of the aircraft control surfaces.

To mitigate the difficulty of powered flight, pilots are subjected to stringent training requirements, including medical tests to determine adequate fitness levels. Commercial pilots have increased medical requirements and receive ongoing training to keep their certificates. Many pilot aids exist, ranging from simple checklists outlining flight procedures, advanced flight training, and automation serving to reduce pilot workload. These aids exist to mitigate and manage opportunities for pilot error.

Aviation during the First World War was extremely dangerous. Investigations made on fatalities at the time found that more pilots were lost due to accidents than combat, usually resulting from a failure of the airframe itself or, more likely, the propulsion unit (Koonce, 1999).

The field of Human Factors (HF) was born to mitigate these glaring problems. Early HF methods focused on the selection and training of pilots, managing physiological stresses of flight, and improvements to the design of equipment to ensure mission success and safety of operation.

Errors are not completely random, and are not decoupled from conditions related to their occurrence. This property means they are classifiable and can have various management methods. Human Factors practitioners study human error, and develop methods of controlling it (Edwards 1998). This process, called error reduction, requires that errors are properly detected and appropriate corrections are applied. The difficulty facing researchers and investigators is that evidence and causes of human error are not immediately apparent, and often are qualitative in nature. Compared this to the crisp, digital nature of mechanical failures, it is easy to see why it so easy to classify all mishaps as "human error." The example given by Wiegmann & Shappell (2003) is that of a fatigued bolt that may have failed, versus a fatigued pilot. Tools and very discrete measures exist to examine metal fatigue to a high degree of accuracy, but identifying mental fatigue requires inference and knowledge about the person's recent activities and rest levels. Quantifying the highly variable human is very challenging; there are no discrete reference standards that apply to all humans or an all-encompassing approach that works for every individual.

Progress introduced by the fields of Human Factors and associated error management tools can be illustrated by the investigations of Engle and Lott (1979) on WWI aviation-related fatality data. For each 100 aviation deaths in WWI, 2 came from enemy action, 8 could be attributed to aircraft failure, and 90 were caused by individual defects. Of these 90, 60 could be attributed to improper training. The main lesson to be learned from these data is that, although technology changes and adapts rapidly, the human operators do not.

Aviation in particular is a practice that demands high reliability. This reliability must be present in the hardware as well as the operator. Due to the highly variable nature of humans, the resulting difficulty of control and classification of this variability, their role, activities, and needs are overlooked in favor of managing the more predictable hardware. Elaborate efforts avoiding human considerations are made, which range from improving training and selection and increasing automation, to "designing out the human" completely. Modern aircraft have placed pilots in more of a supervisory role, they preside over the automatic systems during normal flight and only become directly involved with aircraft control when problems or unexpected situations arise. It is likely these situations require actions that have not been practiced, and so errors may occur as the pilot must either create new knowledge or recall infrequently accessed knowledge.

Human error can be managed and attempts are constantly being made to “design out” possibilities for human error. These strategies include, but are not limited to, training, hardware changes, software optimization, and regulation of activities. These interventions have been shown to be effective in increasing safety and reducing error. One example of this is the development of the angled aircraft carrier deck (Wiegmann & Shappell, 2003). This innovation eliminated hazardous opportunities for collisions where an aircraft taking off from the bow aborted the take-off while another was landing. The angled deck changed the direction of take-off aircraft while others could still land safely. An understanding of factors contributing to unsafe conditions, anomalies, or accidents was necessary to properly manage the undesirable results.

Edwards (1988) questions the future role of pilots: should they be “automated out of the system?” Humans possess abilities that current computer systems do not have, such as large associative memories, abilities to deduce relevant information, and propensity for pattern recognition. Modern Artificial Intelligence (AI) systems attempt to grant this ability to computer systems. Already, unmanned, automated flight is possible in military drone systems; just as Edwards (1988) described a future where “ground-based managers supervise several flights in a geographical sector or airborne supervisors manage individual flights.”

This supervisory activity of the human has led to the creation of a supervisory control model, consisting of five steps. These control functions were identified and broken down into their respective human and automation components by Sheridan (2010).

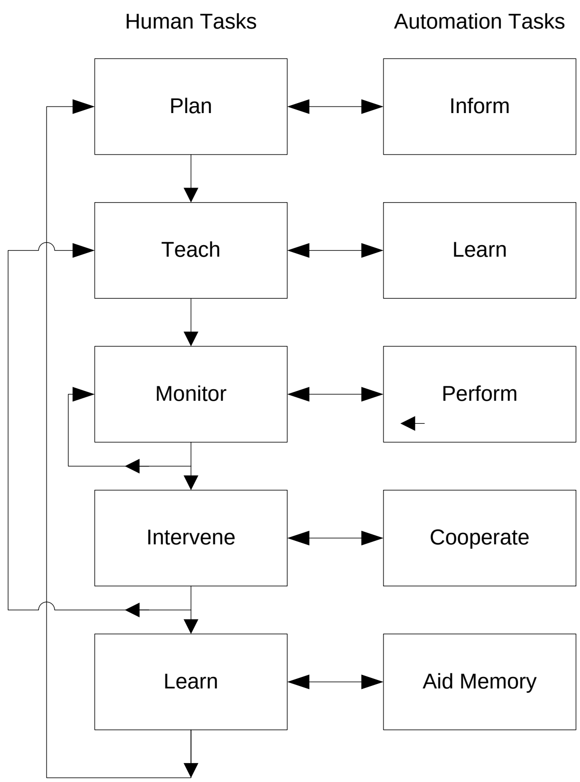


Figure 2: Supervisory control framework adapted from Sheridan (2010)

The five functions that humans perform with the aid of automation are as follows:

1. Plan – the human must predict and represent the end goal whilst the automation relays relevant information to completing the task at hand

2. Teach – the human must manipulate controls and symbols to create a representation of the end goal and completion states that the automation "understands"

3. Monitor – supervise the execution of the task, subject to intermediate constraints and

performance measures, the automation carries out the task

4. Intervene – when current state variables or conditions to not match those prescribed, modify the automation's functioning or disconnect it entirely
5. Learn – develop heuristics, shortcuts, and employ information learned to improve future performance; the automation, with its digital storage capacity, aids in relevant information retrieval

At any point in the above figure and list of activities, there may be inconsistencies, aberrations, loss of information or complete signal degradation, or outright incorrect information. Making generalizations or classifying all-encompassing breakdowns in this process is very difficult, due to the varied complexity of automation systems, different levels of pilot training, and multitudes of mission states and parameters.

Precursor Analysis

Accident investigations are not new practices. Humans naturally are curious and will continually pursue determining causes of events. However, precursor analyses can be more effective and informative in determining causation. Precursor analysis attempts to investigate prior indicators, missed signals, and dismissed alerts that, if heeded, could have prevented the accident. Bird and Germain (1996) claim that there are many more precursor events available for analysis than there are actual accidents. Precursor analyses are attractive also in the sense that they are performed on an intact system, without the added pressures or interference caused by responsible parties with something to hide, resource constraints, or unreliable witness accounts (Phimister et al, 2004). The recognition of these precursors clearly benefits safety studies in the sense that actions should be taken when such precursors manifest themselves.

Precursor analysis benefits investigators by exploiting information from near-misses. For most accident situations, near-misses occur more frequently than the actual catastrophic event (Bird and Germain, 1996). Near-misses often mirror actual accident conditions without the event occurring, affording analysis of contributing factors or scenario construction. All precursor analysis methods aim to find associations in existing conditions to reveal novel insights into accident causes and risk levels.

There are a few caveats to precursor investigations. Hindsight bias is a condition that occurs when individuals who were present at the accident would artificially inflate the risk levels associated with conditions at the accident. These individuals likely consider the accident to have been highly likely (Hawkins and Hastie, 1990). Precursor analysis could contain many disparate elements; these elements may not be given the appropriate attention and level of analysis that is normally reserved for accident investigation. Precursors are sometimes dismissed as the "lessons learned," as their application may not be as strong or salient as findings from other types of studies. However, one of the most attractive elements of precursor analyses is their low cost as compared to accident investigations. Data gleaned from precursor analyses can be used to create or enhance automated safety monitoring programs and detectors. Precursor analysis, when planned, defined, and deployed properly, can create viable action plans and recommendations to increase safety (Phimister et al, 2004).

Taxonomies, Ontologies, and Classification Schemes

Taxonomies serve as maps and guides for a group of related concepts. Individual simplified elements are mapped and plotted, and relationships are created between them to illustrate similarities and dependencies.

Classifying and organizing concepts makes them easier to work with, as human learning can be reduced to matching input patterns to output patterns. When the relationships between concepts are easily understood, the pattern matching activity is simplified. As demonstrated by the literature, error types can be classified; as aircraft became more complex, the practice of classifying pilot error became more prevalent (Stephenson, 1991). Classifying errors can form relationships between accident conditions, causes, and management strategies.

For a classification system to be effective, its outputs should have the same meaning for all users of the system (Fleishman et al 1984). That is, more usable classification systems are those where higher correlations between elements belonging to the same category reflect actual user perceptions of those items. The main goal of this classification is to provide usable and functional relationships, affording the creation of previously unknown relationships. Aircraft accident analyses that make use of databases use categorical or analytical methods. Categorical analyses group accidents by circumstances, while analytical analyses describe causal factors.

When classifying records, measures of performance are required to ensure and indicate reliable classification. One such measure of classification reliability is Cohen's *kappa*, which measures levels of agreement between classifying agents as compared to chance. The range of the *kappa* coefficient is 0 to 1, where 1 indicates perfect reliability. Acceptable levels of *kappa* range from 0.6 to 0.74; systems should strive for a value of 0.6 or better, with values over 0.8 deemed "excellent."

A taxonomy related to errors was created by Swain and Guttman (1983). Their research investigated accidents at nuclear power plants. The categories created were:

1. Errors of omission are those errors where an individual omits the entire task or omits a step in the task. These errors are failures to perform an action.
2. Errors of Commission are errors that are due primarily to poor selection. The individual selects a wrong control incorrectly, incorrectly manipulates said control, or issues an incorrect command or input.
3. Errors of Sequence are errors caused by actions that are performed out of the proper sequential order.
4. Errors of Timing are errors caused by actions that are either too early or too late.

Classifying error from accident reports and creating data on error counts can provide insights into error causes, and from these data strategies on error prevention can be created. Furthermore, these taxonomies can be linked directly to organizational processes.

There are various methods of data analysis to use when creating taxonomies. These range from technically rational approaches that look *through* the data, to exploratory approaches that look *at* the data (Wallace and Ross, 2006). Information sources do not matter to the taxonomy creation process; it must only be of sufficient quality to be analyzable. Analyzing rare and large-scale single events such as catastrophes may not reveal the most useful information as such incidents are not typical or representative of reality. General patterns and results can only be visible when large numbers of data points are analyzed.

Databases containing short reports or data can only be used effectively if classified reliably. In an applied setting, a database containing short accident reports (such as the ASRS) is only useful if the coding and classification of the reports would be coded the same way by different people (or rules). If no consensus exists in how the data are classified, then the amount of noise in the database would quickly diminish its usefulness. Wallace and Ross (2006) investigated the reliability of coding in a database for nuclear root cause event studies. They found that by sampling previously coded reports, three experienced coders had a reliability of 42 percent, where less than half of the coders applied the same codes to the previously coded reports. These findings emphasize the importance of a robust taxonomy and classification scheme.

Jarvis and Harris (2010) created a custom human factors taxonomy for inexperienced glider pilots involved in accidents. Fifty-nine categories of accident causes attributable to the pilot were created, grouped into HF elements such as judgment, handling, strategy, and attention. These topics were then linked to specific, non-general flight tasks such as approach control and failure to judge distances when flaring. Another outcome of the creation of this taxonomy was that inexperienced pilots had more accidents across all general HF elements except for "strategy," possibly implying that training or safety management should focus on other HF elements.

A taxonomy created by Wiegmann and Shappell (2003) outlines broad categories for intervention of error. This taxonomy eventually led to the creation of the Human Factors Analysis and Classification System (HFACS).

1. Environment - control of temperature, noise, vibration, lighting
2. Human - personnel selection, incentives, training, teamwork, communication
3. Machine - engineering design, capacity
4. Task - ordering or timing of events, procedures, standardization

Latent failures are defined as already existing, and they usually comprise contributing factors. Reason (1990) defines latent failures as inadequate training programs, fatigue, or negligent supervision. Active failures are those that contribute to the incident at the time of occurrence, such as reduced perceptual ability or skipping an item on a checklist. These failures "add up" or all come together to contribute to a major accident or incident. Reason (1990) illustrates this using an analogy popularly referred to as the "Swiss cheese" model.

The Swiss cheese model's cheese slices depict accident or error prevention factors, while the holes in the cheese represent failures or inabilities to manage given conditions. The model illustrates that, under everyday normal "operations" -- be they aviation, engineering, construction, or even primal behaviors such as self-preservation -- certain latent safeguards are in place to prevent failures or errors. When these latent and active conditions are performing properly, conditions leading to error are unable to pass through. There are also active behaviors or actions at work at any given moment. However, these latent and active conditions may contain aberrations or complete lacks or deficiencies, and, given the appropriate combination and interaction of these failures, a dangerous condition may pass through all of these safeguards.

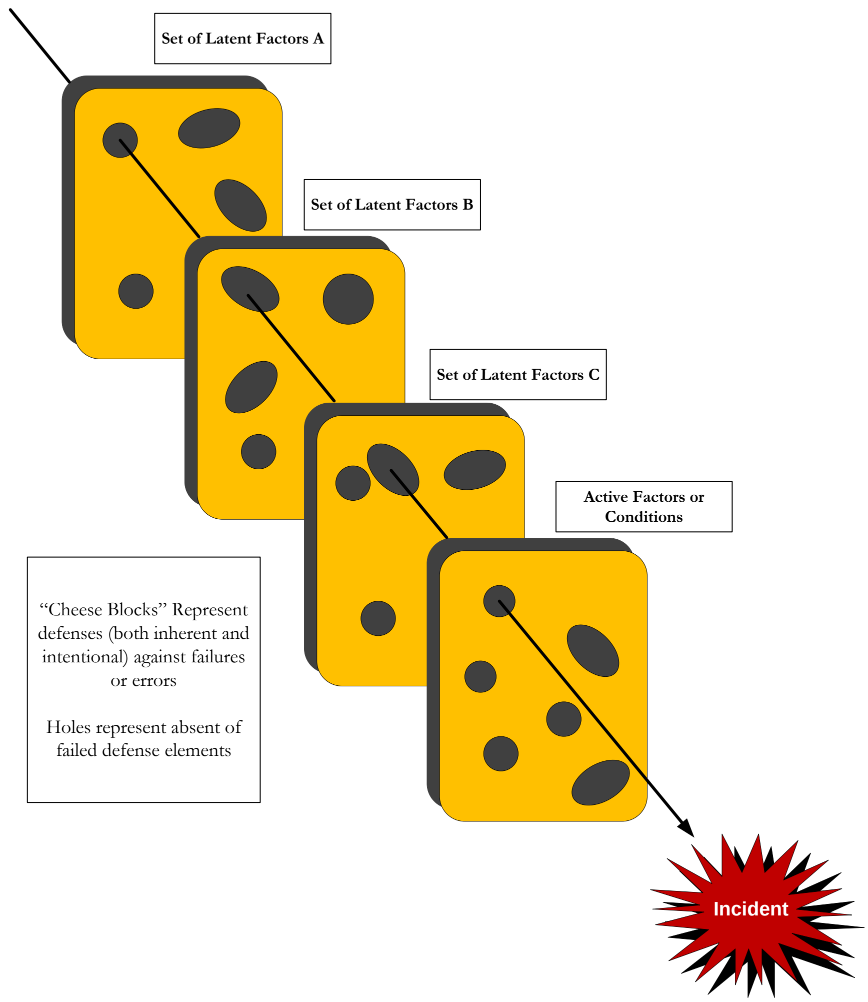


Figure 3: Swiss cheese model of latent and active failures (Adopted from Reason, 1990)

The model reminds investigators and analysts that the blame for an accident cannot always be attributed to one area, or worse, one person. Inherent safety problems can start much higher up than at the operations level: the holes in the cheese model may be introduced by management, maintenance, or design engineering even prior to operation. These inadequacies may be more dangerous than operator error (Reason & Viale, 2006).

The classification system employs four tiers, with three tiers involving latent failures, and one tier describing active failures of prevention. The three latent failure tiers describe existing conditions. The first tier describes how organizational policies may provide opportunities for decreased preventative measures (more holes in the "cheese") in the form of flawed break or rest policies, the lack of a safety-oriented culture, or undue focus on performance.

Unsafe supervisory practices can range from unreasonable organizational pressures, to lack of policy enforcement, to decreasing safety standard monitoring.

Preconditions for unsafe acts include such problems as decreased mental states resulting from fatigue, or discord emanating from personal problems. Supervisors may contribute to dangerous situations by failing to intervene when individuals argue or fight habitually. Such long-term animosities between employees may eventually end violently.

The final, "active" tier classifies operator behavior into either errors resulting from the types previously described, or by willful violations of procedures. Violations are further split into those that are routine or minor. Minor violations are not large deviations from established protocols, and usually habitual and not enforced by management, while major or exceptional violations as a result of their drastic nature, usually result in the death of the operator or large-scale destruction of equipment.

One example of a minor violation would be speeding in an automobile, exceeding the speed limit by 5-10 percent. It is important to note that routine violations may actually be condoned by management. Anecdotal evidence suggests that many state troopers will not stop a vehicle travelling 5-10 miles per hour over the speed limit, as studies show such drivers are more alert due to their constant attention to the fact they are violating the speed limit. Driving at double or triple the posted speed limit would constitute an exceptional violation, as incidents or losses of control at such speeds are usually fatal.

Wiegmann & Shappell (2003) created the Human Factors Analysis and Classification System (HFACS). HFACS employs a systematic approach that encompasses both latent and active failures that lead to an incident (Wiegmann & Shappell, 2003).

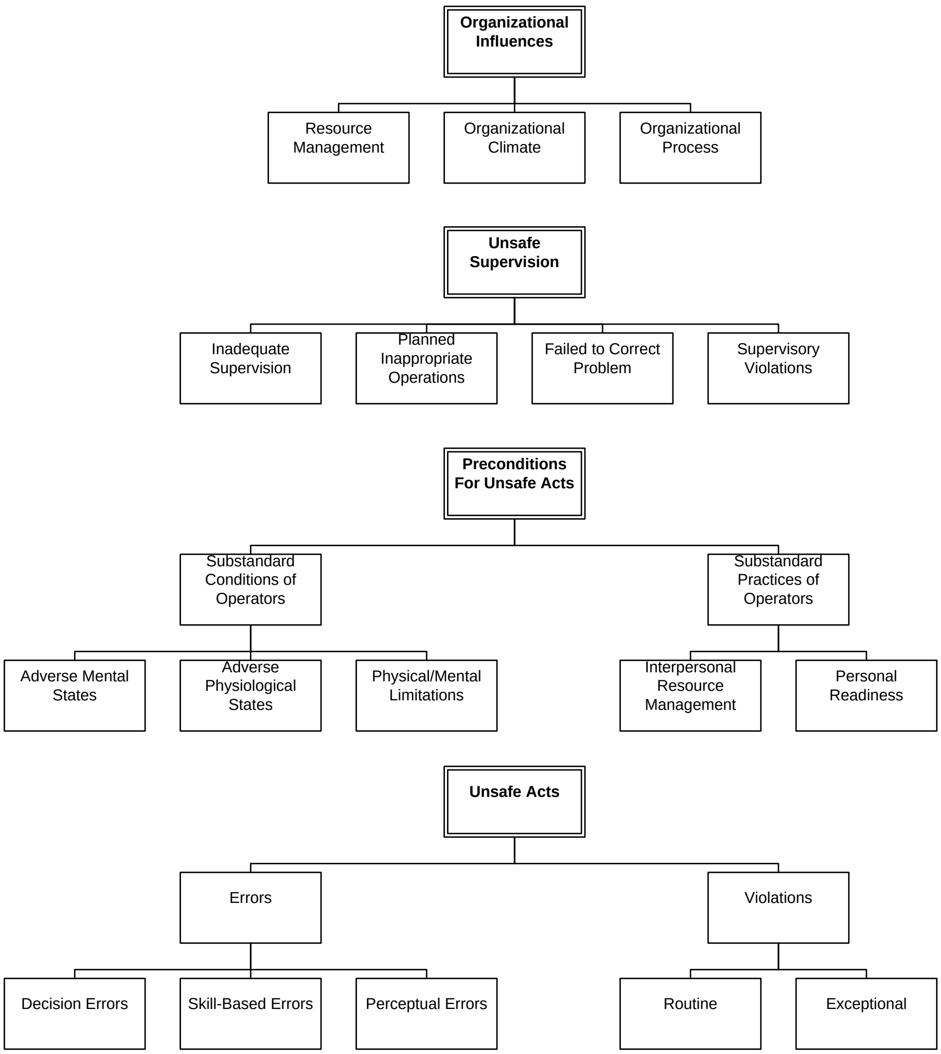


Figure 4: HFACS taxonomy breakdown

HFACS has already been employed to analyze incidents in the commercial aviation industry, including an investigatory period between 1990 and 1996. The results of the investigation defied expectations and common understandings of aviation accidents. Flight training at small flight schools and large institutions such as Embry-Riddle Aeronautical University, as well as FAA initiatives, has long drawn upon Aeronautical Decision Making (ADM), as "the wrong choice" has often been thought to be the greatest contributing factor to accidents. However, Wiegmann and

Shappell (2001) found that close to 60 percent of aircrew-related accidents were due to skill-based errors. In their study, the proportion of accidents attributable to decision-errors was less than 30 percent. Perceptual errors were not commonly accounted for in the study of 119 accidents, as only 17 of the accidents studied contained perceptual error. Upon further inspection, this idea should not be surprising. Much of everyday human life, including the operation of machinery, tends to be largely automatic, stemming from learned or trained behaviors (Boquet et al, 2004). Using this thought process, it is logical to conclude that the bulk of incidents will occur when these highly rehearsed, automatic behaviors go wrong, simply because they occupy most of human behavior. Decisions, by relative comparison, are rare, with the latent preventative factors of training, education, and reason providing significant buffers against making the wrong decision.

HFACS was again employed by Wiegmann and Rantanen (2003) to investigate the effectiveness of NASA's Aviation Safety Program (AvSP). Interestingly, the study concluded that current NASA safety intervention strategies target the machine, and not the human, environment, or task. Furthermore, almost half of the technologies developed were rated as having little to no effect on aircrew error. The study found that only one product was geared towards reducing skill-based error - the most common air crew error committed. It also found that no intervening technologies addressed civilian air violations in aircraft operation. These are among the many insights that arise from categorizing data and investigating relationships between categories.

Meaningful associations between errors and contributing factors can be made once they have been appropriately categorized. Hobbs & Williamson (2003) differentiated between causal accident models that are based on contributing factors, and their approach, which linked specific errors to underlying contributing factors. Their study was based in the domain of aviation aircraft maintenance. Aircraft maintenance is an interesting field of study as the errors that occur in maintenance can be truly latent; they may manifest themselves after much time has passed since the error, or when other conditions exist.

The authors stated that most studies of safety databases report errors and contributing factors independently of each other, listing them in separate, unlinked tables. A taxonomy of errors based on the Boeing Maintenance Error Decision Aid (MEDA) was created to describe the terminating condition of each incident. The MEDA tool is a form that aids investigations where aircraft maintenance is a contributing factor to the incident or accident. The aim of the MEDA tool is to improve aircraft maintenance operations by removing the contributing factors that lead to the incident. It departs from traditional investigations by not being solely focused on assigning blame and punitive measures, departing significantly from traditional accident investigations and alterations to current policies, which end immediately after the technician found to be at fault is punished (Rankin & Allen, 1996). Without change to company policies, the contributing factors that led to the maintenance incident still exist, and can cause similar incidents in the future because "the system is still broken."

The Hobbs & Williamson (2003) contributing factors taxonomy employed the items in the below table:

Table 2: Factors contributing to maintenance error Hobbs & Williamson (2003)

Factor	Definition
Fatigue	Inability to function at optimal levels, due to lack of rest or other physical or mental deficiency
Pressure	Increased expectations induced by management or situational conditions, shortened time available for work
Coordination	Incorrect, inadequate, or lack of communication and integration between team members
Training	Shortcomings in education and familiarity with work processes
Supervision	Improper, lack, or inadequate control of workers by management
Previous Deviation	Unsatisfactory work from a prior occurrence; error remained and was not reported or recorded
Procedures	Unsuitably designed, documented, not properly shared or not adequately enforced standard
Equipment	Poorly designed, maintained, or apparent lack of tools or other necessary implements or aids for task performance
Environment	Physical surroundings of the worker that are beyond his or her control - lighting levels, noise, temperature, etc.
Physiological	Biological problems stemming from medical conditions or individual limitations

The resulting error taxonomy used in the Hobbs & Williamson (2003) study was loosely based on familiar taxonomies created by Reason (1990), Wiegmann & Shappell, (2001), and added one category called "mischance," where a seemingly "correct" procedure was followed. This action created an unsafe condition because the procedure was incorrect or otherwise flawed.

Table 3: Errors encountered during maintenance operations Hobbs & Williamson (2003)

Error	Definition
Perceptual Error	A failure to detect a symbol or sign
Memory Lapse	Omission of an intended action
Slip	Failure of the performance of a routine, highly trained and automatic procedure
Rule-Based Error	Failure to employ familiar process used in processes already experienced or trained
Violation	Intentional deviation from standards
Knowledge-Based Error	Error resulting from inadequate or incomplete knowledge of a task
Mischance	Correct procedures were followed but behavior still lead to anomalous occurrence

After the development of the two taxonomies was complete, cross tabulations of errors and contributing factors were calculated. Relationships between the categorical variables were determined using correspondence analyses as documented by Clausen (1998). Logistic regression analyses were used to estimate association strength between contributing factors and each error type. The most prevalent associations found were those links between memory lapses and fatigue, and between rule violations and time pressure.

Concept Maps

Concept maps are similar to taxonomies. However, they describe the relationships between concepts with more detail. Terms such as, "is a part of," "gives rise to," or "results in" aid in visualizing and relating data. Concept maps are hierarchical in nature, with a main concept at the top of the map, and related elements and sub elements linked downward on the concept map page (Novak & Cañas 2006).

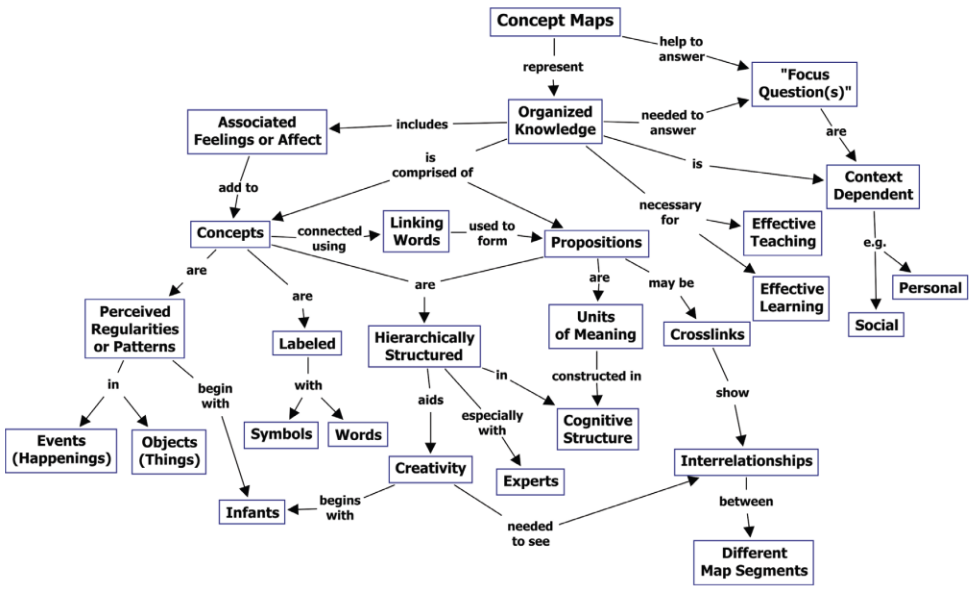


Figure 5: Concept maps: reproduced from Wikipedia Concept Maps (2010)

Concept maps are usually organized around a focus question, and this focus creates the context for the map. Concept maps were developed in 1972 as a tool to understand the evolution of a child's knowledge of science (Novak & Musonda, 1991).

Systems Thinking

Technological advancement brings with it complexity. This complexity is afforded by systems – namely, elements are grouped together as systems, each performing a designated function. These complex systems comprise many interrelated components. Though advantageous and even necessary, systems, and Systems of Systems (SoS), can bring about previously impossible and inconceivable failures due to their highly coupled nature (Perrow, 1984).

The systems approach bounds an element of reality into identifiable parts and examines the interaction between these parts (Sheridan, 1988). The art of systems thinking is found in the appropriate bounding of the system: bounds can result in systems that are trivial for real-world application or simply too complex for analysis, not unlike the problem of model development in simulation practices. Appropriate bounding and filtering of system elements depend on the purpose of investigation. The power of the systems approach comes from this bounding property, its iterative nature, and visualization ability.

One systems model conceived by Edwards (1972) is the SHEL model. The SHEL model is used when applying Human Factors principles towards design. The components of this model are:

1. Software – rules, regulations, laws, operating procedures and general methods of organization and implementation of information comprises the software component
2. Hardware – physical equipment, buildings, vehicles and materials
3. Environment – surrounding conditions, be they physical, social, political, or economic factors where the other resources are employed
4. "Liveware" – human beings, their capabilities, limitations, and performance measures

All of these components have interfaces with each other, with environment encompassing the other three. Each interface is discussed and it is the job of the human factors practitioner to optimize these interactions. Good interfaces likely prevent error, especially when control or information must be passed between SHEL model components.

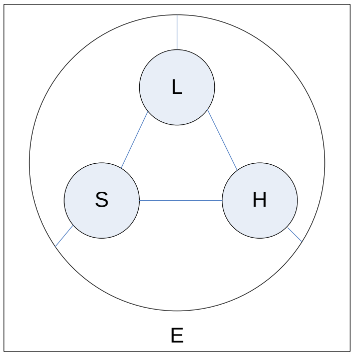


Figure 6: SHEL Model

The Liveware-Hardware and Liveware-Software interfaces are critical to aviation. Buttons, switches, displays, and controls must accurately and reliably represent information to the human operator, and these interfaces do not stop just at the airplane level. Software also encompasses regulations and operational parameters, and these regulations and external operations cannot be neglected during error analysis and prevention. The SHEL model is a good starting point for creating a new taxonomy for factors contributing to accidents or anomalies, as it can begin to account for pre-existing unsafe conditions, including mental states or organizational policies.

A well-designed, stable system exists in a state of equilibrium (Edwards, 1988). Dynamic systems with changing interfaces present a challenge to designers and managers, as well as to the operators of the system themselves. A continuous review of the system is necessary, so that variables beyond the control of designers and operators are properly managed.

Reason (1997) speculates that systems operate for long periods with inherent safety flaws present within them. These flaws are not revealed until an accident or incident occurs. Sträter (2005) claims that stakeholders may even know and tolerate these flaws because of the following two states:

Nothing happened so far in my system => it is safe

My system is safe => nothing will happen in my system

Statement (b) is a simple reversion of inference from (a); however, this state may not actually be achieved. The appropriate way to consider either of the above statements is:

There are unsafe elements in my system

Identifying and managing these elements, in most cases, prevents incidents and leads to increased safety and higher performance, as the system no longer has to recover from unsafe conditions.

Modeling and Classification

Classification models assume that the cases investigated are heterogeneous. The classification model's end goal is to segregate as many cases into each defined category as they are observed. The model is effective when it correctly predicts assignments to given categories. The model fails if it describes the sample as completely homogeneous (e.g., each individual case comprises a category). Classification models are usually employed when identical treatment of all groups is not possible (Menard 1995).

Selection models are concerned with accepting or rejecting cases from a given group or category. The cases are selected based on whether they fulfill a requirement or satisfy a criterion for inclusion in the group. Selection models may include a cutoff number for the amount of cases to be included within a group. Such models are used when a certain number of cases must be chosen from a larger population (Wiggins 1973).

Among the outputs of logistic regression is an *odds ratio*. The odds ratio is associated with a factor for an independent variable. If the odds ratio is greater than 1, the chance that the dichotomous dependent variable will be 1 increases when the associated independent variable increases (Menard, 1995). Put another way, an odds ratio greater than one means the presence or the increase of an independent variable increases the chance or odds that the dependent variable will be positive or present. Odds ratios are usually employed in epidemiological research when indicating whether the presence or severity of a factor increases the chances of a disease or disorder to be present.

A study by Thoraval et al (1997) employed logistic regression to investigate factors contributing to the severity of Hepatitis C Virus (HCV)-related liver disease. Their findings calculated the odds ratios of various factors. Those that had ratios greater than 1 and had thus contributed to the presence of liver disease were over age 30, exhibited excessive alcohol intake, and had HBsAg, the surface antigen signaling the presence of the Hepatitis B Virus (HBV). The findings of this study caused the authors to mandate abstinence from alcohol intake for HCV-infected patients, as abstaining from alcohol halved the risk of disease progression to cirrhosis in these patients.

Logistic regression models have been employed successfully to identify risk factors and their severity in bacterial colonization of chickens by Kapperud et al (1993). The authors examined flocks of chickens at various farms for presence of *campylobacter* colonization at time of slaughter. Managers of the farms were interviewed regarding hygiene and husbandry practices for the chickens. Factors found to be associated with the presence of the bacterial colonization included providing the chickens with water that had not been disinfected, tending to other poultry prior to entering the chicken house under investigation, geographic region, and season. The study concluded that the most effective preventative measure against bacterial colonization was to disinfect water provided to the chickens.

Machine Learning

The discipline of machine learning concerns itself with a type of algorithm development that allows computers to behave in a desired, usually unsupervised fashion. A major goal of machine learning is to automatically recognize complex patterns and subsequently formulate appropriate actions, decisions, or behaviors given the nature of the data. Machine learning has ties to AI, as well as probability theory and data mining. The major obstacle to machine learning is that modern programming languages and data structures are still not complex or robust enough to accommodate all possible choices for behavior. Machine learning involves statistics and computer science, but there are two major problems those disciplines must solve: efficient algorithms are required to optimize training, and efficient storage is needed for findings and related data (Alpaydin, 2004).

The field of AI has long struggled with developing an efficient means for machine learning. Vast amounts of data were fed to "expert systems" in the 1960s in an effort to give machines all the data necessary for making good decisions and approximating human behavior. These systems failed.

Lormand (1990) described the difficulties AI faces in a paper on the frame problem. The frame problem describes the issue of “teaching” a computer or artificial system to make relevant choices. The author outlines iterations of computer and robot designs, starting from C2, the causation computer, to R2D1, a robot-relevant "deducer." C2 is unable differentiate causality between different time states. That is, given the simple problem of asking the computer whether an object exists within a room before and after it is removed from the room, the computer is unable to determine the correct answer. For a human, this question is trivial.

Skipping through design process iterations to R2D1 shows a machine that is able to determine the implications of its actions when given a goal to complete. However, R2D1 considers all the possibilities of its actions, and then tags them as relevant or not. The rather comical task given to R2D1 by Lormand is to escape a room with a bomb in it that is attached to a cart containing R2D1’s power source. Poor R2D1 is still in the room deciding whether the colors of the walls will change if it decides to move when the bomb explodes. The fictional R2D2, the ideal for robot designers, one who can quickly and accurately deduce the effects of its actions and choose relevant actions given a problem, is the ultimate goal (Dennett, 1987). This problem, of using empirical data to determine appropriate actions, is the very essence of machine learning.

There are various types of machine learning:

1. Supervised learning – a function that maps inputs to outputs, akin to human learning, where a desired response given a specific input is required
2. Unsupervised learning – a model that groups inputs, very similar to clustering

3. Semi-supervised learning – algorithms that operate on data that is both labeled and unlabeled that create and verify the learning algorithm at the same time
4. Reinforcement learning – algorithms that operate on iterations where subsequent observations guide the algorithm through feedback mechanisms

Machine learning has been applied to analyzing functional magnetic resonance imaging (fMRI) data. Algorithms have been developed that create classifiers that can decode stimuli, behaviors, mental states, and otherwise extract information from complex data structures. Pereira et al (2008) created classifiers that were able to discriminate patterns, localize those patterns, and even characterize them. To do this, a classifier function that operated on various features (independent variables) of the example fMRI data was created to predict the class that the data belonged to (dependent variable). For this particular study, the features used were voxels, data elements that represent volume elements in three dimensional space. Voxels are analogous to their 2-dimensional pixels, which represent 2D image data. The voxel data were arranged in row vectors and fed to the algorithm, which classified that particular area of the fMRI dataset. Machine learning is known to operate best when data are in numerical, vector form. Using these classifiers allowed authors to produce meaningful information out of otherwise complex, high-dimensional data.

Structural Equation Modeling

Structural Equation Modeling (SEM) is a statistical method for determining and estimating causal relationships by employing statistical data and qualitative data based on given assumptions, such as survey data. SEM requires interval data that defines abstract constructs to be effective. The method of SEM allows the creation of latent variables. These are not measured directly; rather they are estimated from the model by related, measured variables. A goodness of fit test is conducted, and model unreliability is thus measured.

There are four core SEM techniques:

1. Confirmatory factor analysis - used to assess loading of variables, and quantify the number of factors present
2. Path analysis - describes dependencies among a set of variables
3. Latent growth curves - a longitudinal analysis technique that estimates growth over time (commonly used in social sciences)
4. Structural equation models - tests and estimates of causal relationships

The field of social work employs structural equation modeling. A review of the use of SEM conducted by Guo, Perron, and Gillespie (2008) focused on SEM methods used in 32 social work studies published between 2001 and 2007. Critiques and suggestions for individual practices and methods resulted from this review. Findings supported the fact that SEM analysis generated models and measures not previously tested empirically. However, a very important trend was discovered in which studies employing SEM would *meaningfully modify models without "theoretical justification or substantive interpretations"* (Guo, Perron, and Gillespie, 2008). When modifying the model, the authors suggested doing so only when theoretical justification existed, and to employ appropriate estimation procedures when doing so. Studies were also investigated for best practices, including proper theoretical specification of the model, which is important as accuracies of parameter estimates depend on both the soundness of the theory and the validity of the measurement. Study findings stressed that poor SEM practices compromise the field's knowledge base.

Self-organizing maps

Connectionist neural networks developed in the 1980s were thought to most mimic and resemble human learning in machines. These networks consist of varying layers of nodes, and, by altering the weights between the nodes, a given input pattern can be "taught" to match an output pattern. These networks work in parallel, through a process called spreading activation. Once the weights are sufficiently altered in the layers of the network, any input can be matched to any output, despite the presence of aberrations, inconsistencies, or even missing data.

Self-organizing maps (SOM) are a subset of artificial neural networks that create two-dimensional surfaces based on training samples. This property is useful to visualize high dimensional data, and large SOMs can even display emergent data (Kohonen and Honkela, 2007). This is because these SOMs are model-free – they do not have an imposed training program – and because of this property, previously undiscoverable patterns can be found, given the lack of introduced investigation bias.

SOM Tools

Self-organizing maps visualize high-dimensional data in a low dimensional grid (Kohonen, 1998). Stated another way, the SOM process converts complex, nonlinear statistical relationships into simple geometrical relationships. The SOM compresses information, but keeps metric and topological relationships between data points intact. Beyond this already-useful property, the SOM can be "taught" to produce abstractions. These abstractions, like in the SEM approach, may reveal previously unknown or unimagined relationships between data elements. Kohonen (1998) describes the following properties of SOM:

1. It consists of a two dimensional grid of nodes, each of which is associated with a model of a given observation
2. It employs an algorithm that optimizes the description of a domain of discrete or continuously distributed observations.
3. Models are formed and organized so that similar models are closer to each other in the grid than the differing ones, offering both the benefits of a similarity graph and a clustering diagram.

The SOM has already seen applied use in a variety of domains. They have been used in data-mining applications, creating similarity graphs for statistical tables and full-text document collections (Kohonen, 1998).

Dimensionality Reduction

Dimensionality reduction is a topic that has received recent attention. There is a staggering amount of data being created every day. Accessing, categorizing, and using this information requires organization. To handle “real-world,” often unstructured, high-dimensional data accurately and with minimal computational load is a challenge for mathematicians and computer scientists alike. In a review by van der Maaten et al (2009) efficient data representations should have a dimensionality level that approaches the intrinsic dimensionality of the original data. Fukunaga (1990) defines intrinsic dimensionality as the minimum number of parameters necessary to represent the functional properties of the data. The below figure displays a taxonomy of dimensionality reduction techniques reproduced from van der Maaten (2009).

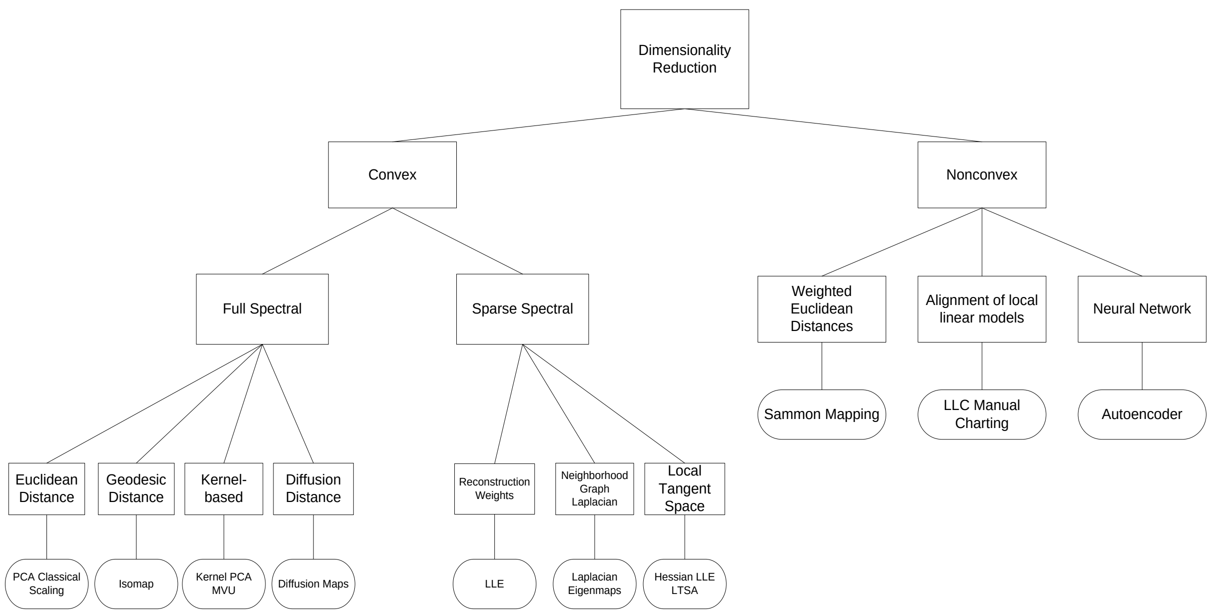


Figure 7: Taxonomy of dimensionality reduction techniques (van der Maaten, 2009)

Diffusion Maps

Diffusion maps are one of many techniques for dimensionality reduction. When dealing with multiple categories of data or many factors, visualizing the data to deduce meaning can be challenging or impossible. Dimensionality reduction assumes that the data observed has some sort of structure or logical order to it, and therefore could be reduced to a dimensional level low enough to be represented or visualized. For this to occur it is assumed that the interesting data can be represented on a non-linear manifold, or mathematical space. A manifold is an abstract mathematical concept where points exist in a domain that resembles Euclidean space. Manifolds of a sufficiently low dimensional level can be plotted or otherwise represented visually. Dimensionality reduction techniques have been successfully employed in machine learning, mapping, and clustering activities. When reducing dimensionality the goal is to maintain any underlying structures or patterns in the high-dimensional data.

Diffusion maps work by embedding high dimensional data onto low dimension Euclidean space. This is done through the eigenvectors of defined random walks performed on the data. The data is assumed to be randomly sampled from an underlying general probability distribution:

$$p(\mathrm{x}) = e^{-U(\mathrm{x})} \tag{1}$$

As the number of samples approaches infinity, eigenvectors of each diffusion map converge to the eigenfunctions of a corresponding differential operator defined on the support of the probability distribution (Nadler et al, 2006). Diffusion maps have the added benefit that, when properly employed, they quickly converge on a meaningful scheme or result. In most applications diffusion maps are unsupervised when employed. Coifman et al (2005) present a general framework for diffusion maps, and demonstrate diffusion maps' effectiveness in exploring geometry, statistics, and functions of data. The authors also demonstrate how diffusion maps afford a low-dimensional embedding of high-dimensional data. This process is naturally suited for visualization, clustering, and regression.

The diffusion map algorithm process as described by Bah (2008) employs four major steps, summarized below. It assumes the data has already been modeled by a weighted graph.

1. An adjacency matrix of the graph is created
2. A diagonal $k \times k$ normalization matrix and Laplacian matrix of the graph are calculated
3. Eigenvalues and eigenvectors are computed of these two matrices
4. The lowest value initial eigenvector is dropped, and the next m eigenvectors are used to represent the n-dimensional space.

One fascinating aspect pointed out by Bah (2008) is that diffusion maps may mirror biological functions. For example, the human brain is constantly bombarded by unstructured, highly dimensional stimuli. Diffusion maps may model the biological analogs that perform natural clustering and categorizing applications to make sense of surroundings.

Lafon and Lee (2006) describe a unified framework for employing diffusion maps to reduce dimensionality and cluster documents according to the words contained within them. Use of k-means clustering in diffusion space allows this categorization, and the authors also propose a measure of clustering accuracy used to assess the results given by the algorithm.

Dimensionality reduction has found recent use in extracting information from a corpus of text documents. Underhill (2007) states that manipulating large amounts of text data can be extremely computationally intensive. A reduced dataset with relevant meanings intact would be extremely useful in information extraction efforts. In addition, such information extraction could be unsupervised and automated, providing a way to manage the incredible amounts of information being generated.

Data mining and text mining

Data mining is a technique used to extract patterns and trends from data stores. Many agencies and individuals employ data mining techniques: these agencies range from surveillance and law enforcement applications, to nuclear power plant operations, to aircraft maintenance. These patterns and trends can be employed in both proactive and reactive methods. Data mining, at its most basic level, serves to match input data to an output cause, through a logical and appropriate mapping function (Kantardzie, 2003).

For example, data on aircraft component reliability can be used to form maintenance intervals by calculating mean time between failure (MTBF) of individual elements, thus justifying extending maintenance on long-lasting components, or implement more frequent inspections on items that fail often or unexpectedly. These trends and insights are only as good as the quality of the data used; inaccurate, falsified, or missing data points may lead to unusable or incorrect conclusions.

Data mining employs a given "sample" of a larger whole of data. One possible shortcoming of this process is that the sample may not contain a property or trend that exists in the population. Likewise, a pattern discovered in the sample may not be present in the population: statistical validation and verification of results must be properly carried out in relation to patterns found in other samples.

Humans naturally form patterns and draw conclusions from their sensory inputs, ranging from observations on trends in weather patterns to the behavior of the opposite sex. It wasn't until modern times (around the 1600-1700s) that more formal, procedural methods were required for the ever-increasing amount of data being created. Early, formal data mining procedures include Bayes' Theorem and Regression analysis (Kantardzie, 2003).

Data mining employs assumptions paired with the division of data to identify trends. Such trends may provide new insights that are not easily discernible from simple averages, means, and other descriptive statistics. Frequency counts and percentages are two frequently employed methods.

The method employed in this study will be based on the Cross Industry Standard Process for Data Mining (CRISP-DM), developed and partially funded by the European Commission under the ESPRIT program.

There are four levels of the data mining process model. These are: phases, generic tasks, specialized tasks, and process instances. The phases organize the lower operations. Generic tasks attempt to cover all data mining necessities, and should be as complete and mature as possible. This means that generic tasks can serve a variety of data mining applications and remain valid, despite the use of novel modeling techniques. The specialized task level further develops generic task steps into operations. They may, for example, clean, reorganize, or reclassify data. The final process instance level is a detailed account of actions and results of the data mining activity. It will detail what actually occurred rather than provide an overview.

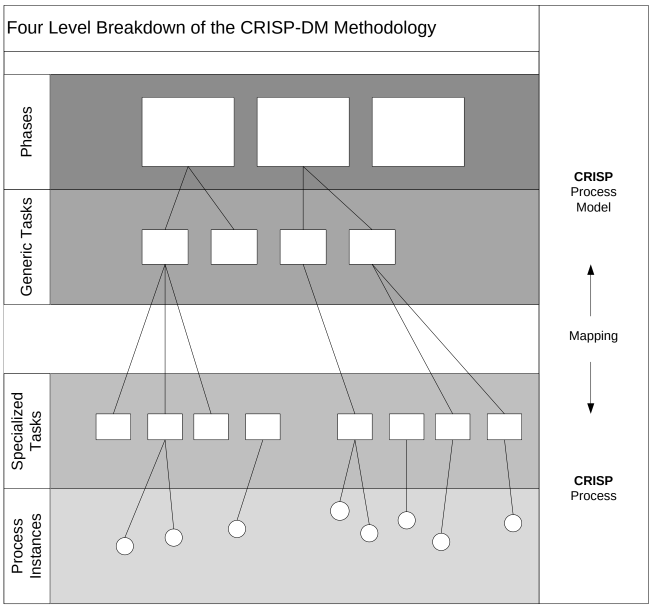


Figure 8: CRISP-DM method for data mining

The CRISP-DM method breaks the data mining process down into six steps, starting with developing a business understanding and culminating with deployment of the findings.

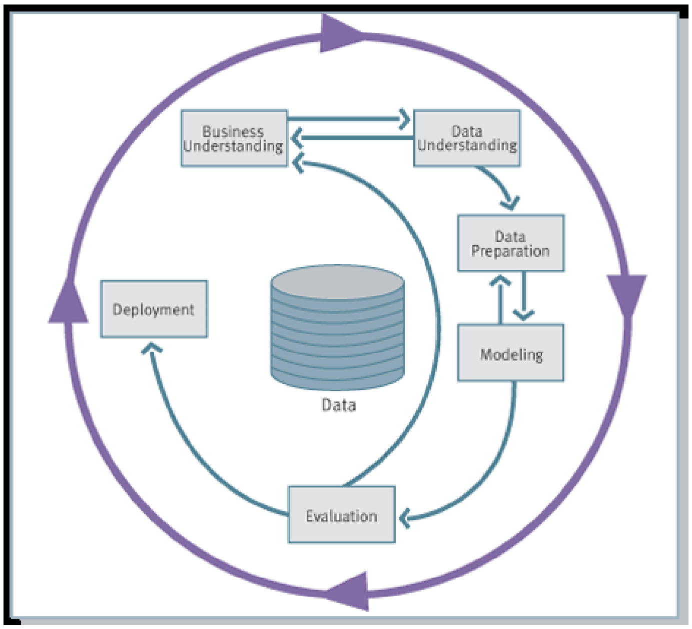


Figure 9: Overview of iterative CRISP-DM method

The six steps are not strictly defined, and the process is iterative at any of the steps. Outcomes of one step determine which step will follow. The entire process is cyclical in nature. The investigator determines which path to take or whether to proceed based on results of the prior operation. The six steps are outlined below:

1. Business understanding - the objectives and requirements of the mining project are translated from business perspectives to an achievable problem definition and preliminary plan
2. Data understanding - this step entails activities with collection and initial familiarization activities with the data; this is where quality problems are addressed and underlying hypotheses are formed
3. Data preparation - the final dataset is constructed and arranged in a way that the

modeling tool may appropriately process it

4. Modeling - appropriate modeling techniques are performed, often necessitating adjustment and validation activities; it is likely that a return to the preparation phase may occur as anomalies are discovered
5. Evaluation - a high quality model is verified to ensure its operational usefulness; a decision is made regarding the final use of the results, as well as suitability from a business perspective
6. Deployment - this phase varies depending on the model's intended use, it can range from report generation to implementation of the process enterprise-wide; the step focuses mainly on presenting the results in a useful fashion

Mining and visualizing data has become very important, as the sheer amount of data being generated in the 21st century limits its usability to draw inferences. It is estimated that every year 1 exabyte of data, mostly digital in nature, is generated by human society. This trend means that, during a study period between 2002 and 2005, more data was generated than had been in all of prior human history (Keim, 2002). Automated machine data mining techniques may not always properly make relevant associations. Keim (2002) describes the usefulness of visual methods, which integrate the human (who has a large associative memory and is able to make relevancy decisions and recognize patterns) into the data mining and exploration process. Departures from automated mining techniques include the ability to deal with noisy and nonhomogeneous data, as well as being intuitive and not requiring in-depth understanding of advanced mathematics or statistics. Visual exploration is faster and provides better results in situations where automatic data mining algorithms fail.

Data mining has been employed successfully by Yang & Chou (2003) in an effort to assign biological functions to genes, proteins, and compounds. Varying data mining methods were utilized, including decision tree and self-organizing map procedures. The decision tree method involved extracting some rules for matching gene characteristics to gene expression, and employing them for decision making. However, this method, heavily reliant on frequency estimation, did not take into account interactions between certain biological constructs. The lack of interaction information reduced the predictive ability of the decision tree method. It was concluded that the predictive rules derived from the SOM method were much more robust than the decision tree method. The SOM rules were also shown to be more accurate than the decision tree rules, and much less sensitive to noise.

Literature Based Discovery (LBD)

Text mining uncovers information from otherwise unorganized document sources. Modern computers can store unprecedented amounts of data due to their vast processing and storage capabilities. However, this information is irrelevant and useless without a means to extract meaningful information and relationships. Human beings are excellent at discovering meanings and patterns, but lack the processing capabilities of computers to investigate large volumes of text. Computers, on the other hand, do not have advanced associative capabilities.

Underhill (2007) describes a serendipitous incident where a medical researcher uncovered a novel treatment for a disease. The researcher, Swanson (1986) was investigating Reynaud's disease. Swanson searched the Medline database for symptoms associated with Reynaud's disease, which were high blood viscosity and platelet aggregation. A second search on just the symptoms revealed that fish oil treats said symptoms. As a result, Swanson reasoned that fish oil can help manage Reynaud's disease and, after some research and clinical trials, this idea was found to be true. This novel thinking and mining of the medical database extracted novel information and created a new association. It is currently extremely difficult for computers to perform such an operation because most text records are represented in unstructured natural language.

Clustering

Clustering is a method of classification: it divides data into logical subsets. The subsets are defined by similarity, and this similarity is defined by some measure of distance between items (Gibson et al, 1998). Clustering is employed in machine learning, pattern recognition, trend analysis, and text mining. It simply groups similar items in a given dataset. Clustering works ideally with numerical data, as Euclidian distance between items is easily calculated. It is far more challenging to cluster categorical or nominal data, as relations can be completely arbitrary, subjective, or otherwise difficult to quantify.

Gibson et al (1998) developed an approach for clustering set data. The names of items, data not numerical in nature, cannot be ordered naturally by a systematic method. The approach assigned and propagated weights ascribed to the categorical values in a table, thus creating a similarity measure. The results were presented as non-linear dynamical systems, thereby creating a connection between tables of categorical data and said systems. The developed system was effective at uncovering similarities and sub-populations in various types of data. As exhibited by most diffusion maps, Gibson's systems quickly converged and did not require large computational effort.

A popular algorithm for clustering is the k-means method. The algorithm was first introduced by Cox (1957). It was later popularized for partitioning large numerical data sets with by MacQueen (1967). The *k*-means method is employed in the fields of statistics and machine learning. The goal of the algorithm is to partition *n* observations into *k* clusters. The clusters are based on means. To successfully implement the algorithm, *k* must be specified. Using too large or too small a *k* value will yield undesirable results.

K-means operates by calculating Euclidean distance between observations. To compute a measure of clustering breadth or scatter, variance of the observations is employed. Algorithms that cluster data into partitions utilize a cost function of the type

$$\zeta = \sum_{i=1}^{n} \left\| d_i - C_j \right\|^q \tag{2}$$

In the above equation, n is the number of elements in the data set, d_i is the data object, C_j is the center of the jth cluster, and q is an integer defining the nature of the distance function (for example, 2 is used for Euclidean distance). K-means simplicity also has its drawbacks, as a correct *k* must be specified and traditional, unmodified k-means clustering only operates on numerical data

(Hartigan, 1979). Categorical variables do not work well as there is no natural ordering among the values of each category.

Huang (1997) altered the k-means algorithm to allow clustering of categorical values without sacrificing the algorithm's efficiency. A new algorithm was introduced, k-prototypes, which seeks to maximize intra cluster similarity of observations entered into it.

$$\zeta = \sum_{i=1}^{n} v(d_i, C_j) \tag{3}$$

The quantity to the right of the summation operator represents the distance of a data object from the closest cluster center. A dissimilarity measure for categorical objects is employed, and modes are used instead of means for clusters. A frequency-based process updates the modes that minimize the cost function. The algorithm comprises three processes: selection, allocation, and re-allocation. The algorithm concludes its iterative process once a local optimum is reached. A typical application to a large data set containing 75,808 records and 20 attributes resulted in 64 similar clusters.

Ahmad and Dey (2007) proposed a new k-mean clustering algorithm that works for mixed numeric and categorical data. A new cost function and distance measure based on co-occurrence of values was proposed. In addition, the proposed algorithm accounted for significance of an attribute towards the clustering operation. Finally, the authors presented a modified definition of a cluster center that no longer required numerical data to characterize the cluster. The new representation of the cluster center demonstrated that it captured cluster characteristics well as contained the distribution of all categorical values in a given cluster. The resulting algorithm was tested with positive results across various real world datasets, ranging from political voting choices by party, to heart disease data, to credit card usage in Australia.

San, Huynh, and Nakamori (2004) introduced an extension to the k-means algorithm that employed a concept called "cluster centers" for dealing with categorical data without sacrificing efficiency. The authors noted that Huang's (1997) algorithm, which replaces means with modes to determine clusters, can become unstable due to the non-unique nature of the modes. Selection of the modes strongly influences the results of the clustering process. By using their "cluster centers" notion, San et al (2004) used fuzziness rather than means to determine clusters. By no longer relying on mode selection and allowing fuzzy clustering to organize the data, the authors found about a 69 percent chance of obtaining a good clustering result across two experiments. A "good result" was defined by a clustering accuracy of $r > 0.87$. The measure of clustering accuracy r was proposed by Huang (1998) and defined as:

$$r = \frac{1}{n}\sum_{l=1}^{k} a_l \tag{4}$$

In the above equation, a_l is defined as the number of data objects that occur in both a cluster and its labeling class, while n is the number of objects in the data set. Research demonstrates that it is possible to cluster and systematically organize categorical data, assuming there is an underlying pattern to the data.

Document Clustering

Text mining is a subset of data mining, with similar goals in mind. It is applied to large volumes of non-structured text files as opposed to numerical or interval data. It is used to discover knowledge from text data (Renz and Franke, 2003). There are two types of operations that text mining performs:

1. Categorization - assigning a given class to a text
2. Clustering - splitting a text collection into topics

Text mining as a discipline arose due to the complexity of human language and accompanying human desire to organize, store, and retrieve relevant information from text sources. Language serves the vital role of communication, but it is not easily analyzed by computers (Rend and Franke, 2003). This is due to the many functions language performs, from description to instruction to conveying of emotions. Language is associative and contains many contexts, which can change rapidly and drastically over the course of a conversation or narrative report. Deducing relevant facts and creating knowledge from these text reports is one goal of text mining. Clustering has been shown to aid information retrieval and summarization, as it facilitates locating relevant material much more quickly.

As stated previously, the sheer number of documents and information being produced poses a serious challenge for organization activities. Natural language has a high dimensionality value. Allah, Grosky, and Aboutajdine (2008) proposed a document clustering method based on diffusion maps and *k*-means clustering. Based on the underlying assumption that related documents – those documents belonging to the same cluster – are likely to share a common trait defined by a distance measure, the authors stated that the similarity distance measure used was more interesting than the representation space. Their process employed diffusion maps and singular value decomposition (SVD) to cluster documents. The main finding was that, if documents are properly defined by accurate and reliable distance measures, diffusion maps are highly effective, as well as computationally efficient, in clustering them.

CHAPTER THREE: METHOD

A review of the literature demonstrates that errors can be classified and appropriate counters implemented, thereby managing error. It further shows that errors are not random, and have associative properties when considered in environmental and situational contexts. It stands to reason that if certain error types exist, and these types can be linked to conditional factors, then a survey of anomalies can reveal insights or imply causal information through correlation, association, and regression analysis. These findings can lead to guidelines for error management and prevention through the investigation of said trends. Furthermore, clustering and grouping of the data can reveal additional insights on incident and factor relationships. A review of the literature thus demonstrates that human error is manageable, thereby implying relationship, predictability, and structure to the patterns of human error.

The method of investigation from the Hobbs & Williamson (2003) study forms the basis of investigation for the present study. The ASRS database identified factors that were reported along with each anomaly. These anomalies were also broken down into types. A categorical breakdown of these anomalies and their associated factors was first conducted.

Data Structures

Data was provided via an ASCII text dump of the Oracle database employed by ASRS. These data were extracted, merged, and categorized by type. All available data as of March 15, 2010 was used. The ASRS database classifies anomalies based on the following types and subsets:

Table 4: ASRS Anomalies and their subsets

ASRS Anomaly Types and Subsets	
Aircraft Equipment	Ground Excursion
Critical	Ramp
Less Severe	Runway
Airspace Violations	Taxiway
ATC Issues	Ground Incursion
Flight Deck/Cabin/Aircraft Event	Runway
Illness	Taxiway
Passenger Electronic Device	Ground Event/Encounter
Passenger Misconduct	Aircraft
Smoke/Fire/Fumes/Odor	FOD
Conflict	Gear Up Landing
NMAC	Ground Strike - Aircraft
Airborne Conflict	Loss of Aircraft Control
Ground Conflict	Object
Critical	Person/Animal/Bird
Less Severe	Vehicle
Deviation - Altitude	In-flight Event/Encounter
Deviation - Speed	CFTT/CFIT
Deviation - Track/Heading Deviation - Procedural	Fuel Issue
Clearance	Loss of Aircraft Control
FAR	Object
Hazardous Material Violation	Bird/Animal
Landing without Clearance	Unstabilized Approach
Maintenance	VFR in IMC (flight into low-visibility
MEL (Minimum Equipment List)	conditions without proper authorization)
Published Material/Policy	Wake Vortex Encounter
Security	Weather/Turbulence
Weight and Balance	No Specific Anomaly Occurred
	Other

Data Understanding

The data for this study arrived in three ASCII text files. The three files contained strings of column data separated by pipe (|) characters. The first file examined was the ALL_ITEMS.txt data file. It was converted to SPSS .sav format using the SPSS data input wizard via the GETTEXT command. The data organized into 5 columns. The columns comprised an ID field, a descriptive ENTITY field, an ATTRIBUTE field and two VALUE fields, a general value and a DISPLAY value field.

ITEM_ID	ENTITY	ENUMERATOR	ATTRIBUTE	VALUE	DISPLAY_VALUE
75085	Aircraft	1	Airspace.Class D	SAN	.
75085	Aircraft	1	ATC / Advisory.TRACON	SAN	.
75085	Aircraft	1	Flight Phase	Cruise	.
75085	Aircraft	1	Make Model Name	Small Aircraft, High Wing	.
75085	Aircraft	1	Crew Size.Number Of Crew	1	.
75085	Aircraft	1	Airspace.Class B	SAN	.
75085	Aircraft	1	Flight Plan	None	.
75085	Aircraft	1	Aircraft Operator	Personal	.
75085	Aircraft	1	ATC / Advisory.Tower	MYF	.
75085	Aircraft	1	ATC / Advisory.Tower	SAN	.
75085	Aircraft	1	Mission.Other	Utility	.
75085	Aircraft	1	Route In Use	Direct	.
75085	Aircraft	1	Flight Phase	Cruise	.
75085	Assessments	1	Primary Problem	Human Factors	.
75085	Environment	1	Light	Daylight	.

Figure 10: Data fields from the ALL_ITEMS file

The ENUMERATOR and DISPLAY_VALUE columns were not utilized in this study, as they were not populated with usable data. The ENUMERATOR field was populated with 1s, while the display value field was only populated when a measurement or instrument reading was associated with the record. The ENTITY field marked relevant records for extraction. The records employed in this study were those marked Anomaly.[Type], Assessments, Person, and Result.[Type] where [Type] indicated a subset element.

The second file of interest was the TEXT.txt file. As with the ALL_ITEMS.txt file, the data was encoded in ASCII text. Each record had an ID field, an ENTITY field, an ENUMERATOR, ATTRIBUTE, and TEXT field. The ENTITY and ENUMERATOR fields were not used, and relevant records were extracted using the ATTRIBUTE field, with only records marked "Narrative" being employed for the text mining part of the study.

ITEM_ID	ENTITY	ENUMERATOR	ATTRIBUTE	TEXT
75085	Text	1	Synopsis	ENTERED THE TCA WITHOUT CLRNC. OPERATION BANNER TOWING.
75085	Text	1	Narrative	APPROX 5 MI NW OF THE MISSION BAY VOR, WHILE AT AN ALT OF APPROX 500' OVER...
79866	Text	1	Synopsis	AFTER LEAVING FL320 MLG INSTRUCTED TO DES TO FL310 AFTER BEING INITIALLY C...
79866	Text	1	Narrative	WE WERE IN OUR CLB FROM CMH TO DFW, WITH A CLRNC TO FL310. PASSING FL280...
79912	Text	1	Synopsis	SMA PENETRATED TCA ON CLIMB OUT.
79912	Text	1	Narrative	THIS WAS MY FIRST DEP FROM BFI ON 31L. MY TURN TO THE SOUTH WAS PREMATUR...
79965	Text	1	Synopsis	SMT PLT DESCENDED TO ARPT UNDERLYING TCA, ACCUSED OF PENETRATING TCA.
79965	Text	1	Narrative	A VFR FLT, BEING CONDUCTED UNDER FAR PART 91, WAS APCHING DENVER CENTE...
79997	Text	1	Synopsis	ATCT LCL CTLR MIXES UP CALL SIGNS.
79997	Text	1	Narrative	SITUATION BEGAN ON A ROUTINE VFR PLEASURE FLT INTO LUK ARPT. ONBOARD ACF...
79998	Text	1	Synopsis	FMS STEERED ACFT SOME 20-25 MI OFF COURSE WITH NO WARNING TO FLT CREW.
79998	Text	1	Narrative	CLRD DIRECT PVT VOR AFTER TKOF BOS. USING R NAV FMS AND OMEGA BOS CENTE...
80003	Text	1	Synopsis	ACFT WAS SERVICED WITH JET A FUEL INSTEAD OF GASOLINE.

Figure 11: Fields from the TEXT.txt file

The nature of the ASRS database was found to be a list of records. Individual reports contained a unique identifier code, found in the ITEM_ID column. This number marked all elements associated with that record. Each record identifier contained multiple rows of data. To proceed, the relevant rows of data had to be separated from the master ALL_ITEMS file, and then recombined in a new file to create additional columns, ensuring that the ITEM_ID field is used as a key to maintain integrity of the record. In this way, a new file could be created that contains all columns deemed necessary for the analysis. The PASW Modeler 13 tool was used for this part of the process.

The ASRS database contained a data field named "Assessments," where contributing factors or situations to the anomaly were identified. The assessments field further identified elements as "Contributing Factors / Situations," with a special emphasis on one item if it was identified as a "Primary Problem." The items that were entered into Contributing Factors / Situations or Primary

Problem were identical save for an additional option for the Primary Problem category, called Ambiguous.

Table 5: ASRS Contributing factors entry types

ASRS Contributing Factors / Situations and Primary Problem Classifications	
Aircraft	Human Factors
Airport	Incorrect / Not Installed / Unavailable Part
Airspace Structure	Logbook Entry
ATC Equip / Nav Facility / Buildings	Manuals
Chart or Publication	MEL
Company Policy	Procedure
Equipment / Tooling	Staffing
Environment - Non Weather Related	Weather
	Ambiguous

Data Preparation

The text files were translated into two large SPSS .sav documents. These files were then transferred to the PASW Modeler 13 application for further manipulation operations. The creation of the analysis files required a multiple step process. The following operations were employed to create a combined analysis file with text data:

1. Extraction of all records with an "Anomaly.*" ATTRIBUTE value
2. Extraction of all records with an "Assessment" ENTITY value
3. Extraction of all records with an ENTITY value of "Person," and an ATTRIBUTE value of "Human Factors" for a separate analysis
4. Extraction of text records whose ATTRIBUTE value matches "Narrative"

5. Merging of "Anomaly," "Assessment," and "Narrative" files
6. Elimination of duplicate entries by assuring that only unique ITEM_ID fields exist (no duplicate ITEM_ID fields)

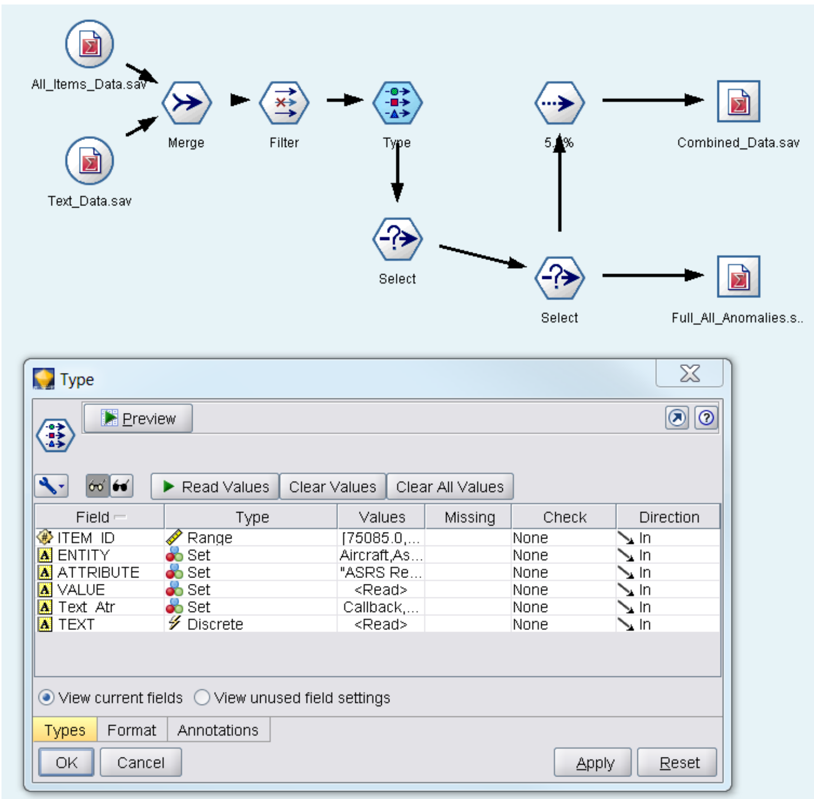


Figure 12: Data merging process (PASW Modeler 13)

After selecting and extracting each type of relevant record (Anomaly, Assessment, Result), the new data element was created as a new column in an output file. This process created duplicate entries during the merging process, as numbers of records in each file were not consistent. One file or the other would contribute more records, thus creating unnecessary duplicate row entries. These duplicates were eliminated by assuring that only one unique ITEM_ID existed per record. This eliminated duplicate narrative text fields, which, if left in the file, would drastically increase file sizes and computation times. The completed, pared down data file was now ready for analysis; it contained entries for anomalies, factors (assessments), and text data detailing the account in narrative form.

The other files containing results and human factors field data were set aside for a separate analysis. Not all records input into the analysis software contained these fields, as they were later added to the ASRS data collection process. For example, Human Factors data was not specifically encoded by the system until June of 2009 (Aviation Safety Reporting System, 2002).

Data Mining

The study employed data mining steps consistent with the CRISP-DM method. The six steps as they pertain to this activity are outlined below:

1. The business case for this activity was to identify what factors are present when pilots submit these anomaly reports. Anomalies were considered for this activity rather than accidents as accidents are rare when contrasted with these reports. Meaningful patterns can be discovered from these reports, rather than rare and non-routine accidents.
2. The ASCII data dump was analyzed and understood. It was decided that the "other" anomaly type did not contribute any useful information, and increased the overall variability and noise in the dataset. This category was discarded for analysis. The type "No Specific Anomaly Occurred" was kept as a report containing admission of

violation of procedure or other difficulty without incident, which could still contribute meaningful information.

3. Data was prepared through appropriate merging, discarding, and type classification activities. The data arrived in separate ASCII text files; these data had to be merged and organized. Report numbers identified individual report elements in rows. Anomalies were divided into their classification, then grouped by details. Factors were also grouped into sets, and these sets were divided into primary cause and secondary or contributing factors. The details of each anomaly were also grouped, and the text reports were kept organized by report number.
4. The modeling techniques performed included self-organizing maps, clustering analysis, and dimensionality reduction using diffusion maps.
5. The evaluation phase - Models were evaluated using goodness of fit analysis, tools internal to PASW Modeler 13, and clustering accuracy measures as proposed in Huang (1998).
6. Deployment phase – results were presented in graphical and tabular form displaying what factors were associated with civil aircraft anomaly reporting.

Structural Equation Modeling

The methods of structural equation modeling described in the literature could not be employed due to the categorical nature of the data. Many accident reports contain qualitative elements such as "bad weather," or "high workload conditions." These are represented as categories, and are either present or not. For SEM to be effective, interval data is required.

Logistic Regression

The data were entered into a multimodal logistic regression model to determine which pre-classified factors contributed most to selected anomaly types. The most prevalent frequencies of anomalies were used for this exercise. Anomalies associated with less than three percent of the data were not used in the analysis. SPSS release 17 was used for the analysis. Specifically, the multinomial logistic regression command employing the NOMREG algorithm calculated the results. This algorithm models illustrate the dependence of a nominal categorical response on a set of discrete predictor variables.

Logistic regression is a statistical method for predicting the probability of the occurrence of an event based on predictor variables (Hosmer, 2000). Logistic regression differs from traditional regression in that it can employ categorical as well as numerical data. Because of these qualities, it has found use in applications which rely on categorical values or dichotomous variables. The strength of the logistic function underlying logistic regression is that it can handle input values of any magnitude, while outputting values between 0 and 1. The logistic function is defined in the below equation, demonstrating how any input z will take the form of a range between 0 and 1. This attribute allows the equation to force outputs to be between 0 and 1.

$$f(z) = \frac{1}{1+e^{-z}}$$

(5)

Logistic regression can be used to create models suited for prediction, classification, and selection. Prediction models group cases according to whether they satisfy a given criterion. There are no constraints on the sizes of these groups; all cases may be predicted to be "positive" or "negative." This means it is possible for all cases to be predicted to belong to one group, i.e., the sample or population is homogenous. Predictive logistic regression is usually applied to those situations where identical treatment, based on the findings of all groups, is a possibility (Menard 1995).

Multinomial Logistic regression builds upon basic logistic models. When a dependent variable has M categories one value is designated as the reference category. The other categories are compared to reference category in terms of probability. For two categories the equation is:

$$\ln\frac{P(Yi = m)}{P(Yi = 1)} = \propto_m + \sum_{k=1}^{K} \beta_{mk} X_{ik} = Z_{mi}$$

(6)

When there are more than 2 groups the probability calculation becomes slightly more complicated as the $P(Y_i)$ term must be solved:

$$P(Yi = m) = \frac{\exp(Z_{mi})}{1 + \sum_{h=2}^{M} \exp(Z_{hi})}$$

(7)

The output is a probability of belonging to a given category. These probabilities are compared to the reference category chosen for the logistic regression analysis, and assignments are based on the comparisons of these probability calculations. These methods are described in more detail in Borooah (2001).

Correspondence Analysis

The statistical software package SPSS r17 was used to perform the dimensionality reduction process in the form of correspondence analysis. Data were coded in alphabetical order. These coded data were then analyzed using the CORRESPONDENCE command within SPSS to generate plots and cross tabulations.

Correspondence analysis operates optimally on categorical data, and is applied to data that is nonnegative and on the same scale. The data used in this study was categorical data coded into single digit numerals. The method operates similarly on rows and columns, and operates on contingency tables. Correspondence analysis decomposes the chi-square statistic based on this table into orthogonal factors. Correspondence analysis is a descriptive technique so it can be used whether it is correct to apply the chi-square statistic or not.

Table 6: Coding scheme for anomaly types

Old Value	New Value	Value Label
Anomaly.Aircraft Equipment Problem	1	Anomaly.Aircraft Equipment Problem
Anomaly.ATC Issue	2	Anomaly.ATC Issue
Anomaly.Conflict	3	Anomaly.Conflict
Anomaly.Deviation - Altitude	4	Anomaly.Deviation - Altitude
Anomaly.Deviation - Procedural	5	Anomaly.Deviation - Procedural
Anomaly.Deviation - Track /	6	Anomaly.Deviation - Track / Heading
Anomaly.Inflight Event / Encounter	7	Anomaly.Inflight Event / Encounter

Table 7: Coding scheme for factor types

Old Value	New Value	Value Label
Aircraft	1	Aircraft
Airport	2	Airport
Airspace Structure	3	Airspace Structure
Ambiguous	4	Ambiguous
ATC Equipment / Nav Facil	5	ATC Equipment / Nav Facil
Chart Or Publication	6	Chart Or Publication
Company Policy	7	Company Policy
Environment - Non Weather	8	Environment - Non Weather
Equipment / Tooling	9	Equipment / Tooling
Human Factors	10	Human Factors
Incorrect / Not Installed	11	Incorrect / Not Installed
Logbook Entry	12	Logbook Entry
Manuals	13	Manuals
MEL	14	MEL
Procedure	15	Procedure
Staffing	16	Staffing
Weather	17	Weather

Text Mining

Various tools were considered for the text mining part of the study. Many tools exist that create contextual links between words in a text source. Khalid et al (2010) employed Leximancer (LXM) and Latent Semantic Analysis (LSA). Leximancer is an online tool to which users can upload sets of documents for analysis. Both tools allow word counts to be performed, and identify concepts which can later be analyzed using statistical methods.

Due to its aviation roots and administrative desire for standardization and efficiency, ASRS data is encoded. For example, concepts such as "Air Carrier" or "Go Around" were encoded into "ACR" and "GAR." Currently available text mining tools do not have dictionaries capable of interpreting these terms. In addition, a "find/replace" operation was deemed infeasible due to the large number of records and the fact that there were over 700 terms to decode. Although the majority of encoded terms were abbreviations of aircraft components and aviation procedures, commonly used words such as "aircraft," "runway," and "landing" were also encoded. Because of these encodings, traditional automated semantic analysis and text meaning tagging operations were not possible using standard tools.

PASW Modeler 13 was chosen because it allows the user to create custom concepts and rules that do not rely on predefined dictionaries. It was therefore possible to extract concepts, though they had to be systematically defined and interpreted by a human analyst. Due to the nature of the data, complex text-linking and automated content-bearing extraction using existing text mining tools was not possible.

O	
OAT	Outside Air Temperature
OJT	On-The-Job Training
OM	Outer Marker
ONS	Omega Navigation System
OP(S)	Operation(S)
OPDEV	Operational Deviation
OPERROR	Operational Error
OTS	Out Of Service
OVCST	Overcast

Figure 13: Selected ASRS encoded terms and their meanings

To begin the text mining portion of the study, the completed data file was input into PASW Modeler's Text Analytics interface. This software tool analyzes all available text in the data and displays the most often encountered words. These often-encountered words were deemed "concepts." These concepts were filtered and grouped by type. For example, words such as "hyd" or "flt cntrls" were put into an "aircraft components" type. These types were used to identify and create rules for classifying text entries; the container element that contained these rules and types was called a "category."

Once types for selected popular concepts were defined, these types were used to build rules to automate record classification. These rules used logical operators on concepts to create relationships that selected and classified certain records. For example, records that contain concepts such as "hyd failure" or "smoke" or "burning smell" likely indicate aircraft issues or malfunctions. The selection and classification rules were created using keywords and data from the HFACS classification system (Wiegmann & Shappell, 2001), Boeing MEDA tool (Rankin & Allen, 1996), and general pilot knowledge keywords elicited from a flight training text (Willits et al, 2004).

Table 8: Categories, rules, and their associated types and descriptors

<table>
<tr><th>Categories</th><th>Rules</th><th></th></tr>
<tr><td></td><td>Descriptors</td><td>Type</td></tr>
<tr><td>Aircraft Issues</td><td>vibration</td><td>Unknown</td></tr>
<tr><td></td><td>low oil pressure</td><td>Rule</td></tr>
<tr><td></td><td>light</td><td>Unknown</td></tr>
<tr><td></td><td>leak</td><td>Rule</td></tr>
<tr><td></td><td>hyd failure | inspection | disconnect | valve | fire bottle | volts</td><td>Rule</td></tr>
<tr><td></td><td>acft & failure</td><td>Rule</td></tr>
<tr><td>Knowledge-Based Errors</td><td><Knowledge-Based></td><td>Knowledge-Based</td></tr>
<tr><td>Perceptual Errors</td><td>spatial disorientation | illusion | visual perception</td><td>Rule</td></tr>
<tr><td></td><td>not heard</td><td>Rule</td></tr>
<tr><td></td><td>misjudge & (distance | flare | altitude | visual | aural)</td><td>Rule</td></tr>
<tr><td></td><td>disoriented</td><td>Rule</td></tr>
<tr><td>Rule-Based Errors</td><td>wrong response | prefight | low fuel</td><td>Rule</td></tr>
<tr><td></td><td>retrospect</td><td>Unknown</td></tr>
<tr><td></td><td>exceeded ability</td><td>Rule</td></tr>
<tr><td></td><td><Rule-Based></td><td>Rule-Based</td></tr>
<tr><td></td><td>(misdiagnose | wrong | inappropriate | bad) & (emergency | maneuver | decision)</td><td>Rule</td></tr>
<tr><td></td><td>(bad | wrong) & (planning | pre-flight | fuel management | decision | go-around | go around)</td><td>Rule</td></tr>
<tr><td>Skill-Based Errors</td><td>stalled | spin</td><td>Rule</td></tr>
<tr><td></td><td>overcompensate | overspeed | wrong direction</td><td>Rule</td></tr>
<tr><td></td><td>inadvertently</td><td>Rule</td></tr>
<tr><td></td><td>forgot | pressed wrong | pushed wrong</td><td>Rule</td></tr>
<tr><td></td><td>did not notice | didn't notice</td><td>Rule</td></tr>
<tr><td></td><td>[bad & (technique | control | visual scan | conditions)]</td><td>Rule</td></tr>
<tr><td></td><td><Skill-Based></td><td>Skill-Based</td></tr>
<tr><td></td><td>(omitted & (step | checklist))</td><td>Rule</td></tr>
<tr><td>Unsafe Conditions</td><td>wake turb</td><td>Unknown</td></tr>
<tr><td></td><td>unsafe</td><td>Rule</td></tr>
<tr><td></td><td>tcasii & ra</td><td>Rule</td></tr>
<tr><td></td><td>self medicating</td><td>Rule</td></tr>
<tr><td></td><td>reaction time | visual limitation</td><td>Rule</td></tr>
<tr><td></td><td>not ready</td><td>Rule</td></tr>
</table>

<table>
<tr><td></td><td>misinterpreted | misinterpretation</td><td>Rule</td></tr>
<tr><td></td><td>incapable | inaptitude | insufficient aptitude</td><td>Rule</td></tr>
<tr><td></td><td>illness | incapacitated | fatigue | fatigued</td><td>Rule</td></tr>
<tr><td></td><td>excessive & (physical | training)</td><td>Rule</td></tr>
<tr><td></td><td>distr</td><td>Rule</td></tr>
<tr><td></td><td>conflict</td><td>Unknown</td></tr>
<tr><td></td><td>attention | complacency | distraction | distracted | bad motivation | overworked | situational awareness | mental fatigue | get home</td><td>Rule</td></tr>
<tr><td></td><td><Non-Routine Behaviors></td><td>Non-Routine Behaviors</td></tr>
<tr><td></td><td>(violation | violated) & (crew rest | rest | sleep | bottle | throttle)</td><td>Rule</td></tr>
<tr><td></td><td>(failed | did not | didn't) & (back up | communicate | brief | briefing | resources | leadership)</td><td>Rule</td></tr>
<tr><td>Unsafe Supervision</td><td>without clrnc</td><td>Rule</td></tr>
<tr><td></td><td>unqualified</td><td>Rule</td></tr>
<tr><td></td><td>oversight</td><td>Rule</td></tr>
<tr><td></td><td>no training</td><td>Rule</td></tr>
<tr><td></td><td>no clrnc</td><td>Rule</td></tr>
<tr><td></td><td>known problem</td><td>Rule</td></tr>
<tr><td></td><td>hurrying</td><td>Rule</td></tr>
<tr><td></td><td>failed to enforce | unqualified | unauthorized</td><td>Rule</td></tr>
<tr><td></td><td>authorized hazard</td><td>Rule</td></tr>
<tr><td></td><td>(supervisor | boss | chief) & (violation)</td><td>Rule</td></tr>
<tr><td>Violations</td><td>VFR into IMC | not current | not qualified | low</td><td>Rule</td></tr>
<tr><td></td><td>altitude flight | unauthorized | hazardous maneuver | aggressive maneuver</td><td>Rule</td></tr>
<tr><td></td><td>follow & (procedure | directive | instruction | command | briefing)</td><td>Rule</td></tr>
<tr><td></td><td>exceeded & limits
canyon | low altitude | overspeed</td><td>Rule</td></tr>
<tr><td></td><td><Violation></td><td>Violation</td></tr>
<tr><td>Weather</td><td>twilight</td><td>Rule</td></tr>
<tr><td></td><td>haze</td><td>Unknown</td></tr>
<tr><td></td><td><Weather></td><td>Weather</td></tr>
</table>

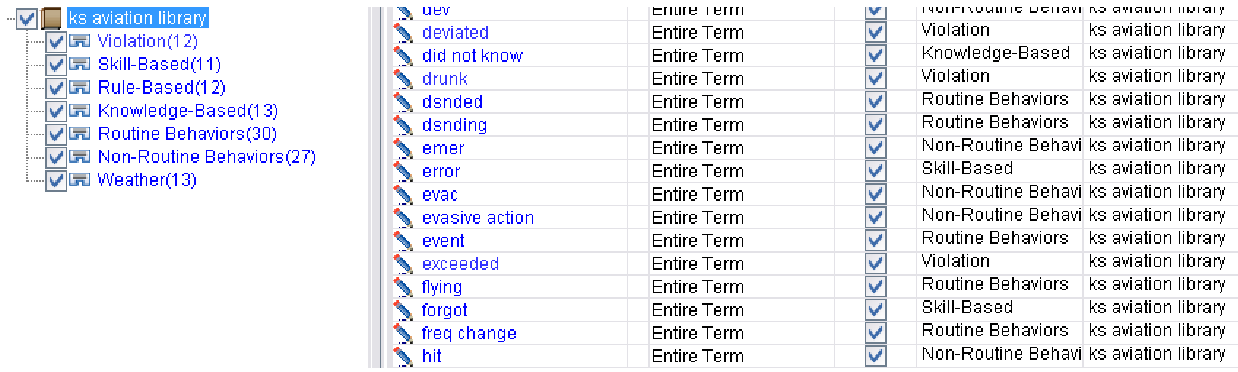


Figure 14: Selected types created from concepts

The above figure shows a selection of types created from concepts. Commonly occurring concepts were grouped into types to facilitate category creation. These types were organized into libraries which were stored and reusable across analyses. A full list of the concepts and types used is presented in the Appendix B.

Clustering and predictive model generation from text mining results

After text extraction and subsequent categorization based on results from the literature review, predictive models were generated using PASW Modeler 13. The text analytics element of PASW Modeler 13 "scored' each text record according to categories created. These categories became selector variables, employing a true or false flag to indicate whether an individual record contained elements of that category. These flags could now act as additional data points that could be used for clustering or predictive model generation activities.

Modeling was achieved by appropriately setting up the "input stream" within PASW 13 Modeler software application. Connections were made between the data source to the various selection, filtering, and typing nodes that prepped the data for analysis. Finally, appropriate modeling nodes that determined the classification and clustering methods used were chosen based on findings from the literature. For example, due to the categorical nature of the data, neural network learning and decision tree models were employed rather than K-means clustering. After the creation of the category flags, Kohonen, Two-Step, and K-means were employed as the "categorical-only" data limitation had been removed.

Dimensionality reduction with diffusion maps

Text data is inherently unstructured and contains data that, if plotted, is of high dimension. Each word in a text document, for example, could be considered a dimension. Thus a 100-word record has 100 dimensions with which to contend. To manage this, methods were modified from those used by Underhill (2007) and Martinez (2002). An unsupervised approach to dimensionality reduction was chosen, as the literature indicated that a need exists for unsupervised dimensionality reduction in text mining. Dimensionality reduction techniques have seen varied uses in clustering and categorizing data. For example, Higgs et al (2006) successfully employed dimensionality reduction through diffusion maps to classify brain images according to species of animal from which brain scans were taken.

The method employed is that described by Lafon and Lee (2006) where a diffusion distance between terms can be approximated to a level of precision given by δ by observing the first few $q(t)$ nontrivial eigenvalues in the following relation:

$$D_t^2(x,z) \cong \sum_{j=1}^{q(t)} \lambda_j^{2t} \left(\psi_j(x) - \psi_j(z)\right)^2$$

(8)

Lafon and Lee (2006) explain that the above relation can be interpreted as a Euclidean distance in the linear map $\mathbf{R}^{q(t)}$ if the right eigenvectors are selected with λ_j^t coordinates on the data. The following diffusion map:

$$\Psi_t : x \rightarrow \begin{pmatrix} \lambda_1^t \psi_1(x) \\ \lambda_2^t \psi_2(x) \\ \vdots \\ \lambda_{q(t)}^t \psi_{q(t)}(x) \end{pmatrix}$$

(9)

describes the relation below

$$D_t^2(x,z) \cong \sum_{j=1}^{q(t)} \lambda_j^{2t} \left(\psi_j(x) - \psi_j(z)\right)^2 = \|\Psi_t(x) - \Psi_t(z)\|^2$$

(10)

The dimensionality reduction and selection of the relevant eigenvectors is dictated by the fall-off of the eigenvalues and other factors described in more detail in Lafon and Lee (2006). The main idea of this process is that the distance measures between Anomaly records are preserved in the dimensionality reduction, these measures then afford classification by *k*-means clustering.

The Term-Frequency Inverse-Document Frequency equation specified by Underhill (2007) was employed to create the input data, and dimensionality reduction using diffusion maps was carried out on the resulting matrix. The computed document difference matrix was chosen as a measure of document dissimilarity. The underlying theory was that the measures of differences between the documents, when reduced, would suggest what level of dimensionality is required to categorize the documents. The below figure, modified from its original version found in Underhill (2007), describes the process.

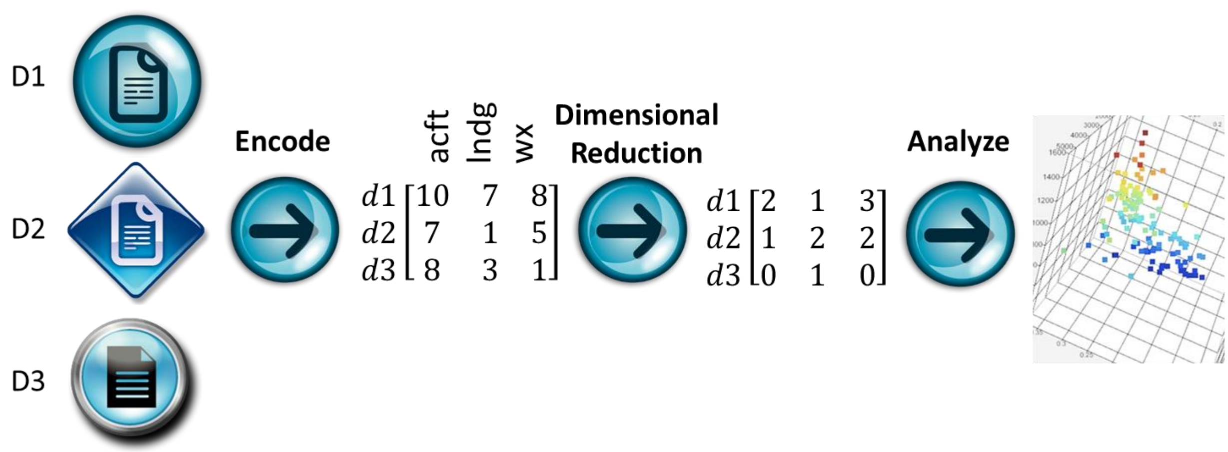


Figure 15: Dimensionality reduction approach (adopted from Underhill, 2007)

Diffusion maps aim to transform distance matrices that highlight local relationships between points (Coifman et al, 2005). These preserved relationships are based on the number of paths that exist between two data points; they describe how anomaly types are connected (Underhill, 2007).

PASW Modeler 13 was used to extract the most common words associated across all records with pilot anomalies. This action returned 127,766 records. Over 5,000 commonly-occurring words were extracted. The words were sorted in descending order by the number of documents that contained them. For example, the most commonly occurring word was “acft” (aircraft), and it was found in 41 percent, or 52,502, records. Due to memory and software constraints which will be described later, only 100 of these words could be used for dimensionality reduction. The 100 words chosen were the most frequently occurring words; these 100 words served as the input dataset for the clustering activity.

A Microsoft Excel document was created that listed these 100 words across the top row arranged by columns, each column containing a word. PASW Modeler 13’s category extraction feature was employed to create a sparsely populated term-document frequency matrix that indicated presence or absence as well as frequency of encounter of a given word within the anomaly record.

This matrix was then used to calculate a document feature. The method chosen was described by Underhill (2007), and is called the weighted term-document matrix. To create this matrix, a slightly modified version of the Term-Frequency Inverse-Document Frequency formula was employed:

$$\boldsymbol{T_{i,j} = (t/T) * ln(D/d)} \tag{11}$$

In this equation, t is the frequency value of a word j appears in document i. The sum total of all words of interest (row sums) that appeared in a given record was term T. The term D is the total number of documents (127,766), and d is the number of documents that contain the term j. This equation led to the creation of a term-document matrix, which then could be visualized and its dimensionality reduced in MATLAB.

MATLAB r2007b was used to carry out the dimensionality reduction, with the actual computation carried out using freely distributable example code developed by Ann B. Lee, Associate Professor within the Department of Statistics at Carnegie Mellon University at the time of writing of this work. The code was accessed from Professor Lee's personal webpage (Lee, 2010). The code was modified in MATLAB to accept 1,000 anomaly report records containing 100-item term-document matrix. A random sample of 1000 anomaly reports was selected. The below table displays a truncated, representative sample of the input data.

Table 9: Truncated input data for MATLAB

apch	rwy	flt	turn	twr	acft	clred
0	1	0	0	0	0	0
1	0	1	1	0	0	0
0	1	0	0	0	1	0
0	1	0	1	0	0	0
0	1	0	1	0	1	0
1	0	0	0	0	1	0
0	0	0	0	0	1	0
0	1	0	0	0	1	0
0	1	0	0	0	1	0
0	1	1	0	0	0	1
0	0	0	0	0	0	0
0	1	0	0	1	1	1
0	1	0	0	0	1	0
0	0	0	1	0	1	0
0	0	1	0	0	1	0

CHAPTER FOUR: RESULTS

Selection of Anomalies and Identification of Factors

The below tables show the total number of entries analyzed, broken down by Anomalies Reported and Total Contributing Factors Reported for those anomalies. These anomalies were selected due to their high occurrence rates; anomalies with less than three percent occurrence rates were not considered in the analysis. In addition, the "Other" anomaly type was omitted as they were records containing non-standardized, non-categorized entries, and did not contribute any meaningful information to this analysis. Furthermore, this was a "write-in" field, meaning that respondents could input any value, thereby making it highly variable and subjective. Data in this field varied widely. It was difficult to interpret these data compared to the ordered sets of data present in the other anomaly types.

Table 10: Anomaly types and frequencies selected for analysis

Anomaly Type	Frequency	Percent
Deviation - Procedural	56087	43.9
Aircraft Equipment Problem	27359	21.4
Conflict	16537	12.9
Inflight Event / Encounter	9676	7.6
Deviation - Altitude	8434	6.6
Deviation - Track / Heading	5294	4.1
ATC Issue	4384	3.4
Total	127771	100.0

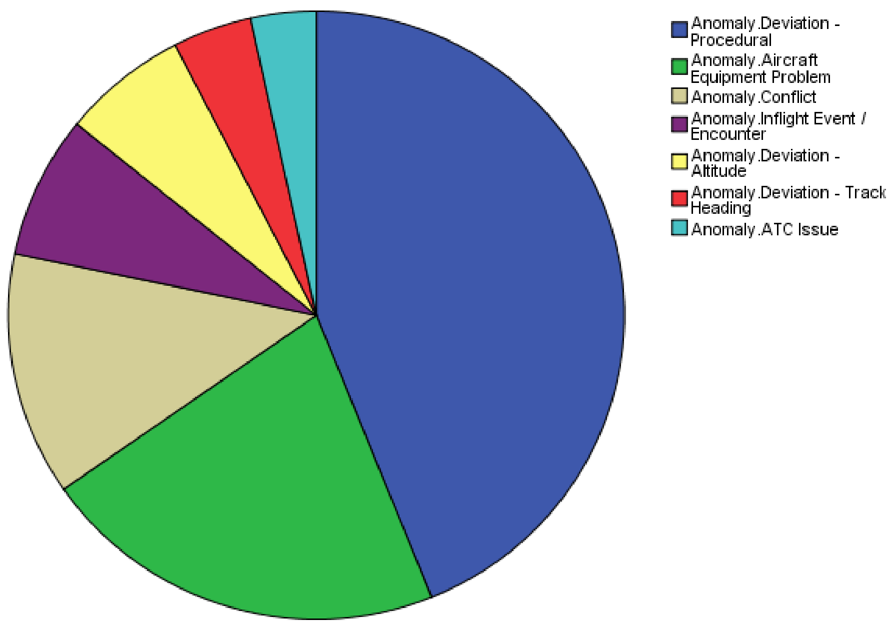


Figure 16: Distribution of anomalies by percentage

Factors contributing to anomalies are shown below. The factors were associated with anomalies; all were considered for analysis to maintain data consistency. The total number of records analyzed was 127,771.

Table 11: Factors associated with anomalies

Factor Type	Frequency	Percent
Human Factors	72607	56.8
Aircraft	27674	21.7
Weather	4818	3.8
Company Policy	4715	3.7
Airport	3757	2.9
Ambiguous	3112	2.4
Chart Or Publication	2484	1.9
Procedure	2246	1.8
Environment - Non Weather	2046	1.6
Airspace Structure	1902	1.5
ATC Equipment / Nav Facil	992	.5
Logbook Entry	608	.5
Incorrect / Not Installed	370	.3
Manuals	325	.3
Staffing	54	.0
Equipment / Tooling	33	.0
MEL	28	.0
Total	127771	100.0

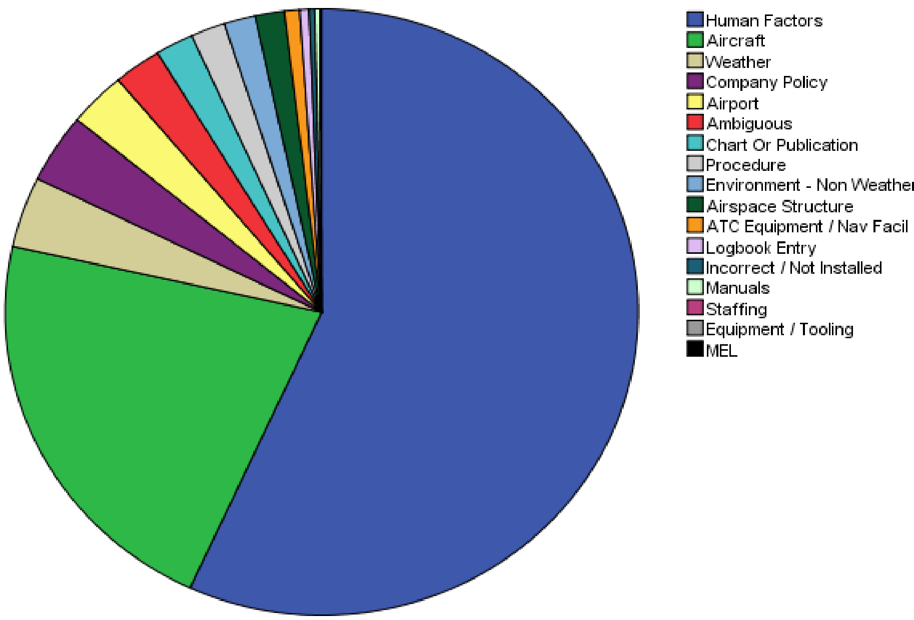


Figure 17: Distribution of factors associated with anomalies

Value	Proportion	%	Count
Anomaly.Deviation - Procedural		43.9	56087
Anomaly.Aircraft Equipment Problem		21.41	27359
Anomaly.Conflict		12.94	16537
Anomaly.Inflight Event / Encounter		7.57	9676
Anomaly.Deviation - Altitude		6.6	8434
Anomaly.Deviation - Track / Heading		4.14	5294
Anomaly.ATC Issue		3.43	4384

Figure 18: Distribution of anomaly types

Value	Proportion	%	Count
Human Factors		56.83	72607
Aircraft		21.66	27674
Weather		3.77	4818
Company Policy		3.69	4715
Airport		2.94	3757
Ambiguous		2.44	3112
Chart Or Publication		1.94	2484
Procedure		1.76	2246
Environment - Non Weather		1.6	2046
Airspace Structure		1.49	1902
ATC Equipment / Nav Facil		0.78	992
Logbook Entry		0.48	608
Incorrect / Not Installed		0.29	370
Manuals		0.25	325
Staffing		0.04	54
Equipment / Tooling		0.03	33
MEL		0.02	28

Figure 19: Distribution of factor types

Consistent with literature findings, the most often cited and classified factor for most anomalies is the human element. The next most popular factor involved aircraft equipment issues. Weather and Company Policies were the next most popular factors underlying anomalies. The least frequently encountered factors were MEL issues, which are issues associated with the Minimum Equipment List. The MEL comprises a list of components deemed absolutely necessary for dispatch, and is created and maintained by the aircraft manufacturer.

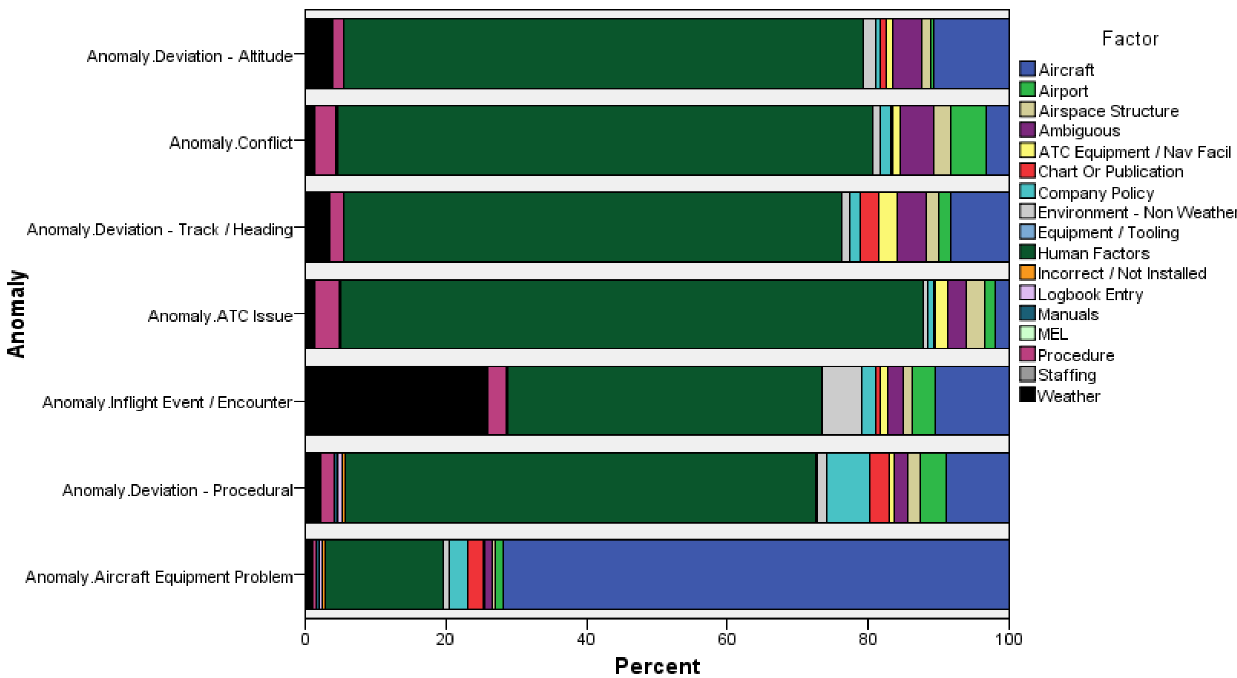


Figure 20: Aircraft anomalies broken down by factor type

The above figure is a visualization of the prevalent factors contributing to each anomaly type. Human Factors is most commonly associated across all anomaly types save for Aircraft Equipment anomalies. One other notable association is that Weather is often associated with Inflight Event / Encounter anomalies.

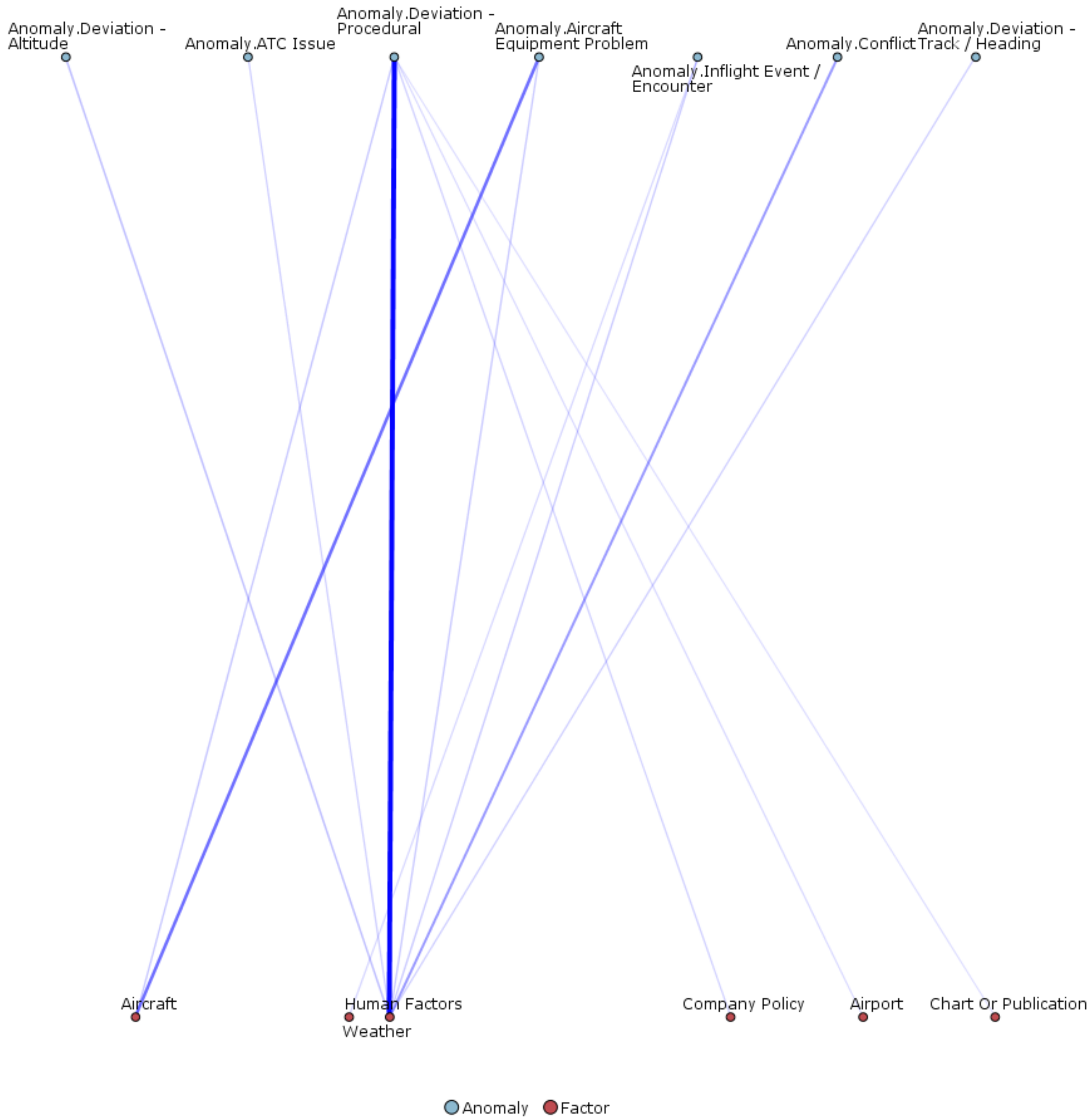


Figure 21: Directed web of relationships between factors and anomalies

The above figure depicts relationship strengths between factors and anomalies. Bolder lines denote stronger associations. These associations were determined by frequencies: the more often a record contained a given factor in an anomaly record, the bolder the line between the two. From the above graph, it is apparent that Human Factors are strongly associated with procedural deviations. Aircraft issues are once again strongly associated with Aircraft Equipment anomalies.

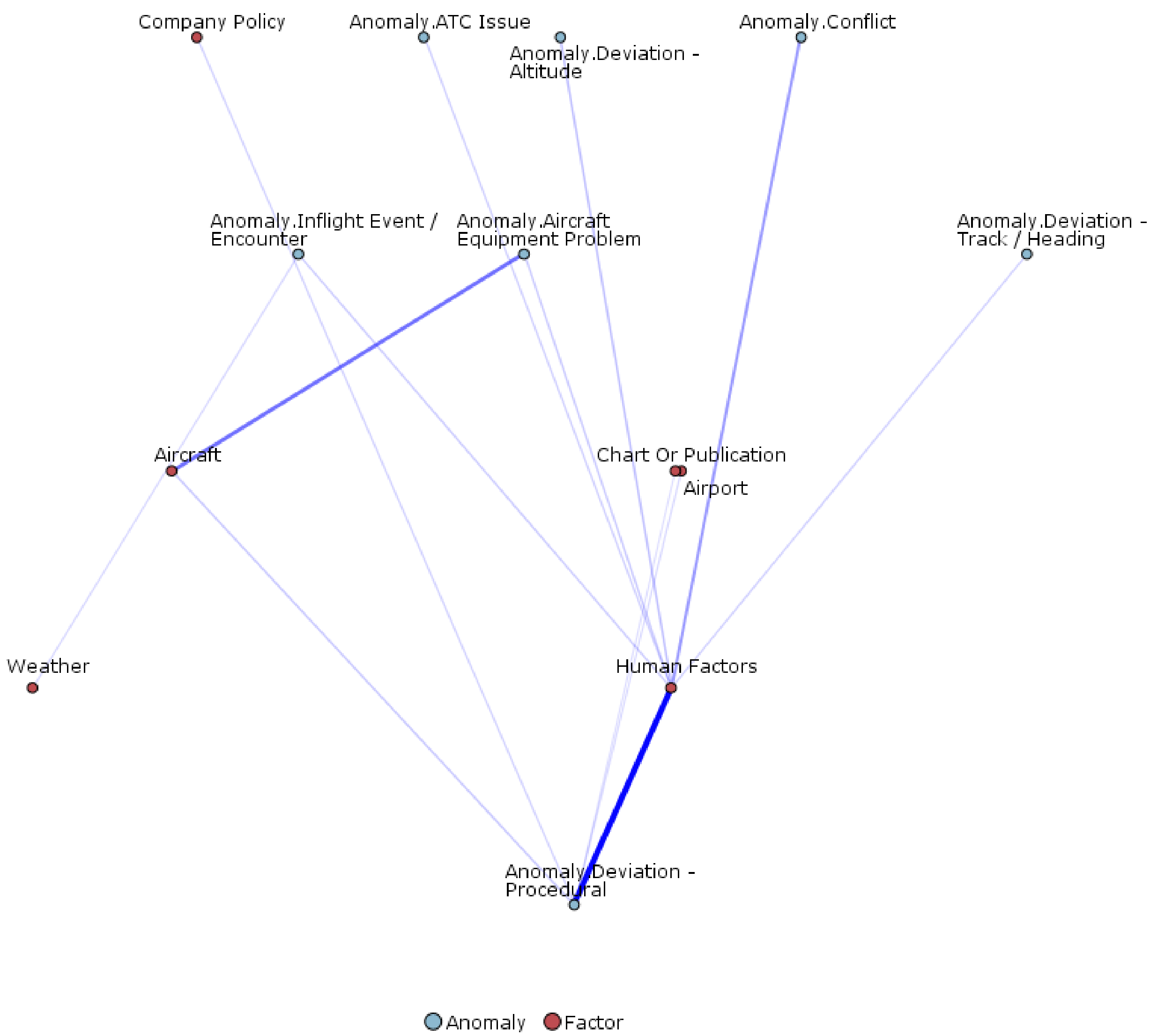


Figure 22: Network web of anomalies and factors

The network web provides another visualization of factor and anomaly associations. In the above figure, several more relationships are easier to see. In addition, partitions between factor types are evident. For example, Weather can contribute to Inflight Events/Encounters, but Company Policy is hardly ever a factor for Inflight Events/Encounters.

Discussion of Statistical Methods Results

This part of the study investigated the pre-determined categories and labels created and maintained by the authors of the ASRS database (Aviation Safety Reporting System, 2002). The anomalies will each be discussed in turn. This study found that the most prevalent factor for all anomalies is Human Factors: over 56 percent of anomalies have Human Factors listed as their principle cause. Dekker (2005) points out that, when an accident investigation fails to find mechanical failures, the investigation inevitably concludes that the problem is human in nature. Based on this evidence, Dekker constructed an equation demonstrating the ratio of accident causes to human and mechanical problems:

$$\text{human error} = f(1 - \text{mechanical failure})$$

This sort of anecdotal evidence is shared across domains where humans must interact with machines. Investigators, mechanics, troubleshooters, and support representatives all are quick to blame the human user for issues if no hardware problems are found.

In the Human Factors data subset, an unexpectedly high number of records indicated that the most frequent issue was Troubleshooting Aircraft Equipment. This is mildly surprising, as Aircraft Equipment anomalies were not the most prevalent anomalies encountered. Time Pressure accounted for the next most frequently encountered Human Factors issue. Decreased Time available to make decisions and enact plans commonly led to mistakes and higher workloads. Communication Breakdowns and losses of Situational Awareness (SA) were related, as they both corresponded to appropriate information flow, integration, and processing. Workload issues were infrequently encountered – another surprising find, given the anecdotal evidence of pilots complaining they were unable to "stay ahead of the airplane" during stressful situations. It is possible that workload issues were encompassed by the Time Pressure Human Factors category. Another notable Human Factors issue was that Fatigue was the least often encountered factor. Fatigue is often synonymous with Human Factors, as fatigue is often cited as a cause for decreased performance levels. These findings from the smaller Human Factors category are only representative of that reduced dataset, and although there are no reasons suggesting that they cannot be generalized to the entire dataset, no validation of this claim was performed.

Aircraft Equipment Problems

Aircraft issues were cited as the primary cause of Aircraft Equipment Problem anomalies in more than 70 percent of anomalies reported. This was an unsurprising finding, as by its very definition an Aircraft Equipment Problem is likely caused by faulty or incorrectly installed or maintained equipment. Human Factors was the second most associated factor. An example of this sort of incident, and one in which an Aircraft Equipment Problem was illustrated, was found in this excerpt of record 80386:

> …Became colder but nowhere near debilitating. Nevertheless, definitely a preoccupation. Thought about advising ctlrs at slc or den center about situation but didn't think it would cause a prob. Clred to descended and then to byson intxn, i thought to cross byson at or above 16000'. I'm fairly certain that i read it back that way. After byson, i was handed off to the next ctlr who said that the previous one wanted to know why i had not crossed at 16000'. I told him that i thought i was to cross at or above. Then, apching the arpt, i noticed that the lights on my avionics were off. Because of the cold i had put on a sport jacket and then a massive, bulky overcoat. A buckle on the coat had snagged the avionics master switch and turned the avionics off…

The above report, though classified as an Aircraft Equipment Problem, was caused by the pilot being too cold and donning a coat that became snagged on a control switch. The pilot did not notice that the avionics package which broadcasted important aircraft data had been turned off until notified by an air traffic controller. It is important to realize that certain seemingly simple anomalies may have unorthodox causes.

Altitude Deviations

The most frequently associated factor for Altitude Deviations was Human Factors. The overwhelming majority of Altitude Deviations, at more than 70 percent, was due to Human Factors issues. The most common reasons for these deviations were breakdowns in communication with air traffic controllers (ATC). Many reports claimed ATC issued requests that were unable to be met, or there were misunderstandings, and these misunderstandings persisted because the read-back to the controllers was not closely scrutinized. The literature supports this finding: modern aviation systems are highly connected information flow systems. If the flow of information is interrupted or corrupted, long-reaching, cascading effects can manifest themselves (Vidulich et al, 2010).

The second most frequently cited factor was Aircraft issues. Problems with aircraft equipment can either indicate improper altitude or provide an indication of another condition that could lead to an altitude deviation. For example, report number 779973 describes a situation where a pilot received a climb indication from a faulty sensor onboard the aircraft. The autopilot was on, and so the aircraft attempted to descend to compensate for this sudden climb indication. The other factor types were too weakly associated to be of any explanatory value.

ATC Issues

Human Factors were overwhelmingly associated with ATC Issues, comprising more than 82 percent of the anomaly reports. Most of these problems cited inattention or incomplete information being relayed between controllers and pilots or ground crew. Specific associated factors were communications breakdowns, delays in communication, or no communication at all.

Report number 901082 illustrates an interesting example of an ATC issue. The pilot landed the airplane but stated that the landing was anomalous. Upon further investigation, the tower's weather equipment was malfunctioning, causing the wrong runway to be assigned by ATC. ASRS classified this incident as "Human Factors," despite a definite problem with equipment. A miss-classified record such as this demonstrates the need for powerful and accurate classification algorithms and methods.

> …I was advised to report a 3 mile final for Runway 22. I requested a wind check. The report from the Tower was "Zero Seven Zero at ..., Gusting 14". I don't recall the exact wind (I believe it was 9 MPH)… …On final approach, the airplane did not "settle" correctly and was very "squirrelly" on short final. I noted that even though my airspeed was 70 KTS, the ground was moving way quicker than it should have been for the given winds from ATC… …At this point, a pilot on the ground radioed the tower and said that the windsock was definitely from the West. At that point, the Tower explained that the wind gauge equipment was up for maintenance the day before and that there was some sort of problem...

Conflicts

Conflict anomalies overwhelmingly had Human Factors cited as a primary contributor. Most conflicts occurred due to contributing factors like communication breakdowns or loss of situational awareness. The only other notable factor to contribute to Conflict anomalies was Weather. Poor visibility and deteriorating conditions leading to deviations or forced landings and inadequate separation all contribute to Conflicts. Text reports from records matching these conditions support these findings.

Inflight Events/Encounters

Most Inflight Events/Encounters were caused by Human Factors issues (67 percent), but a strong contributor was Weather (26 percent of anomaly reports). These findings are also consistent with literature reports. Limited or divided attention to the environment, both internal and external to the airplane, could lead to situations that cascade out of control. One situation analogous to certain concepts found in the literature is that selective attention in humans operates in a serial fashion, whereas external events are parallel. The spreading of responsibility to controllers and other crew members, automation, and integration of displays are all methods that can be employed to manage these types of situations (Vidulich et al, 2010).

Procedural Deviations

The frequency analysis revealed that the most prevalent anomaly type encountered was Procedural Deviation. This is consistent with many findings in the literature, which cite deviations from protocol, missed steps in a checklist, (Dekker, 2005) or other breakdowns in human resource management.

Track/Heading Deviations

The Deviations-type anomalies all had an overwhelming percentage of Human Factors issues listed as the primary contributor. Track/Heading Deviations ascribed 70 percent of records to Human Factors issues. Most of these issues comprised divided attention, fatigue, breakdowns in communication with ATC, and loss of SA. These findings again coincide with those of Durso and Alexander (2010), who began their paper with a Track/Heading deviation fictional scenario to illustrate how SA can be lost during both high and low workload conditions.

The only other notable contributor to anomalies concerning Track/Heading is that of Aircraft issues (8 percent). Similar to the Altitude Deviation anomaly, malfunctioning, inoperative, or incorrectly indicating equipment can very easily lead to an Altitude/Track Deviation, especially in inclement weather conditions.

Identification of Factors and Anomalies in the Human Factors Subset

The ASRS created categories and began tracking Human Factors data in June of 2009. This reduced dataset containing Human Factors category data contains 8,817 records. These records were analyzed separately to glean additional insights not found in the larger dataset.

Table 12: Human factors categories frequencies

	Frequency	Percent
Troubleshooting	3459	39.2
Time Pressure	2487	28.2
Communication Breakdown	760	8.6
Situational Awareness	730	8.3
Workload	386	4.4
Training / Qualification	350	4.0
Other / Unknown	207	2.3
Human-Machine Interface	190	2.2
Confusion	124	1.4
Distraction	70	.8
Physiological - Other	32	.4
Fatigue	22	.2
Total	8817	100.0

The above categories show that trouble shooting aircraft equipment and time pressure are the most common contributors to aircraft anomalies for the study period. Surprisingly, fatigue, distraction, and confusion are not categorized particularly prominently as contributors to aircraft anomalies.

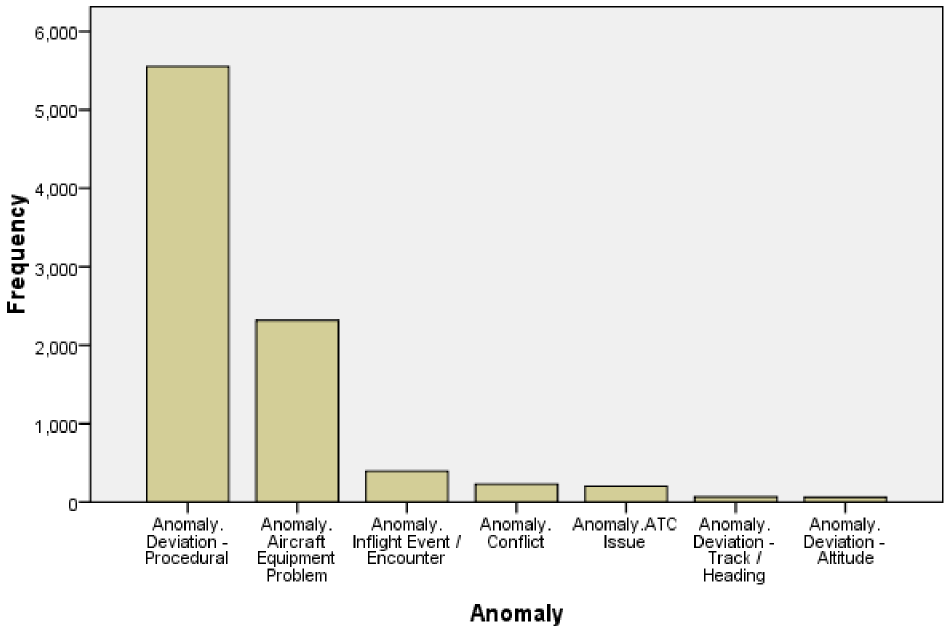


Figure 23: Frequencies of anomaly types in the human factors dataset

The above figure shows the frequencies of anomalies in the reduced Human Factors dataset. The most prominent anomaly types found in the Human Factors dataset were Procedural Deviations and Aircraft Equipment Problems. In light of these frequencies, the high counts of troubleshooting and time pressure as human factors contributors to anomalies appear to follow logically.

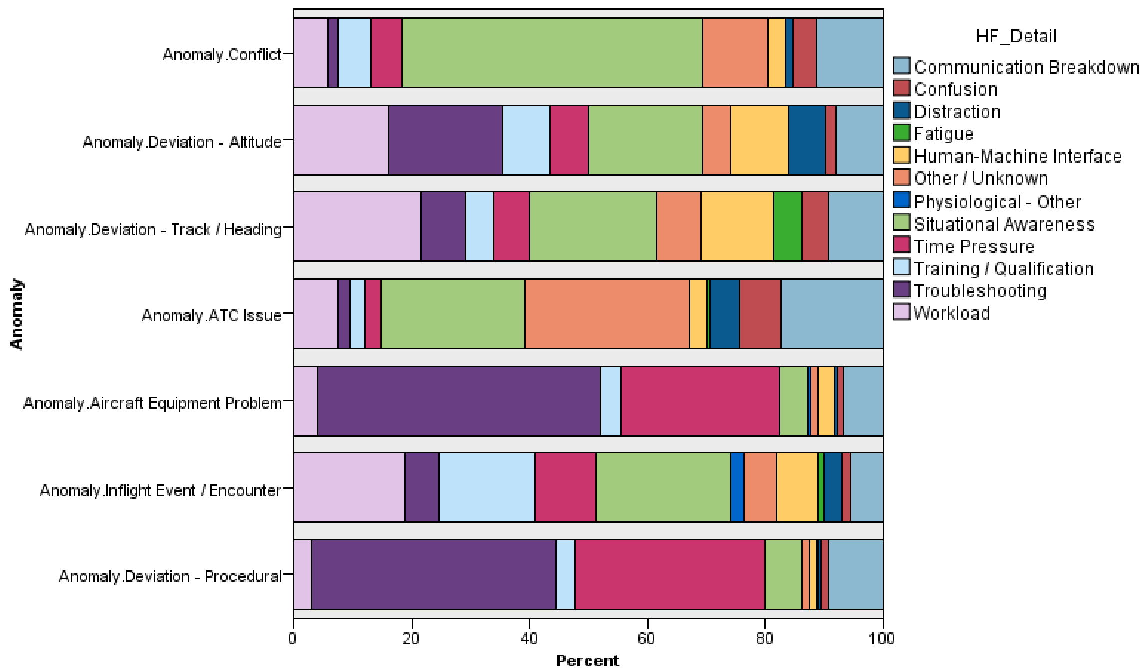


Figure 24: Normalized frequencies of anomalies and human factors contributors

The above frequency plot shows which human factors issues contributed to anomalies. Troubleshooting and Time Pressure were often associated with Aircraft Equipment Problems and Procedural Deviations. The normalized plot reveals that Situational Awareness was often a contributor to airborne Conflict anomalies, an association that the standard frequency counts do not readily reveal. Human-Machine Interface issues, which constitute a wrong button press or incorrect control operation, were also present across all anomaly types. The category Other / Unknown did not contribute any meaningful knowledge to the study.

The reduced dataset containing roughly 9,000 anomaly reports provided limited insights on what specific Human Factors issues exist that contribute to these anomaly occurrences. The most frequently encountered Human Factors issue was Troubleshooting. Troubleshooting involves conditions where aircraft equipment is malfunctioning, and the flight crew must take time away from managing flight to ascertaining what the issue is. Most anomaly reports involved the pilots troubleshooting the automation. The vast majority of problems were problems with the autopilot.

Most records had more than one Human Factors issue associated with them. Common co-associations were Time Pressure and Troubleshooting, Situational Awareness and Time Pressure, and Human-Machine Interface and Troubleshooting. Because of the way in which ASRS maintains records, those anomaly reports that do contain Human Factors data contain multiple categorizations; no one factor is singled out as the primary issue. Unfortunately, the ASRS database only recently began categorizing Human Factors data. Human behaviors are highly complex; placing any one issue into a category is very difficult. This was the primary reason why this study did not limit itself to human factors-related anomalies.

Relating Factors to Anomalies Trough Cross-tabulation

			Anomalies							
			Aircraft Equipment Problem	Altitude Deviation	ATC Issue	Conflict	Inflight Event / Encounter	Procedural Deviation	Track / Heading Deviation	Total
Factors	Aircraft	Count	19663	905	90	561	1021	4989	445	27674
		% of Total	15.4%	.7%	.1%	.4%	.8%	3.9%	.3%	21.7%
	Airport	Count	336	34	64	816	313	2113	81	3757
		% of Total	.3%	.0%	.1%	.6%	.2%	1.7%	.1%	2.9%
	Airspace Structure	Count	79	112	110	399	114	989	99	1902
		% of Total	.1%	.1%	.1%	.3%	.1%	.8%	.1%	1.5%
	Ambiguous	Count	304	351	122	797	226	1098	214	3112
		% of Total	.2%	.3%	.1%	.6%	.2%	.9%	.2%	2.4%
	ATC Equipment / Nav Facil	Count	71	61	69	159	94	394	144	992
		% of Total	.1%	.0%	.1%	.1%	.1%	.3%	.1%	.8%
	Chart Or Publication	Count	563	83	12	44	58	1586	138	2484
		% of Total	.4%	.1%	.0%	.0%	.0%	1.2%	.1%	1.9%
	Company Policy	Count	746	61	36	253	189	3355	75	4715
		% of Total	.6%	.0%	.0%	.2%	.1%	2.6%	.1%	3.7%
	Environment - Non Weather	Count	259	131	32	184	565	810	65	2046
		% of Total	.2%	.1%	.0%	.1%	.4%	.6%	.1%	1.6%
	Equipment / Tooling	Count	5	0	2	1	1	24	0	33
		% of Total	.0%	.0%	.0%	.0%	.0%	.0%	.0%	.0%
	Human Factors	Count	4559	6232	3627	12585	4321	37543	3740	72607
		% of Total	3.6%	4.9%	2.8%	9.8%	3.4%	29.4%	2.9%	56.8%
	Incorrect / Not Installed	Count	108	8	6	31	9	205	3	370
		% of Total	.1%	.0%	.0%	.0%	.0%	.2%	.0%	.3%
	Logbook Entry	Count	148	0	0	0	1	458	1	608
		% of Total	.1%	.0%	.0%	.0%	.0%	.4%	.0%	.5%
	Manuals	Count	85	1	2	0	7	230	0	325
		% of Total	.1%	.0%	.0%	.0%	.0%	.2%	.0%	.3%
	MEL	Count	8	0	0	0	0	20	0	28
		% of Total	.0%	.0%	.0%	.0%	.0%	.0%	.0%	.0%
	Procedure	Count	112	126	158	477	242	1031	100	2246
		% of Total	.1%	.1%	.1%	.4%	.2%	.8%	.1%	1.8%
	Staffing	Count	3	0	5	7	5	32	2	54
		% of Total	.0%	.0%	.0%	.0%	.0%	.0%	.0%	.0%
	Weather	Count	310	329	49	223	2510	1210	187	4818
		% of Total	.2%	.3%	.0%	.2%	2.0%	.9%	.1%	3.8%
Total		Count	27359	8434	4384	16537	9676	56087	5294	127771
		% of Total	21.4%	6.6%	3.4%	12.9%	7.6%	43.9%	4.1%	100.0%

Figure 25: Results of cross tabulations in factor and anomaly frequencies

The above figure displays the results of the cross tabulation activity. The column totals reveal which factors are most prevalent in each anomaly category. Notable frequencies in the Human Factors row (which makes up the majority of the counts) are Airborne Conflicts and Altitude Deviations. The chart also shows that most factor categories aside from the Human Factors row account for less than 10 percent of anomalies. The last, somewhat trivial, relationship is that Aircraft Issues account for 15 percent of Aircraft Equipment Problems.

Cross tabulations results did not contribute any additional findings to those identified by the statistical frequency analysis. The main findings were that Human Factors issues are present in most anomaly reports. Human Factors are cited as the primary cause more than half of the time, in five out of the seven anomalies investigated. Of these two anomalies, only Aircraft Equipment Problem had another factor more strongly associated with it than Human Factors.

This finding reveals that either human error or decision making is truly to blame for most aircraft anomalies, or else database administrators and those completing the reports are quick to assign blame to the human for problems. One interesting aside to accident investigation might be to ask how often human contributions "saved" the situation or prevented a serious incident. A notable example of this is US Airways Flight 1549, which made a water landing in the Hudson River in New York City after suffering a catastrophic engine failure due to multiple bird ingestion. In that situation, the pilot made the "correct" decision to attempt a water landing rather than try to divert to another airfield without any engine power. The pilot came to this decision and executed the landing while ATC was struggling to catch up and clear the airspace (Eisen et al, 2009).

Relationships between Factors and Anomalies

Consistent with the methods defined by Clausen (1998) and applied by Hobbs & Williamson (2003), correspondence analysis was carried out using SPSS r17. The categories shown below are representations of the chi-square distances between the categories. Those categories appearing closer together on the plot are more closely associated than those further apart.

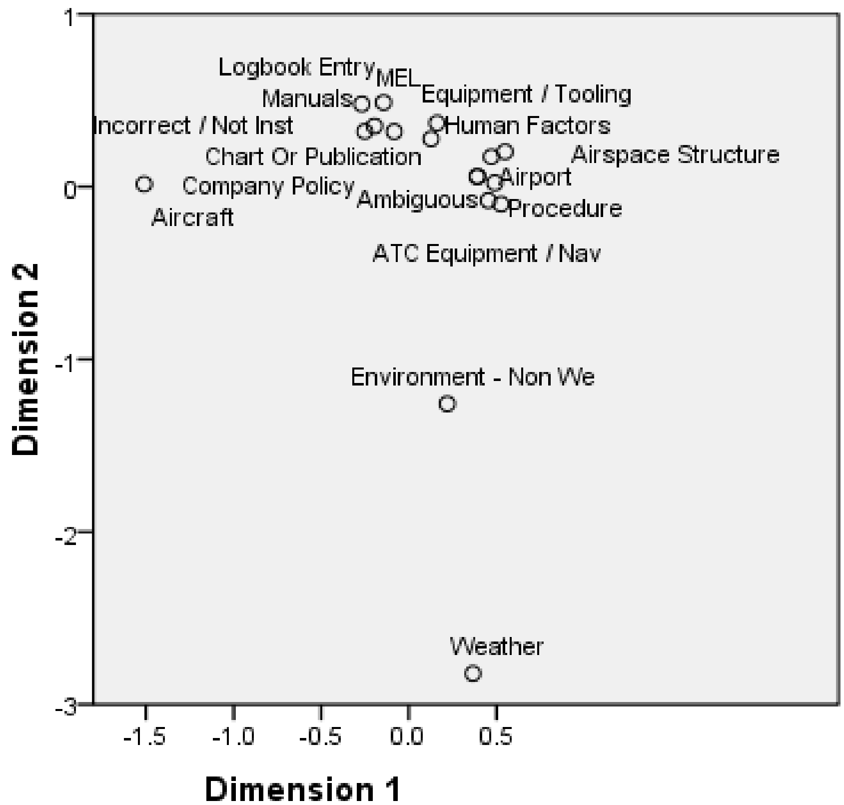


Figure 26: Correspondence analysis plot for factors

The above correspondence plot shows how closely associated individual factors are based on their chi-square distance. Non-Weather related Environmental issues and Weather were most disassociated from the other factor types. Procedural factors, Human Factors issues, Procedural issues, ATC Equipment and Navigational Facility issues, and Airspace Structure all were highly associated. Logbook Entry problems, unclear Manuals, issues pertaining Master Equipment List, Equipment/Tooling problems, and Installation problems were all closely associated, as these are maintenance-related factors. Aircraft issues were slightly disassociated from the maintenance factors group, but still showed signs of association with these factors.

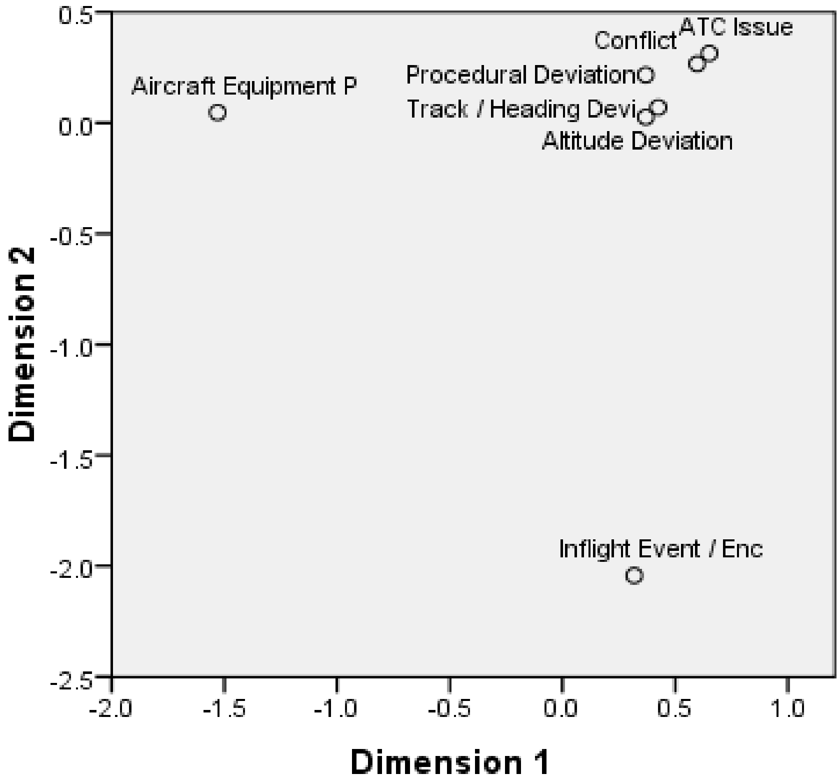


Figure 27: Correspondence analysis plot for anomalies

The above correspondence plot shows a high level of association between the inflight-related events ATC issues, Procedural Deviations, Track/Heading Deviations, and Altitude Deviations. Further disassociated from these are Aircraft Equipment Problems, and Inflight Events/Encounters.

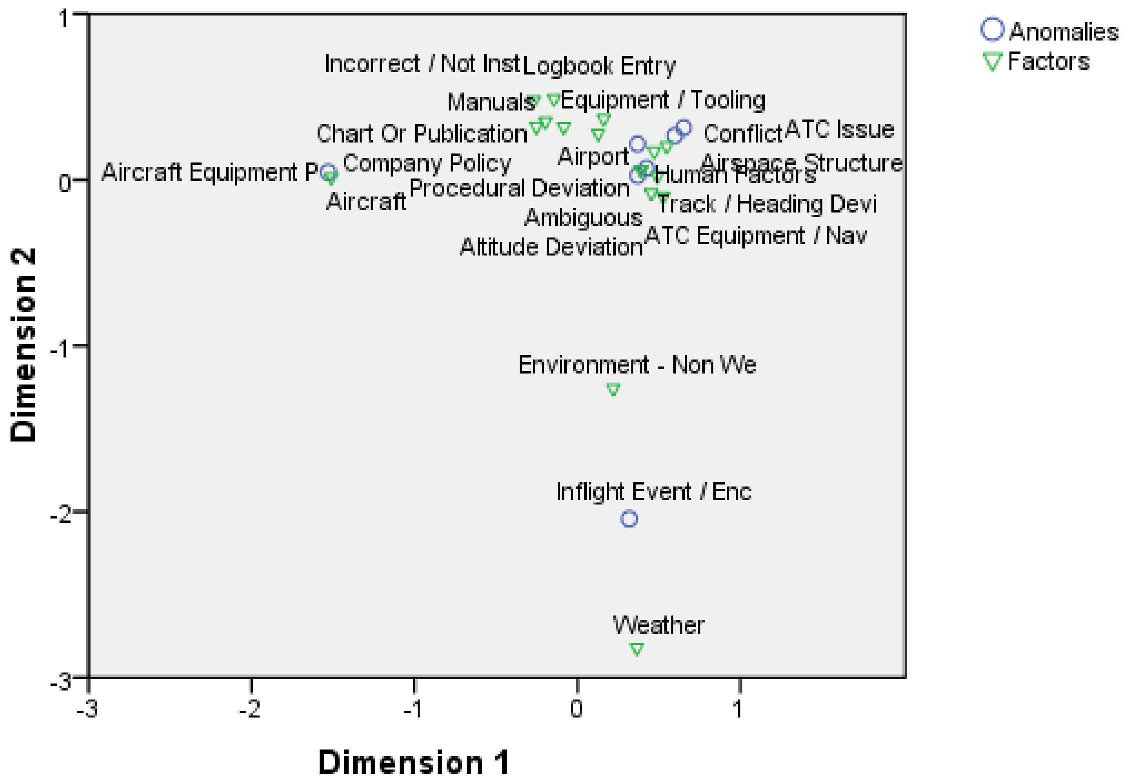


Figure 28: Correspondence analysis plot of factors and anomalies

The above figure shows the anomalies and factors plots superimposed on each other. It demonstrates closely the associations between Aircraft issues and Aircraft Equipment anomalies. Next, Chart or Publication issues, problems with Manuals, Airport issues, Logbook Entry errors, and discrepancies with Company Policies are all closely associated with the three Deviation anomalies (Track/Heading, Altitude, and Procedural). Human Factors as a factor is very closely associated with all anomalies save for Aircraft Equipment problems and Inflight Events/Encounters. The disassociated factors are Non-Weather related Environmental issues, and Weather issues. Inflight Events/Encounters is associated between the two Environmental factors.

The value of the findings of correspondence analysis was not limited to identifying known associations between anomalies and factors, but also described a powerful visualization of correlations and associations. Concepts that were intuitively related and supported by data such as Aircraft Equipment Problems and Aircraft issues were displayed in close proximity to each other, as might have been expected. All of the human decision and procedural anomalies and factors also clustered together. Seemingly random events beyond the control of the flight crew, such as Weather issues or Non-Weather Environmental problems were distanced from the other related anomaly and factor types. This method has already seen successful use in linking contributing factors to maintenance errors in Hobbs & Williamson (2003). The results here further demonstrate its usefulness in visualizing categorical data in order to make inferences about associations and possible causes.

Predicting Anomalies from Factor Data

The anomalies and factor classifications were placed into a multinomial logistic regression algorithm in an attempt to create a predictive model capable of classifying anomalies based on contributing factors information. Standard binomial logistic regression could not be employed because the dependent variable being predicted had more than two values.

Table 13: Multinomial logistic regression results

Factor	Anomaly	Frequency Observed	Percentage Predicted
Aircraft	Aircraft Equipment Problem	19663	71.1%
	Altitude Deviation	905	3.3%
	ATC Issue	90	.3%
	Conflict	561	2.0%
	Inflight Event / Encounter	1021	3.7%
	Procedural Deviation	4989	18.0%
	Track / Heading Deviation	445	1.6%
Airport	Aircraft Equipment Problem	336	8.9%
	Altitude Deviation	34	.9%
	ATC Issue	64	1.7%
	Conflict	816	21.7%
	Inflight Event / Encounter	313	8.3%
	Procedural Deviation	2113	56.2%
	Track / Heading Deviation	81	2.2%
Airspace Structure	Aircraft Equipment Problem	79	4.2%
	Altitude Deviation	112	5.9%
	ATC Issue	110	5.8%
	Conflict	399	21.0%
	Inflight Event / Encounter	114	6.0%
	Procedural Deviation	989	52.0%
	Track / Heading Deviation	99	5.2%
Ambiguous	Aircraft Equipment Problem	304	9.8%
	Altitude Deviation	351	11.3%
	ATC Issue	122	3.9%
	Conflict	797	25.6%
	Inflight Event / Encounter	226	7.3%
	Procedural Deviation	1098	35.3%
	Track / Heading Deviation	214	6.9%
ATC Equipment / Nav Facil	Aircraft Equipment Problem	71	7.2%
	Altitude Deviation	61	6.1%

	ATC Issue	69	7.0%
	Conflict	159	16.0%
	Inflight Event / Encounter	94	9.5%
	Procedural Deviation	394	39.7%
	Track / Heading Deviation	144	14.5%
Chart Or Publication	Aircraft Equipment Problem	563	22.7%
	Altitude Deviation	83	3.3%
	ATC Issue	12	.5%
	Conflict	44	1.8%
	Inflight Event / Encounter	58	2.3%
	Procedural Deviation	1586	63.8%
	Track / Heading Deviation	138	5.6%
Company Policy	Aircraft Equipment Problem	746	15.8%
	Altitude Deviation	61	1.3%
	ATC Issue	36	.8%
	Conflict	253	5.4%
	Inflight Event / Encounter	189	4.0%
	Procedural Deviation	3355	71.2%
	Track / Heading Deviation	75	1.6%
Environment - Non Weather	Aircraft Equipment Problem	259	12.7%
	Altitude Deviation	131	6.4%
	ATC Issue	32	1.6%
	Conflict	184	9.0%
	Inflight Event / Encounter	565	27.6%
	Procedural Deviation	810	39.6%
	Track / Heading Deviation	65	3.2%
Equipment / Tooling	Aircraft Equipment Problem	5	15.2%
	Altitude Deviation	0	.0%
	ATC Issue	2	6.1%
	Conflict	1	3.0%
	Inflight Event / Encounter	1	3.0%
	Procedural Deviation	24	72.7%
	Track / Heading Deviation	0	.0%
Human Factors	Aircraft Equipment Problem	4559	6.3%

	Altitude Deviation	6232	8.6%
	ATC Issue	3627	5.0%
	Conflict	12585	17.3%
	Inflight Event / Encounter	4321	6.0%
	Procedural Deviation	37543	51.7%
	Track / Heading Deviation	3740	5.2%
Incorrect / Not Installed	Aircraft Equipment Problem	108	29.2%
	Altitude Deviation	8	2.2%
	ATC Issue	6	1.6%
	Conflict	31	8.4%
	Inflight Event / Encounter	9	2.4%
	Procedural Deviation	205	55.4%
	Track / Heading Deviation	3	.8%
Logbook Entry	Aircraft Equipment Problem	148	24.3%
	Altitude Deviation	0	.0%
	ATC Issue	0	.0%
	Conflict	0	.0%
	Inflight Event / Encounter	1	.2%
	Procedural Deviation	458	75.3%
	Track / Heading Deviation	1	.2%
Manuals	Aircraft Equipment Problem	85	26.2%
	Altitude Deviation	1	.3%
	ATC Issue	2	.6%
	Conflict	0	.0%
	Inflight Event / Encounter	7	2.2%
	Procedural Deviation	230	70.8%
	Track / Heading Deviation	0	.0%
MEL	Aircraft Equipment Problem	8	28.6%
	Altitude Deviation	0	.0%
	ATC Issue	0	.0%
	Conflict	0	.0%
	Inflight Event / Encounter	0	.0%
	Procedural Deviation	20	71.4%
	Track / Heading Deviation	0	.0%

Procedure	Aircraft Equipment Problem	112	5.0%
	Altitude Deviation	126	5.6%
	ATC Issue	158	7.0%
	Conflict	477	21.2%
	Inflight Event / Encounter	242	10.8%
	Procedural Deviation	1031	45.9%
	Track / Heading Deviation	100	4.5%
Staffing	Aircraft Equipment Problem	3	5.6%
	Altitude Deviation	0	.0%
	ATC Issue	5	9.3%
	Conflict	7	13.0%
	Inflight Event / Encounter	5	9.3%
	Procedural Deviation	32	59.3%
	Track / Heading Deviation	2	3.7%
Weather	Aircraft Equipment Problem	310	6.4%
	Altitude Deviation	329	6.8%
	ATC Issue	49	1.0%
	Conflict	223	4.6%
	Inflight Event / Encounter	2510	52.1%
	Procedural Deviation	1210	25.1%
	Track / Heading Deviation	187	3.9%

Table 14: Classification table of aircraft anomalies

Observed	Predicted Aircraft Equipment Problem	Altitude Deviation	ATC Issue	Conflict	Inflight Event / Encounter	Procedural Deviation	Track / Heading Deviation	Percent Correct
Aircraft Equipment Problem	19663	0	0	0	310	7386	0	71.9%
Altitude Deviation	905	0	0	0	329	7200	0	.0%
ATC Issue	90	0	0	0	49	4245	0	.0%
Conflict	561	0	0	0	223	15753	0	.0%
Inflight Event / Encounter	1021	0	0	0	2510	6145	0	25.9%
Procedural Deviation	4989	0	0	0	1210	49888	0	88.9%
Track / Heading Deviation	445	0	0	0	187	4662	0	.0%
Overall Percentage	21.7%	.0%	.0%	.0%	3.8%	74.6%	.0%	56.4%

Table 15: Goodness-of-fit measures

Pseudo R-Square	
Cox and Snell	.378
Nagelkerke	.395
McFadden	.150

The above classification table shows mixed results in the model's ability to predict aircraft anomalies from factors data. The overall accuracy of the model is 56.4 percent, and the model is best at predicting Procedural Deviation Anomalies and Aircraft Equipment problems. This is due to the high levels of association between single factors such as Human Factors issues or Aircraft

Equipment problems and the predicted anomaly. Those anomalies that have fewer occurrences of individual factors are harder to predict, as less information exists.

The pseudo R-squared scores provide a rough measure of the model's ability to account for variability in the data. The two commonly used scores shown in the above table, Cox and Snell and Nagelkerke; both demonstrate that the model can explain roughly 40 percent of the variability in the input data.

The classification table resulting from the regression findings demonstrated an overall model accuracy of 56 percent. This is significant because it assigns some underlying pattern to factors and anomalies. When some information about the factors that contribute to an anomaly exists, the anomaly can be predicted at levels better than chance.

This figure is misleading, as individual model accuracy for single anomalies was much higher. For example, the model was able to identify almost 89 percent of Procedural Deviations, and nearly 72 percent of Aircraft Equipment Problem anomalies. This was likely due to the fact that these two anomalies had overwhelmingly single factors. Specifically, Human Factors issues usually indicated a Procedural Deviation, as these types of anomalies were by far the most often encountered. Namely, when Human Factors issues are present, it is highly likely that a Procedural Deviation will occur.

The goodness of fit analysis, at best, reported that the logistic model could only account for about 40 percent of the variability in the data. This is a reasonable "real-world" figure, as most "acceptable" or "good" R-Squared values hover around 0.5, meaning they can explain half of the variability in the data. The model was better at correctly identifying those anomalies that had more contributing anomaly records. The anomalies that had few definite factors associated with them were harder for the model to predict correctly. These findings demonstrate that, given enough input data, multinomial logistic regression can identify anomalies from factors, albeit with a level of reliability dependent on amount of data available.

This type of approach has been employed by epidemiological studies, which investigated what factors were present to determine whether they contributed to a disease or not. The main difference between the two is that epidemiological studies look for a "flag" value, i.e., "yes" the patient has the disease or "no" he does not, assuming certain risk factors exist. Thoraval et al (1997), for example, published a study linking risk factors that affected the severity of Hepatitis C.

Identification of Active and Latent Factors from Anomaly Reports

Categories

PASW Modeler 13 analyzed the 127,776 text records associated with anomalies. The results of the categories extracted from the records are shown below:

Table 16: Concept counts extracted from text data

Concepts	Records Containing	Percentage
Unsafe Conditions	50625	39.60%
Rule-Based Errors	33161	26.00%
Uncategorized	27855	21.80%
Skill-Based Errors	25073	19.60%
Weather	21818	17.10%
Knowledge-Based Errors	19054	14.90%
Aircraft Issues	10110	7.90%
Violations	5287	4.10%
Unsafe Supervision	271	0.20%
Perceptual Errors	55	0.00%
Total	127766	100.0%

The most commonly extracted category was the one containing concepts related to unsafe conditions, followed by rule-based errors, then skill-based errors, followed by knowledge based errors. Approximately 22 percent of the records were not categorized by the text mining algorithm, meaning they did not contain any of the keywords sought out by the algorithm.

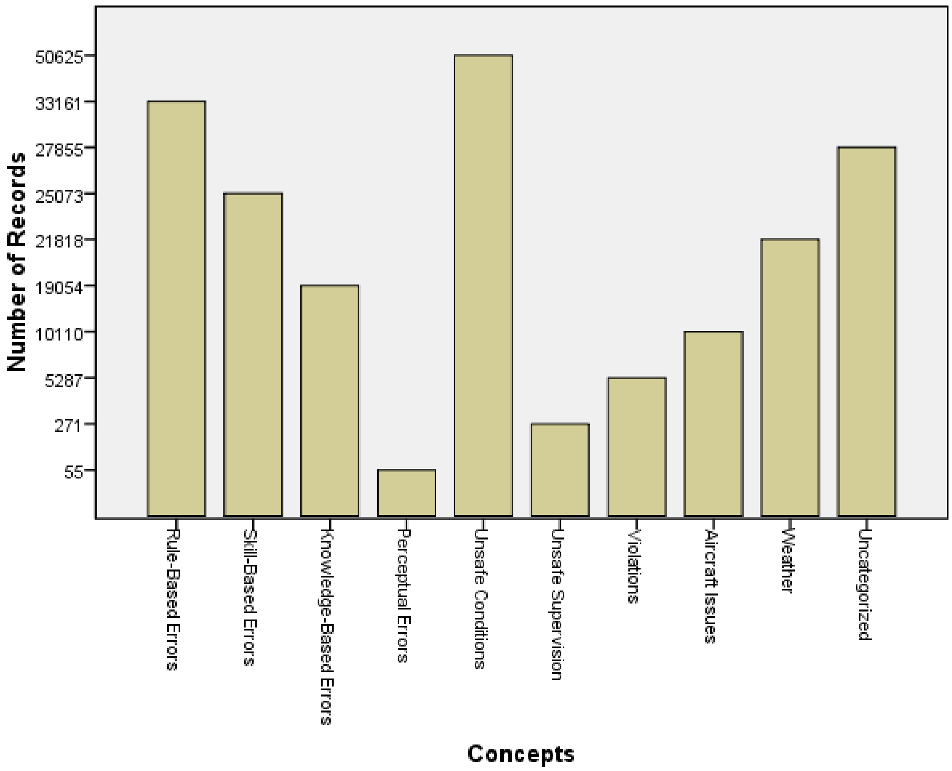


Figure 29: Frequencies of categories extracted from text data

Types

The below table displays the types of concepts extracted, counts of how many occurrences of a given type there were, and how many records contained that type. Types were constructed from concepts (frequently occurring words in the corpus); a group of related concepts sharing a common meaning or theme was grouped into a type.

Table 17: Types of concepts extracted

Types	Number of Occurrences	Number of Records Containing This Type
<Unknown>	3116095	127761
<Routine Behaviors>	192086	87198
<Non-Routine Behaviors>	65826	43003
<Rule-Based>	45520	32584
<Skill-Based>	37748	25004
<Weather>	36243	21508
<Knowledge-Based>	24076	19054
<Violation>	6150	5192

The unknown (or uncategorized) type was the most often encountered type within the anomaly report corpus. Concepts associated with routine behaviors made up the first meaningful type. These were concepts such as approach, cruise flight, and landing. The next most prevalent and utilized type was Non-Routine Behaviors; more than 43,000 records contained a concept or word associated with actions taken by the pilot that were extraordinary in nature. Concepts grouped into Rule-based and Skill-based error types were the next most often encountered, followed by Weather and Knowledge-Based types. Violation types were encountered in 5,192 of the records mined.

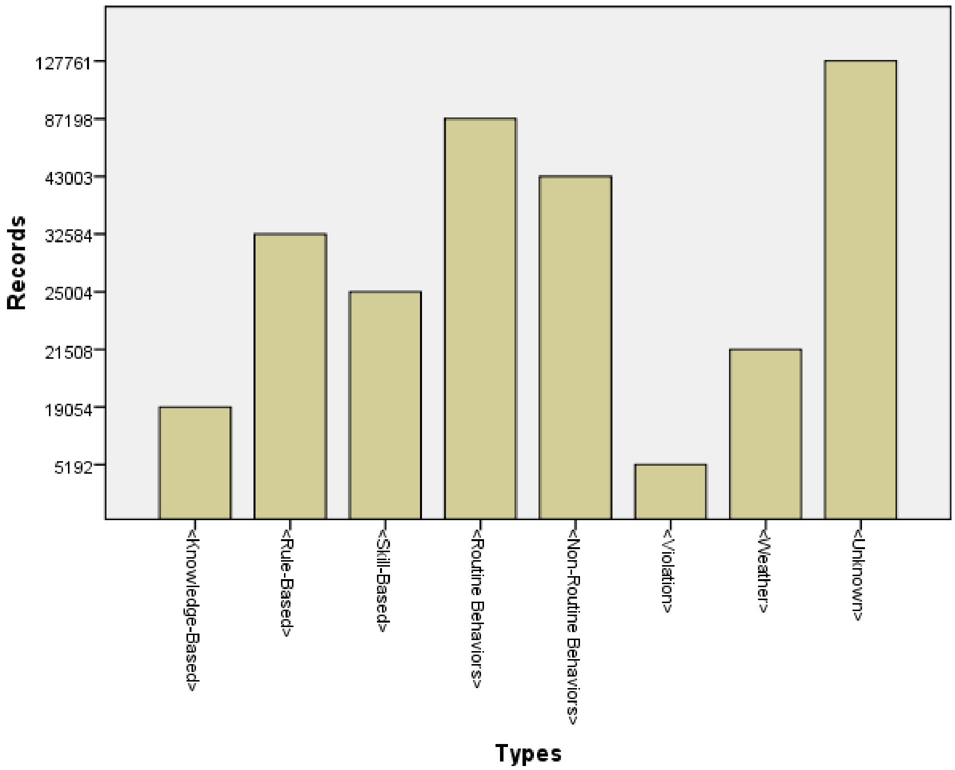


Figure 30: Counts of number of records in which each type appears

The above chart displays the number of documents that contained each type of concept. Routine Behaviors made up the vast majority of these records. The next most prevalent type found was Non-Routine Behaviors. Finally, the Skill-Based, Rule-Based, and Knowledge-Based error types followed. Weather types were found in about 22,000 records.

The types extracted from the record corpus were made up of "concepts," or frequently encountered words representing a grander whole or concept. As with the categories, most words extracted had no type associated with them, and were left unclassified as <Unknown>. A full list of the types created and their component keywords can be found in the Appendix.

Routine Behaviors was the “control” type, created initially as a test to see how the system would handle parsing concepts into words. No analysis or links were performed with this type, but it was left as an indicator to assure the system was performing correctly.

Non-Routine Behaviors constituted extreme maneuvers, troubleshooting, evasion, unruly passengers, and other nonstandard keywords identified by the literature. Over 43,000 anomaly reports contained at least one reference to non-routine behaviors, many containing multiple instances of Non-Routine Behaviors. This was demonstrated by the existence of more than 65,000 recorded instances of this type.

Rule-Based concepts were the next most frequently encountered type. This type was linked to the Rule-Based Error category, and contained concepts such as bad decision, wrong impression, “I was sure,” and other keywords that would lead to an incorrect action. Pilots making Rule-Based errors often apply an inappropriate action plan or decision to a situation, but are sure they are making the right decision. Reason (1990) calls these types of impressions “strong but wrong.’ These types were hard to identify, as most Rule-Based errors are domain specific and described in terms restricted to a given domain; it is very challenging to create generic keywords to identify Rule-Based error types.

Skill-Based concepts comprised the next most frequently encountered type. Skill-based error types were easier to identify than Rule-Based types, as they were usually generic actions such as “pushed the wrong button,” “pulled the wrong lever,” “I forgot,” “oversight,” and other such generic error terms. Reason (1990) uses these terms in several scenarios to describe the differences between the three error types.

Weather types were the “low-hanging fruit” in mining the record corpus; however, weather as a factor in anomalies would generally have to be undesirable weather. This prompted keywords such as low visibility, turbulence, rain, ice, and IMC, or Instrument Meteorological Conditions, signifying that safe flight by using information available through the windshield was not possible. Weather was a factor in 21,508 anomaly reports. This count was higher than the number of anomaly reports that had Weather listed as a primary factor associated with the anomaly. This suggests that either the Weather type was too generic, or perhaps weather was a more serious contributing factor to anomalies than the frequency counts suggest.

Knowledge-Based types were the next most frequently encountered type. These types comprised keywords such as unfamiliar, unsure, did not know, and student. Student was chosen because many anomaly reports were from instructors training student pilots. These student pilots would commit knowledge-based errors, as many did not have the requisite knowledge of a procedure or checklist, which in turn contributed to anomalies. Knowledge-based errors usually involve problem-solving activities, as new or not frequently accessed or applied knowledge must be recalled. This can take time, or the knowledge recalled may be inaccurate or incomplete, thus contributing to onsets of anomalies.

The final type extracted was Violations. Violations were identified with keywords such as exceeded, inebriated, overspeed, too fast, and violated. Unfortunately, many of the aviation violations were encoded or abbreviated and all could not be identified. Generic terms implying violations were used for this part of the study.

Identification of Latent and Active Factors

The PASW Modeler 13 Text Analytics feature was used to construct document webs examining strengths of associations of categories within records. These are shown in the web diagrams in the below figures. Lines that are bolder denote stronger relationships, that is, higher frequencies of occurrences of categories within records. The size of the Category dot representation itself represents the number of anomaly reports containing the category.

Human Error Categories

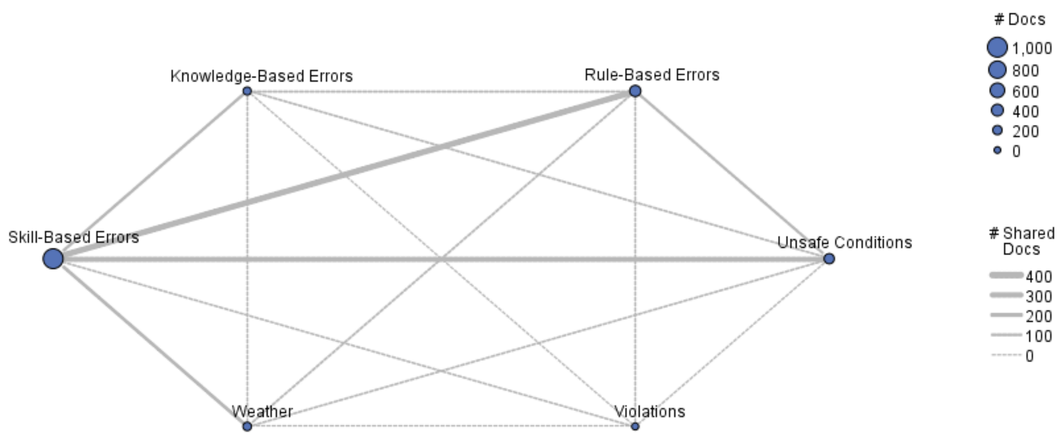


Figure 31: Skill-based errors category web

Skill-based errors are frequently found within anomaly reports along with Rule-Based errors, as well as Unsafe Conditions. There is a weaker association found with Weather, as well as a tenuous association with Violations.

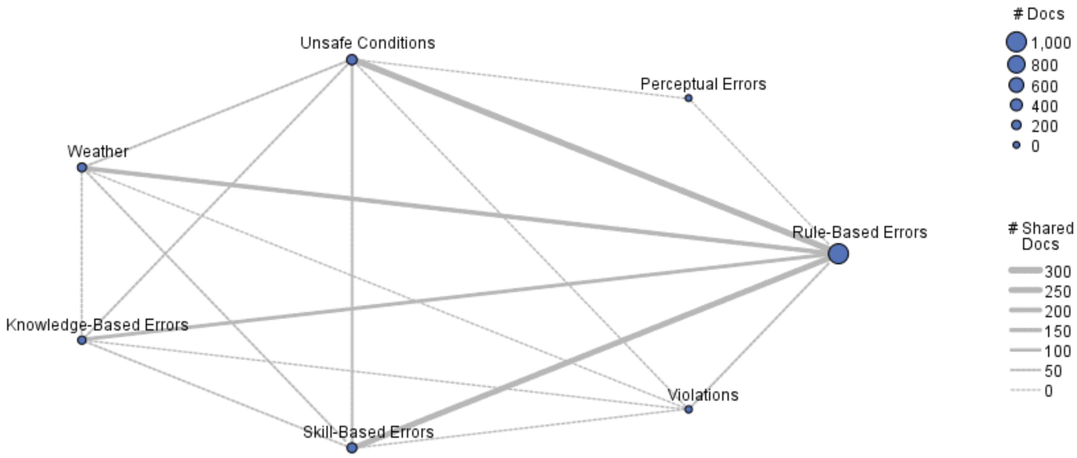


Figure 32: Rule-based errors category web

Rule-Based Error types are found alongside Skill-Based errors, and have a high association with Unsafe Conditions. There are weak associations with Perceptual Errors and Violations. Rule-Based Error types also share some association with Weather and Knowledge-Based Errors in anomaly reports.

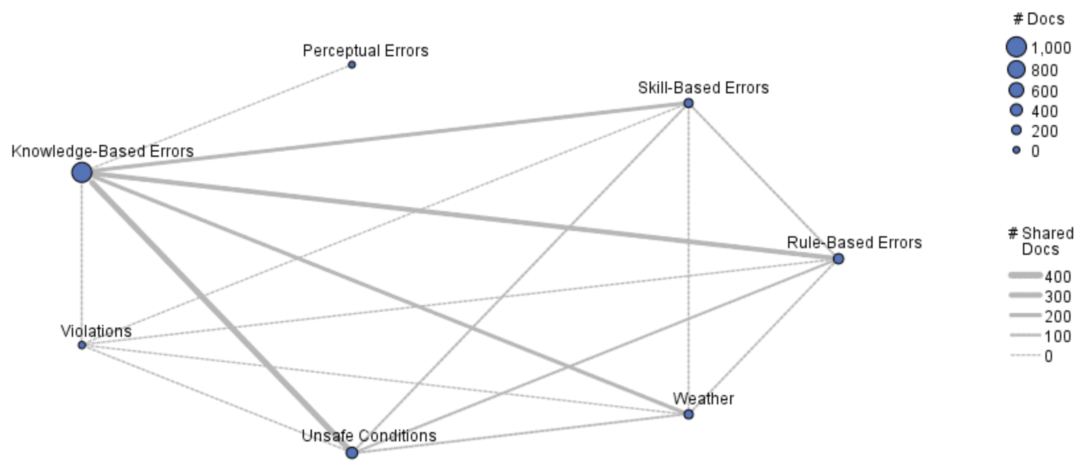


Figure 33: Knowledge-based errors category web

Knowledge-Based Error types appear to be weakly associated with Perceptual Errors and Violations. There are strong associations with Unsafe Conditions types, Rule-Based Errors, and Skill-Based Errors. Knowledge-Based Error types have some associative qualities with Weather types as well.

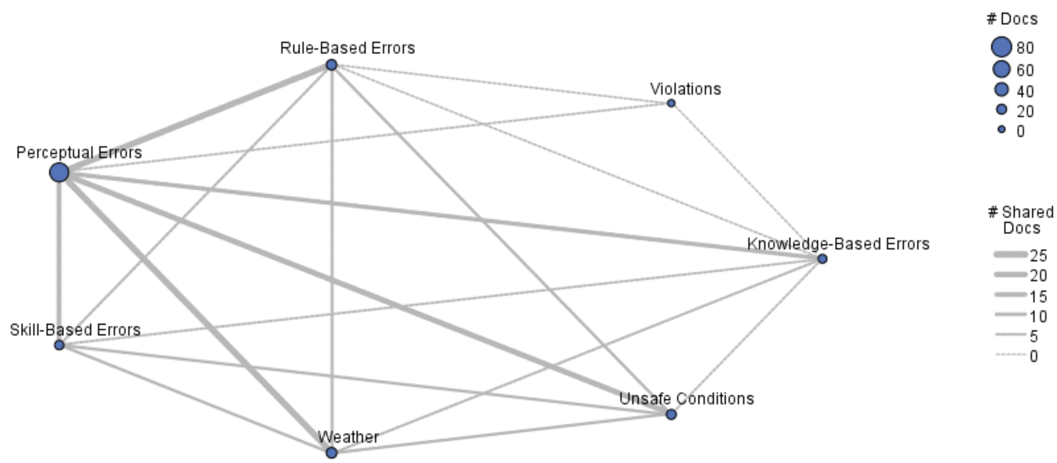


Figure 34: Perceptual errors category web

Perceptual Error types were strongly associated with Weather types, as well as Rule-Based Error types and Unsafe Conditions. There were light associations to Skill-Based Error types and Knowledge-Based Error types. The weakest associations were with Violation types.

Supervisory and Administrative Categories

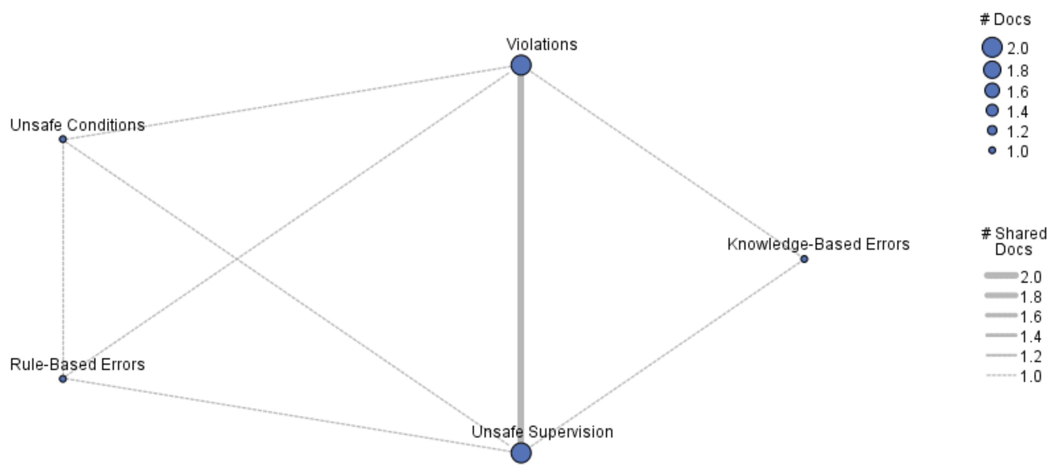


Figure 35: Unsafe supervision category web

Unsafe Supervision types were strongly associated with Violation types. The other associations were too trivial and few in number to merit consideration.

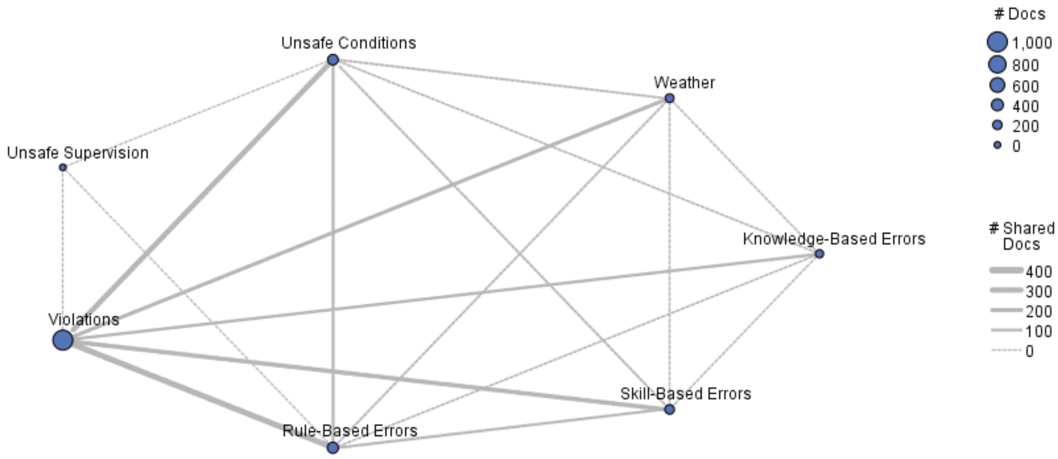


Figure 36: Violations category web

Violations types were associated with many other categories, including all three types of Human Error types, as well as Weather and Unsafe Conditions types. Although violations were not common by magnitude in number of records, they were frequently associated with other factor types.

Latent and Active Factors Categories

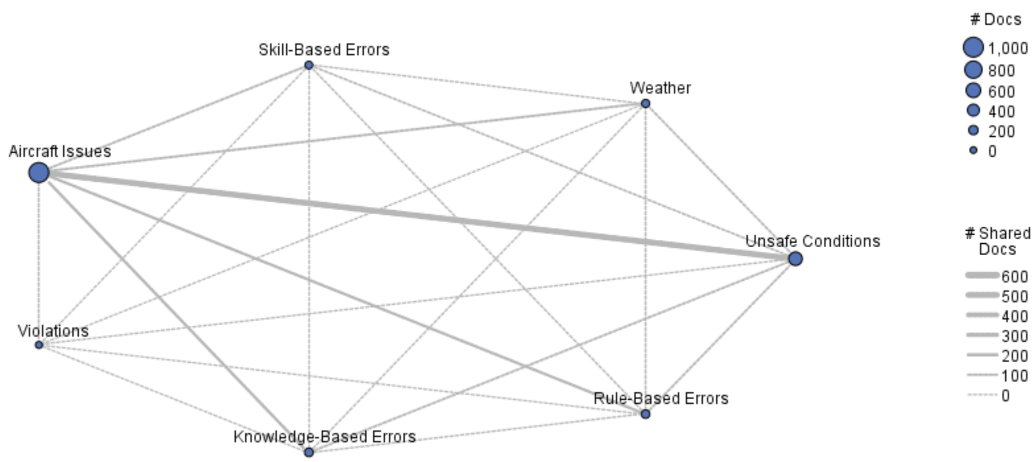


Figure 37: Aircraft issues category web

Aircraft issues types were those associated with equipment malfunctions. They were strongly associated with Unsafe Conditions types, and also were loosely associated with the three Error types as well as Weather types. There was a very weak association with violations.

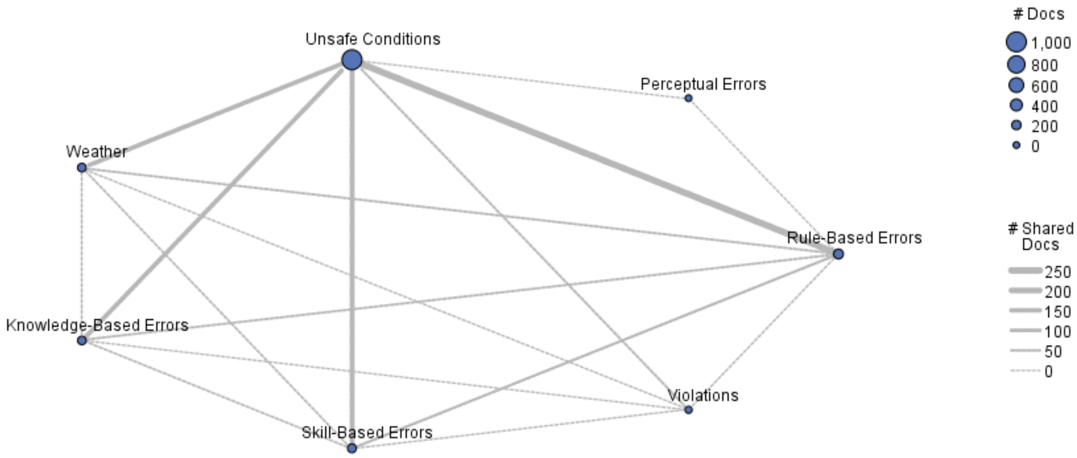


Figure 38: Unsafe conditions category web

The Unsafe Conditions type had associations with all three Human Error types, with the strongest association being the Rule-Based Error type. It also was associated with Weather types. There was a weak link with Perceptual Errors.

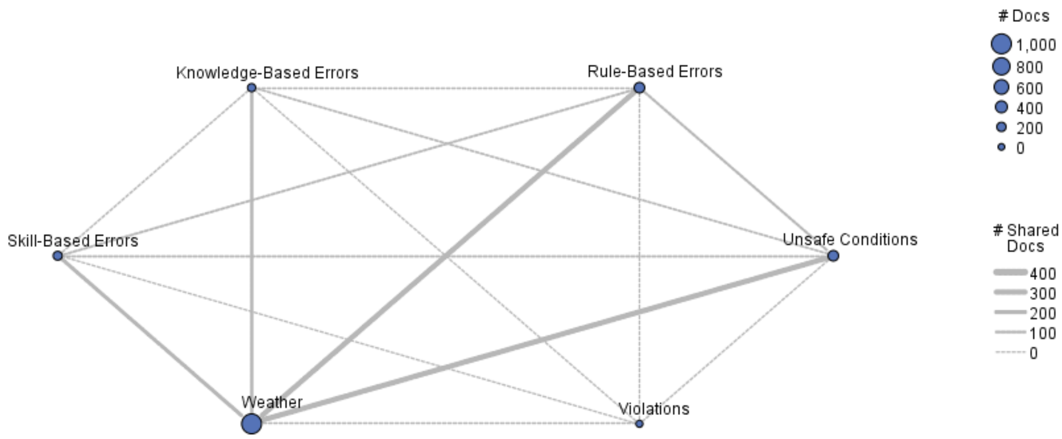


Figure 39: Weather category web

The Weather category type was loosely associated with Skill and Knowledge-Based Error types. It also was strongly associated with Rule-Based Errors and Unsafe Conditions. Weather had a tenuous relationship with Violations.

Associations between Latent and Active Contributors to Anomalies

The categories created were associated with each individual anomaly report, and were used to create a structured means of clustering the data. The categories were placed into columns, and a flag was assigned to each column and row indicating the presence of that category within the anomaly report. These flags thereby provided the non-categorical data element that could be used to create clusters of the data, in order to demonstrate how supervised text mining could be used to create new associations for factors contributing to anomaly reports.

This part of the analysis employed the "Assessments" data column, and the output cluster column from each respective clustering method. The clustering methods employed were Kohonen Self-Organizing Maps, Two-Step Clustering, and K-Means. After running the clustering operation, the generated clusters were compared to the manually categorized anomaly data to evaluate clustering performance. All operations were run for one pass on the data.

Cluster Summary Data

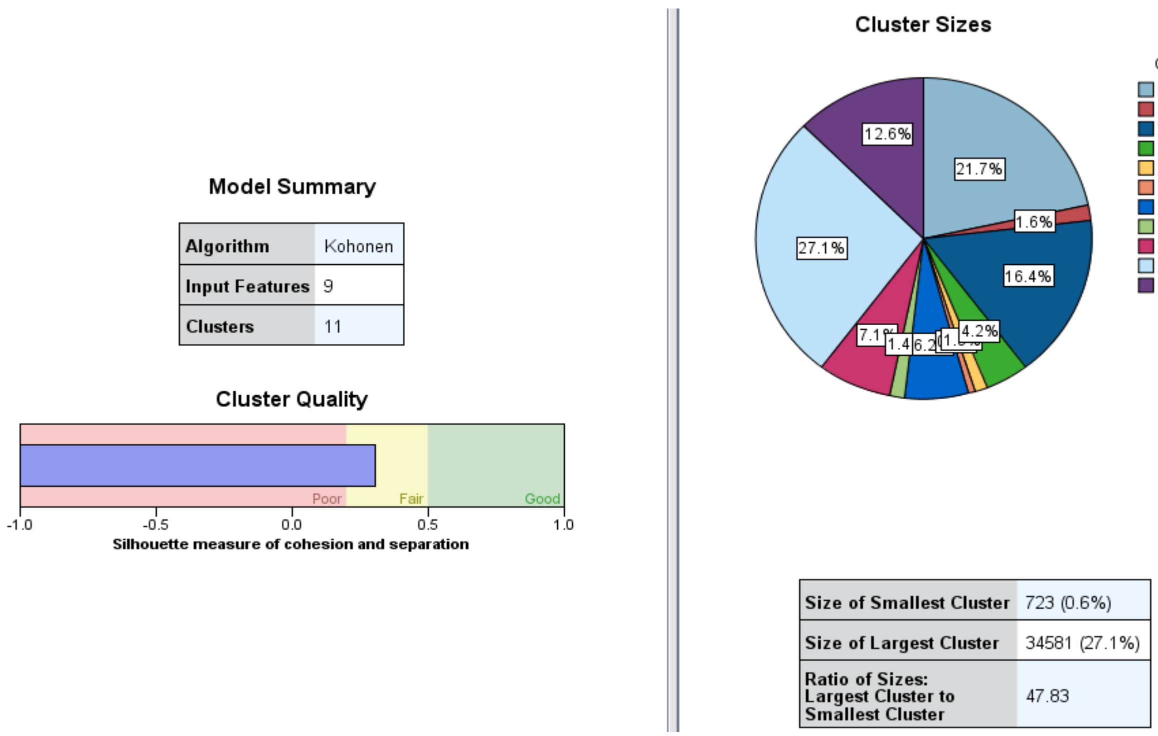


Figure 40: Summary of the Kohonen clustering operation

The Kohonen clustering algorithm transformed the 9 category flags into 9 clusters. The internal measure of cluster cohesion and separation was regarded as "fair" by the software.

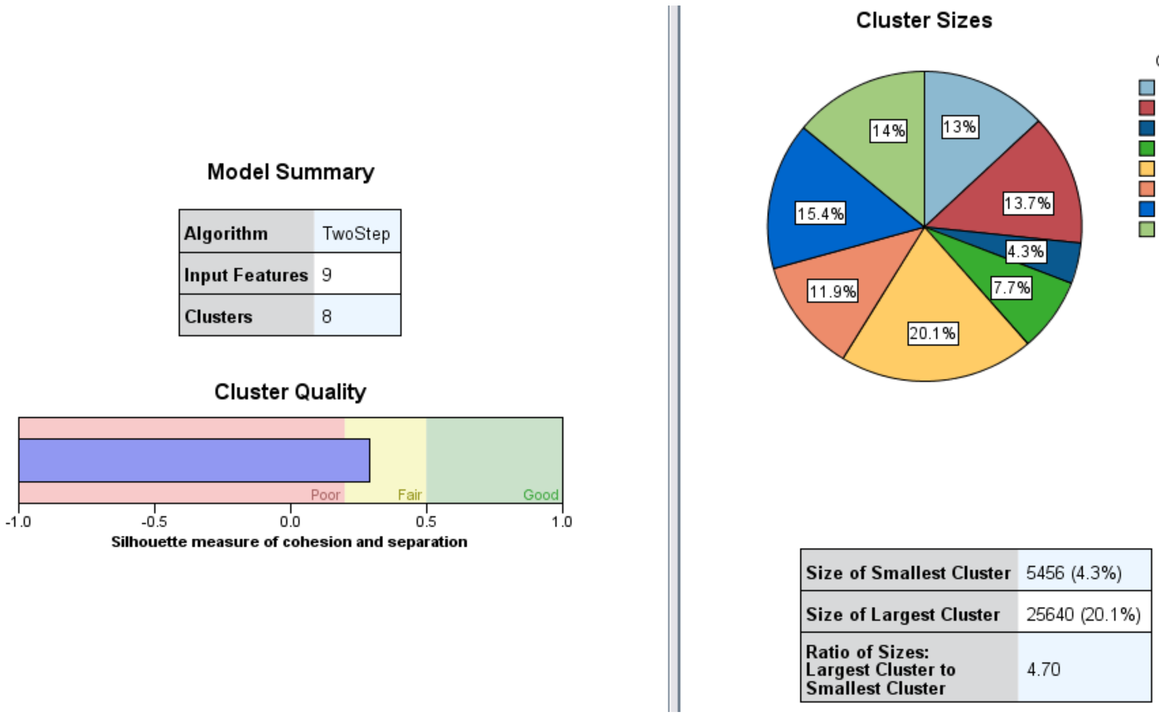


Figure 41: Summary of the two-step clustering operation

The Two-Step clustering algorithm produced 8 clusters; these 8 clusters scored a "fair" measure of cohesion and separation. The clusters were much more even in terms of size as compared with the Kohonen clusters generated.

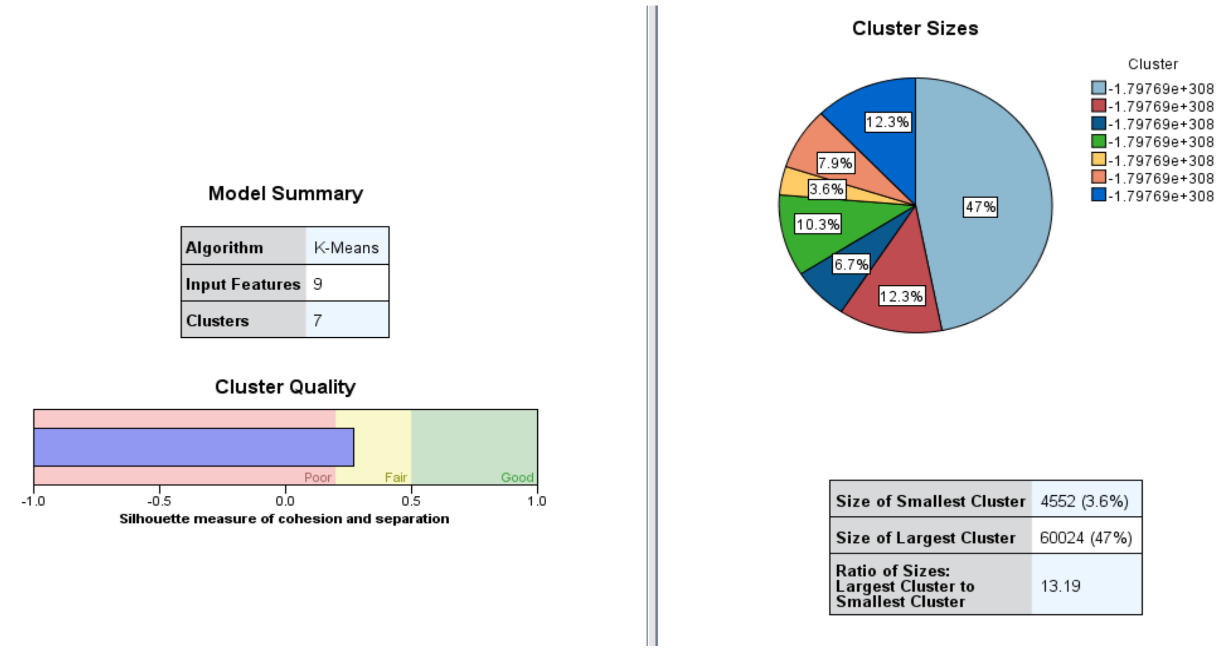


Figure 42: Summary of the K-Means clustering operation

The K-Means clustering operation produced 7 clusters, with one containing nearly 47 percent of the data analyzed. The measure of cohesion and separation is fair/poor for this clustering operation.

Cluster Comparisons with Assessments

The clusters generated were compared with the assessments assigned by the maintainers of the ASRS database in order to determine whether any of the algorithms employed could serve as automated classification tools. The reference "clustering" in the form of assessment categories generated by the ASRS development team is presented in the below figure:

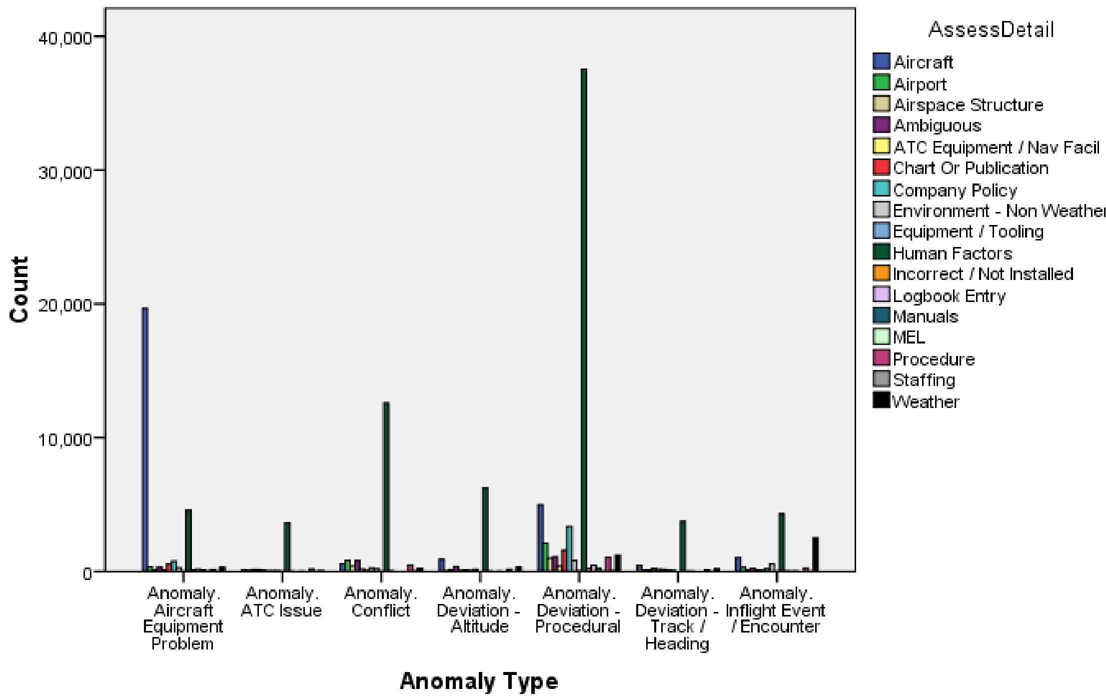


Figure 43: "Clustering" as organized by ASRS anomaly types

The above figure shows how the various Assessments are broken down across Anomaly types. Human Factors is the most prevalent factor for all anomalies save for Aircraft Equipment problems, whose most commonly occurring factor is Aircraft issues. The largest "cluster" that was created by the Anomaly partitioning method was Procedural Deviations, with 56,087 anomaly reports. Aircraft Equipment problems and Conflicts were the next largest clusters, with 27,539 and 16,537 anomaly reports, respectively. The remaining clusters all had fewer than 10,000 anomaly reports associated with them.

Kohonen Self-Organizing Map Clustering

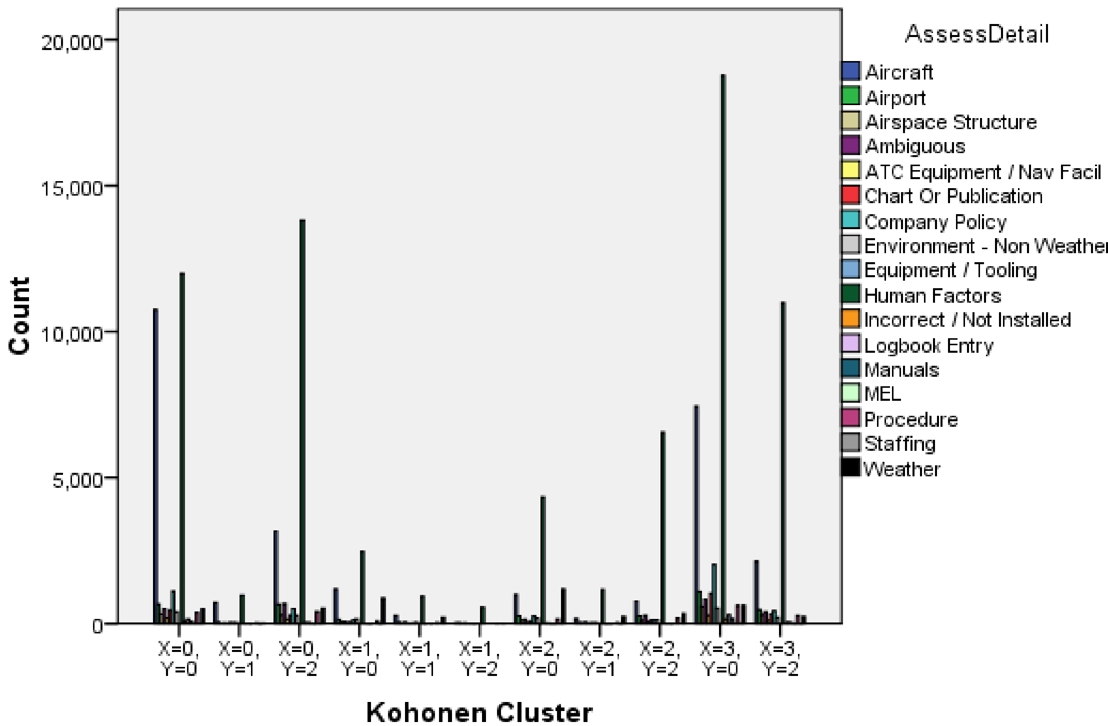


Figure 44: Clustered bar chart of Kohonen clustering operation

The Kohonen clustering operation created 11 clusters, each with a varied assortment of Assessments. The most prevalent Assessment found in all clusters was Weather, followed by Human Factors. Some clusters were much larger than others. The largest cluster was the X=3, Y=0 cluster, with 34,581 anomaly reports categorized. There were three other large clusters, containing 27,700, 20,962, and 16,093 anomaly reports. The remaining clusters all had fewer than 10,000 anomaly reports categorized within them.

Two-Step Clustering

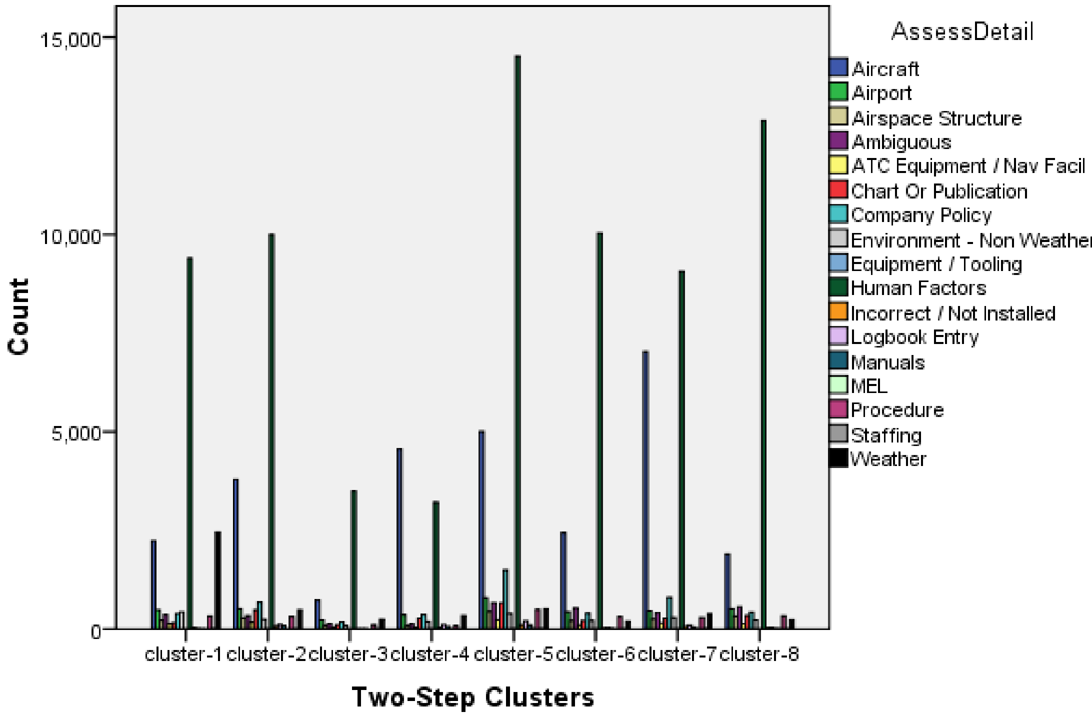


Figure 45: Clustered bar chart of the Two-Step clustering operation

The Two-Step clustering algorithm produced 8 clusters, each containing a sizeable portion of the anomaly reports. The most prevalent assessment by count found in each cluster was Human Factors. Weather was the second most prevalent assessment found in each cluster. This clustering method spread the various assessments throughout each cluster. Most clusters contained about 20,000 anomaly reports, with only clusters 3 and 4 having 4,456 and 9,816 anomaly reports associated with them, respectively.

K-Means Clustering

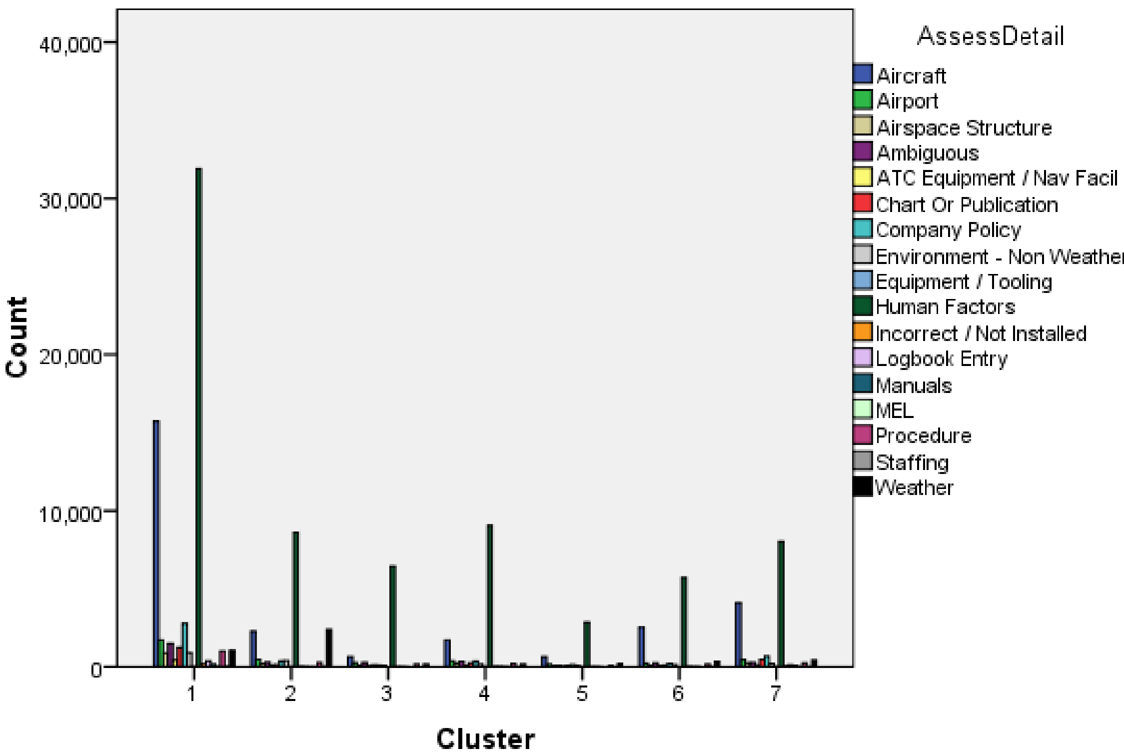


Figure 46: Clustered bar chart of the K-Means clustering operation

The K-Means clustering algorithm produced the above output. Cluster 1 had over 87,920 anomaly reports associated with it, with Human Factors and Aircraft issues being the most prevalent assessments found associated with that cluster. The other clusters, save for the largest Cluster 2, which contained 17,456 records, all contained fewer than 10,000 anomaly reports. This clustering method created an output very similar to the classification used by the ASRS database administrators. However, the output was not identical; the value of Cohen's *kappa* for clustering agreement was -0.015. The results of the classification are shown below.

Table 18: Classification analysis of k-means clustering algorithm

		K-Means Cluster Number							
		1	2	3	4	5	6	7	Total
Anomaly Type	Deviation - Procedural	26152	5974	4213	6168	2562	4019	6999	56087
	Aircraft Equipment Problem	15679	2258	526	1593	535	2522	4246	27359
	Deviation - Track / Heading	1581	678	923	1125	182	339	466	5294
	Inflight Event / Encounter	2859	3792	403	456	362	694	1110	9676
	ATC Issue	2024	385	408	700	106	331	430	4384
	Deviation - Altitude	4024	836	689	804	326	978	777	8434
	Conflict	7705	1779	1411	2319	479	1193	1651	16537
Total		60024	15702	8573	13165	4552	10076	15679	127771

Table 19: Cohen's kappa for k-means cluster agreement

Measures of Agreement					
		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Measure of Agreement	Kappa	-.015	.001	-10.656	.000
N of Valid Cases	127771				

Text Mining Predictive Models

The categories that were created and flagged for each anomaly report record were then used to create a predictive model. The intended aim was to determine whether, by mining for the created categories, and given identified categories as an input, the model could correctly classify the record according to Anomaly type. A Neural Network predictive model was employed within PASW Modeler 13. The model summary is shown below. The output column was compared to the actual Anomaly category column.

Table 20: Neural Network model accuracy

Results for output field Anomaly		
Comparing $N-Anomaly with Anomaly		
Correct	72,162	56.48%
Wrong	55,609	43.52%
Total	127,771	

The model was only able to correctly classify about 56 percent of the Anomaly reports based on the 9 inputs of the category flags. The model was trained 5 times.

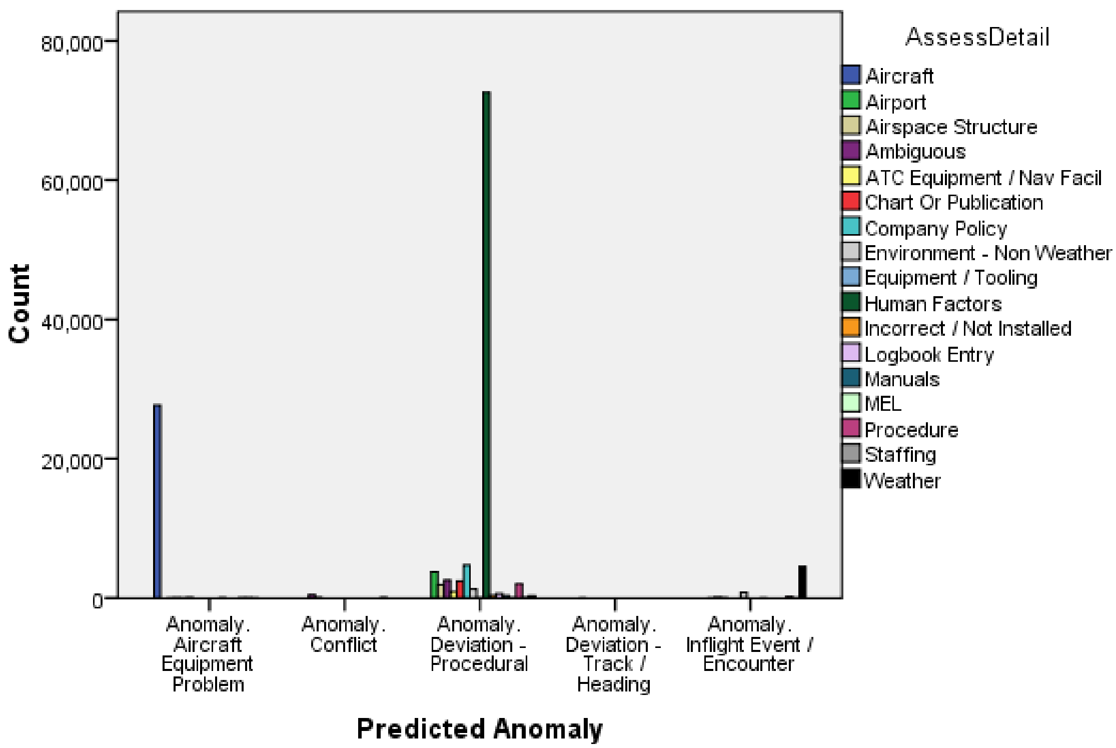


Figure 47: Predictive performance of neural network model

The Neural Network model was able to accurately identify Aircraft Equipment Problems and Procedural Deviations. However, it was unable to reproduce the other three anomalies based on the category flag data alone.

The three clustering algorithms produced markedly different results. All of the algorithms employed were able to generate clusters with a "Fair" rating on cohesion and separation.

Kohonen

The Kohonen clustering algorithm generated 11 clusters. These clusters were very varied in size, ranging from 723 anomaly reports to 34,581 anomaly reports. No meaning or structure was apparent in these generated clusters.

Kohonen clustering trains a neural network to cluster elements that tend to be similar to each other. The Kohonen clustering algorithm used within PASW Modeler 13 employs a network with an input layer and an output map, consisting of nodes with connections between the nodes. Kohonen maps are also used for dimensionality reduction, as many inputs become summarized into a few output features that preserve the underlying structure of the inputs; Kohonen maps can produce similar outputs to factor analysis or principle component analysis (PCA) (Kohonen, 1998).

The network is "trained" through the presentation of data to the input layer, and these values are propagated through the network to the output layer. The network is "trained" repeatedly by adjusting the connection strengths between the input and output connection values, so that the outputs better match patterns of inputs presented to the network. The training results in the creation of a two-dimensional "map," which organizes itself based on the training activity (hence the term self-organizing-map, or SOM).

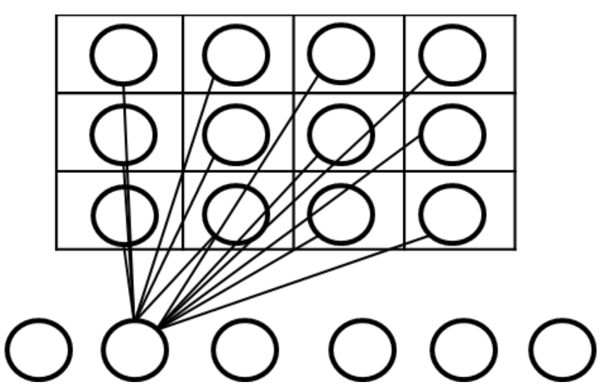


Figure 48: Input and output nodes in a Kohonen network

The above figure demonstrates connections from one node to the output nodes in a

Kohonen map. The input nodes are the nodes below the grid, and the grid represents the output layer. All input nodes are connected to the output layer – but these connections have been omitted for clarity. Kohonen maps, when trained, uncover which input records summarize or are similar to the majority of records, called "strong units." Conversely, the records that are different from the majority form "weak units." The strong units become the cluster centers (Kohonen, 1998).

Kohonen self-organizing maps were not the best choice for document clustering operations. The high dimensionality and variability of the input patterns likely left the algorithm with many "weak units," and few "strong units" around which to build appropriate clusters.

Two-Step

The Two-Step algorithm produced 8 clusters that were more uniform in size compared to those produced by the Kohonen algorithm. The Two-Step algorithm appeared to create each cluster to be similar to the next; all clusters created were similar to each other in terms of their content. The algorithm appeared to partition the data into 8 uniform clusters with similar characteristics.

The Two-Step clustering operation clusters a dataset into distinct groups; it does not attempt to predict an outcome. The Two-Steps that give the method its name operate as follows:

1. A first pass compresses raw input data into a manageable set of sub clusters
2. The second step employs a hierarchical clustering method that merges the small sub clusters into larger and larger clusters

The literature indicates that traditional hierarchical clustering methods are not well-suited to document clustering without modification (Zhao et al, 2005). Two-Step clustering belongs to a class of Agglomerative Algorithms, which have been shown to focus too closely on certain input variables, thereby prematurely and inaccurately placing items into a cluster where they may not ideally belong. Zhao et al (2005) present a new class of clustering algorithms, called, "Constrained Agglomerative Algorithms," that reduce this early stage "tunnel vision" effect that introduces errors in clustering. Unfortunately, these new algorithms were not available for use in PASW Modeler 13.

K-Means

The K-Means algorithm produced seven clusters. These clusters resembled the original seven anomalies. Upon further inspection, however, the clustering algorithm was only able to coincide with the original anomaly categorizations 26 percent of the time. The value of Cohen's *kappa* was -0.015, a very low level of agreement between the clustering algorithm and the database administrators of the ASRS (Cohen, 1960).

The K-Means clustering algorithm may have produced more desirable results than the other algorithms because of its iterative nature. Records are sorted initially by a set of "starter clusters" created from their data. Based on the input values from each record, it is assigned to a cluster. The algorithm then "re-checks" the record to see if it belongs in the cluster assigned to it. Steinbach, Karypis, and Kumar (2000) claim that K-Means clustering is suited for document clustering activities as it applies a seemingly more global approach than the other algorithms. K-Means clustering techniques do not focus on a single nearest neighbor variable; hierarchical clustering methods may exhibit a "tunnel vision" property by focusing too closely on one input variable. The nature of hierarchical clustering techniques is that they may "lock in" a document's cluster early on in the clustering process, and it is impossible to correct this inadvertent early clustering operation as the algorithm progresses.

Classification of Anomaly Reports by Content Alone

The dimensionality reduction classification activity was performed on 100 of the most commonly occurring words. This was done to make the effort as "unsupervised" as possible, meaning that no investigator input was required to choose the inputs. The 100 words were transformed into a document difference matrix, and this matrix was reduced by the Diffusion Maps method described by Lafon and Lee (2006). The results of the classification are shown below. The diffusion maps algorithm creates large matrices, namely a pairwise distance matrix using MATLAB's pdist() command. This command has memory and computation limitations that become quickly apparent when large datasets are employed. To accommodate this limitation a reduced dataset of only 1,000 records was used.

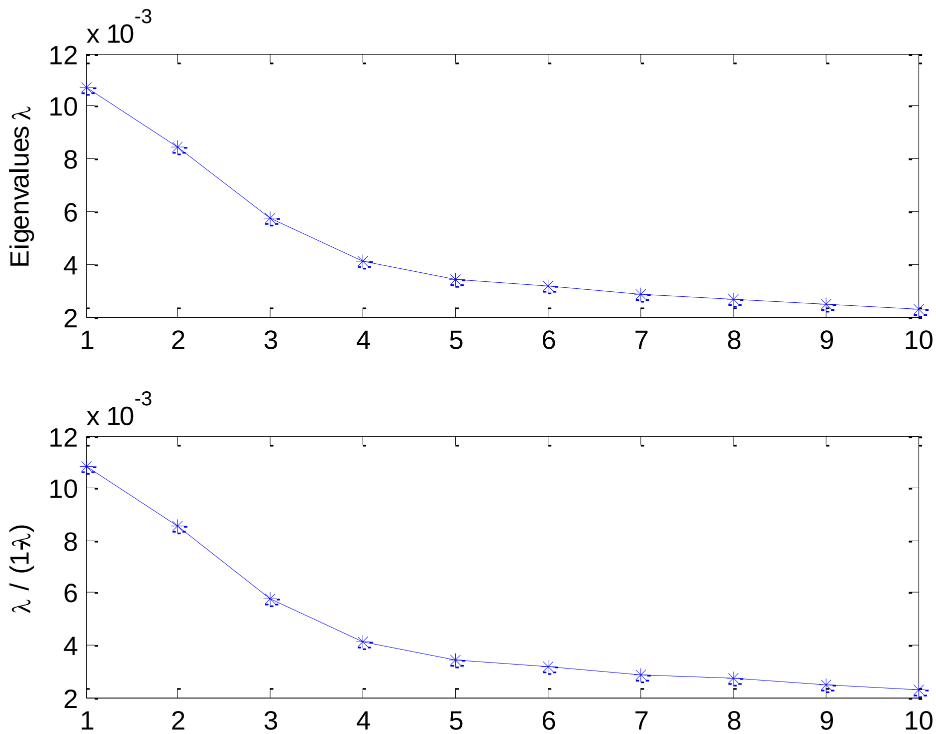


Figure 49: Falling eigenvalues plot from dimensionality reduction

The dimension reduction activity produced the above figure. The first nontrivial eigenvalues describe paths that the underlying structure of the data implies. The various vectors that describe the data in the forms of word presence or absence may follow common paths, and these eigenvalues describe those vectors that account for much of the underlying structure (Lafon & Lee, 2006). The fall-off of these eigenvalues dictates the dimensionality reduction and weighting of relevant eigenvectors.

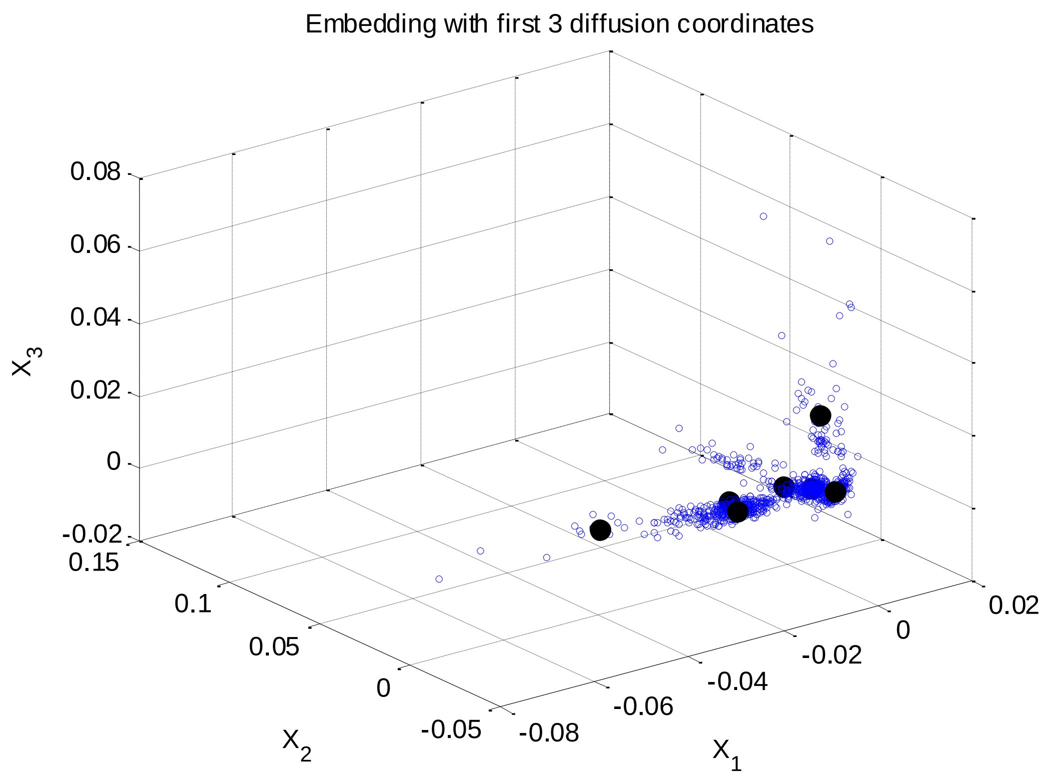


Figure 50: Diffusion map of the first three embedding coordinates

The above figure displays the embedding of the first three diffusion coordinates. The data are clustered around these coordinates. These arrangements describe the underlying structure of the data. These data were employed in a *k*-means clustering algorithm to classify the anomaly records by words contained within them according to anomaly types. The above graph is a realization of a cloud of points where the rescaled eigenvectors are the coordinates. The above graph is a lower dimensionality representation of the data that reveals underlying structure.

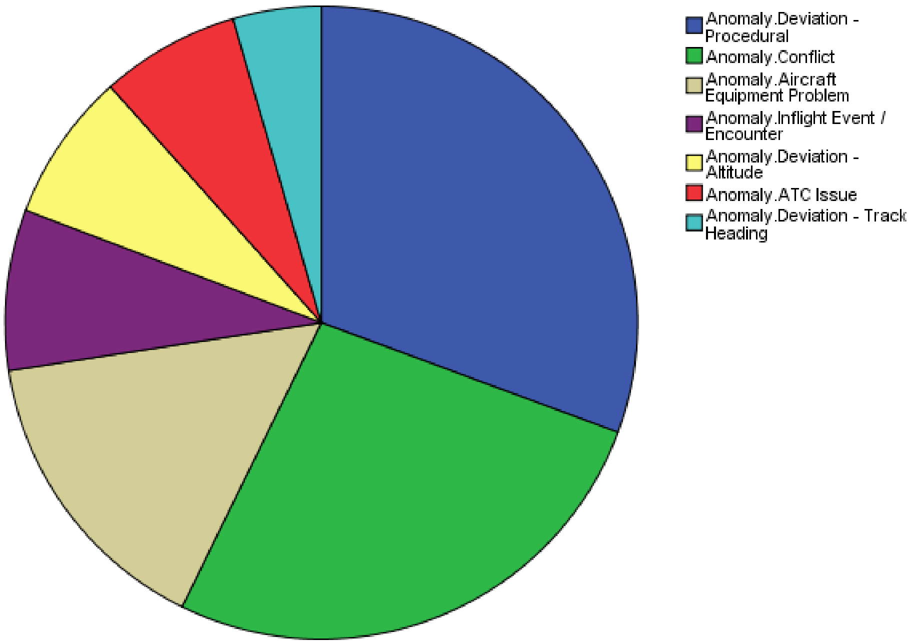


Figure 51: Composition of the dimensionality reduction dataset – actual anomalies

Table 21: Actual anomaly frequencies from reduced dataset

Anomaly Type	Frequency	Percent
Anomaly.Deviation - Procedural	307	30.6
Anomaly.Conflict	266	26.6
Anomaly.Aircraft Equipment Problem	153	15.3
Anomaly.Inflight Event / Encounter	82	8.2
Anomaly.Deviation - Altitude	76	7.6
Anomaly.ATC Issue	71	7.1
Anomaly.Deviation - Track / Heading	45	4.5
Total	1000	100.0

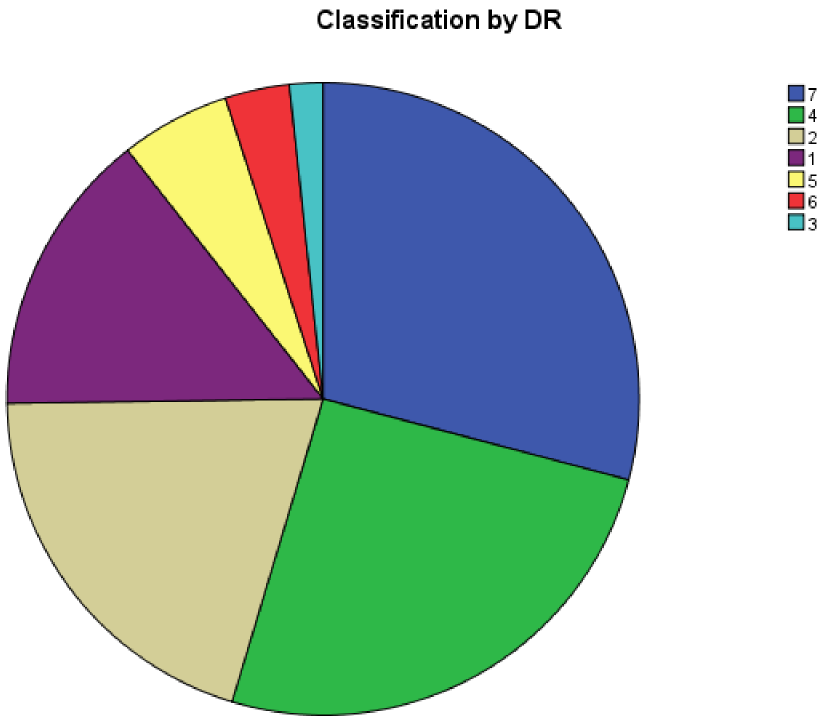


Figure 52: Composition of the dimensionality reduction dataset – classification results

Table 22: Category frequencies as created by *k*-means labeling

k-means label	Frequency	Percent
7	292	29.1
4	255	25.5
2	201	20.1
1	146	14.6
5	56	5.6
6	33	3.3
3	17	1.7
Total	1000	100.0

The reduced 1,000 record dataset contained a representative sample of the data. The above figures and tables describe the data in terms of frequencies and visual distributions. After running the algorithm, a *k*-means labeling of the outputs was run to classify the records.

The algorithm created a distribution of records very similar to the records classification distribution given by the ASRS database administrators. When comparing the output of the algorithm to the predetermined categories assigned by the ASRS database administrators the accuracy is 23.8 percent. These findings suggest that dimensionality reduction can be used to classify anomaly reports in an unsupervised fashion. By using more salient, relevant keywords in the process, this method could yield far better results as the algorithm focuses on using these features to classify the data. The fact that the frequencies of classification are so similar to classification activities performed by the ASRS database administrators suggests that the algorithm identifies features in the dataset. Most records had multiple classifications and contributing factors (however only the first one was used) and the dimensionality reduction activity may focus on a feature that was not used by the human ASRS administrator.

Taxonomy of Factors and Anomalies

Using the data garnered from the frequency analyses, text mining, and clustering, a taxonomy was created linking associations between Factors and Anomalies. The taxonomy comprises the differing anomaly types, with contributing factors located hierarchically below them. The taxonomy is presented in parts for clarity.

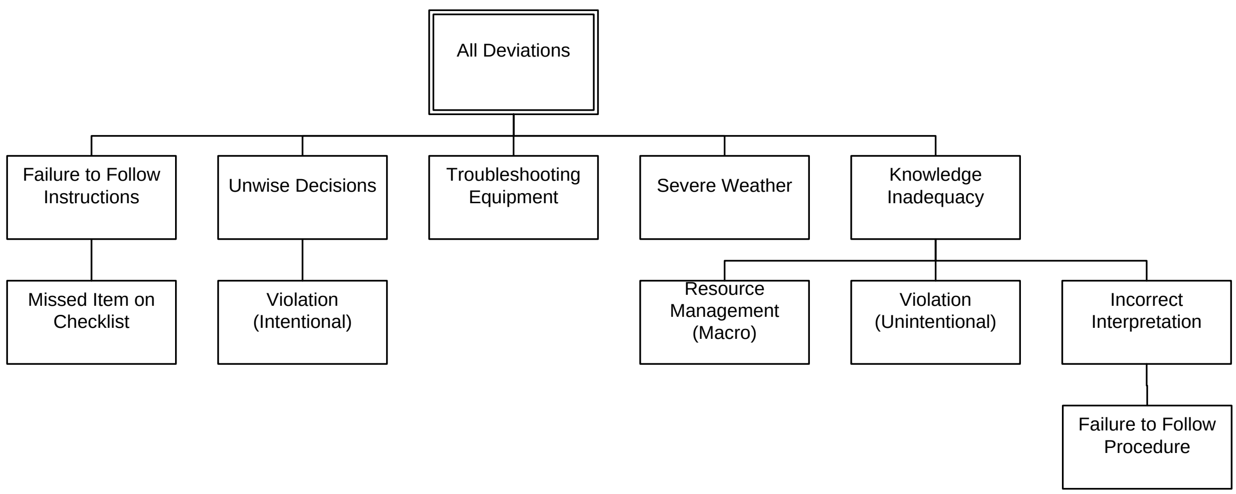


Figure 53: Taxonomy of deviation anomalies

The above figure describes factors associated with deviations, including Procedural, Track / Heading, and Altitude Deviations. Weather was often a factor that would cause the automation to shut off or otherwise induce an intervention that deviated from protocol or assigned flight path. Many anomaly reports suggested that crew members would not follow recommendations from weather briefers, ATC, or other crew members. Missing an item on a checklist would lead to a failure of following established procedures. Aircraft equipment problems were often responsible for procedural deviations, often requiring human intervention in cases of automation failure. Finally, inadequate knowledge of procedures or operations would lead to Deviations, as well as unintentional violations, i.e., situations where a pilot violated an established regulation without conscious knowledge. An example of this would be an unintentional airspace violation. Incorrect interpretations of procedures or communications would also lead to deviation anomalies.

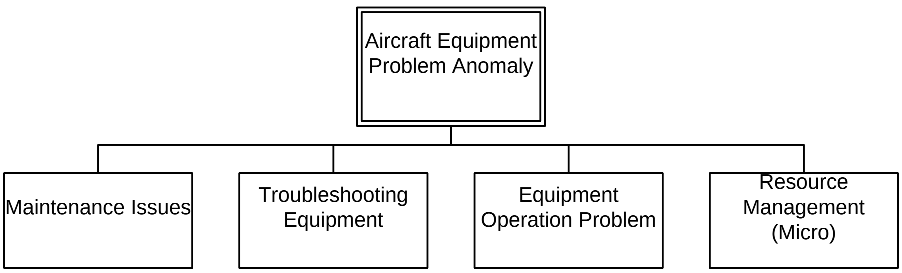


Figure 54: Taxonomy of aircraft equipment problem anomalies

Aircraft Equipment anomalies were the result of equipment malfunctions or failures. Another contributor to anomalies included improper or lack of maintenance. Many anomaly reports contained statements about pilots troubleshooting equipment that had failed or was otherwise inoperative. Equipment operation problems such as lack of knowledge or employing the wrong function also caused anomalies. One anomaly report contained an anecdote in which removing a jacket inadvertently triggered a switch. The last factor, resource management on a micro level, described situations in which pilots were unable to operate aircraft equipment quickly enough to keep up with the rapidly changing situation. The rate of actions or operations performed by the crew (Actions per Minute) was insufficient for a given situation.

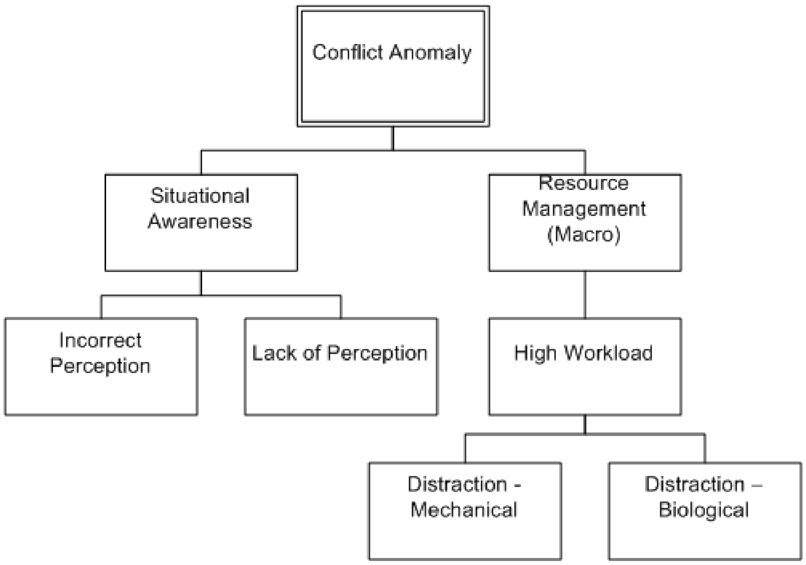


Figure 55: Taxonomy of conflict anomalies

The most frequently cited factor for Conflict anomalies was loss of situational awareness (SA). This loss of SA was blamed on either focusing on irrelevant information, erroneous perception, or not perceiving important aspects of the flight environment at all. Resource management is management of the aircraft environment, or "staying ahead of the airplane" in pilot's terminology. This inability to properly manage and interact with all resources was attributed to high workloads, which could come from distractions such as equipment failures, personal issues, or interactions with other crew members.

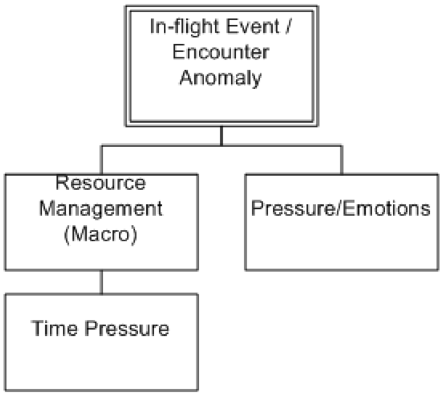


Figure 56: Taxonomy of inflight event / encounter anomalies

Inflight Event anomalies were associated with an inability to manage the cockpit environment. Workload and time pressure were cited as factors. In addition, many Inflight Event / Encounter anomalies were caused by unruly passengers with highly excited mental states. For this reason, pressure and emotions were identified as factors.

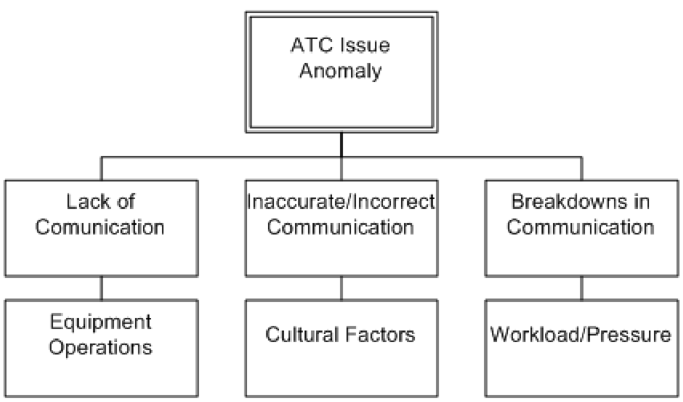


Figure 57: Taxonomy of ATC issue anomalies

The final anomaly type, ATC Issue anomalies, was often caused by breakdowns in communication or a complete lack of communication. In the General Aviation environment, this lack of communication was often the result of incorrect frequencies being used. Other factors included language barriers or high-traffic areas with many pilots attempting to communicate at once.

CHAPTER FIVE: CONCLUSION

A study of factors influencing the submission of anomalous pilot reports was conducted. The study was unique because it investigated the ASRS database, to investigate latent, active, and human factors contributions to incidents, not necessarily accidents. By examining a great number of records (127,776), new associations and interpretations were made.

The study employed traditional statistical analyses, correspondence analysis, text mining, clustering, and dimensionality reduction activities to investigate what factors contributed to civil aviation anomalies.

Major Findings

The study uncovered results that were similar to those of most aviation accident studies (Wiegmann & Shappell 2003, Edwards 1988, Wallace 2006). A vast majority of aircraft anomalies have been shown to be attributable to human error or human factors issues. However, these findings were narrowed and made much more specific, further illustrated by frequencies and distributions. New associations were uncovered that described which types of factors and anomalies typically occurred together, and the strengths of these associations were also found. Supervised and unsupervised data mining methods were also tested. The data supported the idea that concept and keyword extraction can be used to create predictive models, thereby automating future information extraction. The order of quality of results was greatest for the supervised analysis methods, then the semi-supervised, and lastly the unsupervised methods. Finally, it was demonstrated that dimensionality reduction techniques can aid in classifying documents, as well as identifying minimum levels of fidelity in data required for successful classification. The taxonomy created from study findings described relationships between factors and anomalies.

Active and Latent Factor Roles in Anomalies

Using keywords and concepts found in Wiegmann & Shappell (2003), categories of contributors to aircraft anomalies were created. These categories comprised leading causes or factors identified by the literature. The anomaly record corpus of about 127,000 records was mined, and counts of these categories were extracted. Words that were extracted but not belonging to any of the defined categories were placed into an "Uncategorized" category.

The most prevalent category found across all anomalies was the Unsafe Conditions category. Rule-based error categories were found more frequently than skill- or knowledge-based error categories. This contradicts past aircraft accident studies (Wiegmann & Shappell, 2001), whether civilian or military, that employed the HFACS system; they found skill-based errors to be most prevalent. However, the present study was one that investigated anomalies, which may have conditions similar to those had by accidents – though there are enough differences between the two to not constitute an accident. Perceptual errors were also not identified as frequently in the present study they were in accident investigations.

Another interesting difference uncovered by the present study is that only four percent of records were identified as having elements that would constitute a violation. The studies by

Wiegmann & Shappell (2001) found that about 25 percent of aviation accidents contained some sort of violation. It is conceivable that, despite the ASRS' assurances to pilots that their information is confidential and not traceable, many pilots are still hesitant to report their own violation activities.

The low occurrence of unsafe supervision is consistent with the Wiegmann & Shappell (2001) study, which also found low percentages of unsafe supervision in aviation accidents. Identifying unsafe supervision is difficult, as assigning blame to an entire organization rather than an individual has much larger consequences for aviation-related operations, especially if serious violations or oversights are found.

Anecdotal evidence from Jack Hessburg during his presentations on running a fictional airline maintenance operation includes humorous accounts of organizations avoiding blame. Most airlines require blame to be assigned when an aircraft is delayed, be it mechanical, crew-related, etc. Stories exist in which mechanics, who are often the culprits for a flight delay or cancellation, will go to extraordinary lengths to assign blame to another organization. Airline line mechanics have been known to bribe van drivers to strand aircraft crew at the hotel or to break off keys to the catering truck, giving them more time to fix the problem and to artificially create a "crew delay," or "catering delay," respectively (Hessburg, 2000). Data collection efforts such as those of the ASRS or the Boeing MEDA (Rankin 1996) offer an anonymous reporting system to focus on fixing problems or reducing contributing factors within the aviation system, rather than assigning blame.

As compared to the other types of Human Factors category issues, perceptual errors were not encountered frequently in this dataset. A probable explanation is that significant research and technology improvements have attempted to overcome human limitations in the cockpit. Weather radar, strict regulations on avoiding areas of decreased perceptual information (clouds, darkness, etc.), and various aids and instruments all strive to decrease or eliminate perceptual errors (Leibowitz, 1988). Humans have become information integrators and processors within the cockpit, especially during commercial airliner operations (Mosier, 2010). The modern airliner has all but done away with "fly by the seat of your pants" flying or creating situations that require immediate action or feedback. Such information processing activities do not require fast reflexes or immediate responses where perceptual errors may occur due to time pressure.

Discussion of Latent and Active Factors

Many of the document associations discovered in the category web activity reinforced the associations found in the literature, in the correspondence analysis activity, and in the cross tabulation analysis. The categories differed from the factor and anomaly classifications studies because they were slightly less rigid, and did not require a human to classify different data. These were "semi-supervised," as a human analyst specified the terms to search for in the record corpus.

Skill-Based and Rule-Based category types were associated with each other, and less so with Knowledge-Based categories. This is supported by Reason's (1990) explanation that Skill-Based and Rule-Based errors are similar in that they share a control mode that is does not exist for Knowledge-Based errors. Rasmussen (1986) defines Skill-Based and Rule-Based errors as having a feed-forward control element that is based on familiar and ingrained internal models of the environment. The category web supports the theory that Skill-Based and Rule-Based Errors are more similar to each other than they are to Knowledge-Based errors.

The Knowledge-Based category web suggests that they occur on their own, when the pilot is in unfamiliar or unknown circumstances. They are highly associated with Unsafe Conditions, and less so with Weather and Violations. This association supports feedback theories of Knowledge-Based errors; where the individual encountering the error must employ slower, more intensive cognitive processing to solve the problem (Reason 1990). It is highly likely that these errors would be encountered in situations where troubleshooting activities were also present.

Perceptual Errors were associated with Unsafe Conditions and Weather. These two categories are likely interrelated, as Weather can lead to Unsafe Conditions. Somewhat surprisingly, associations were strong with Rule-Based errors. This may be because perceptual errors can fool a pilot into taking incorrect actions, or making wrong decisions due to inaccurate or misrepresented information. An inaccurate reading of a heading or altitude indicator could lead a pilot to take an unnecessary action, albeit one that is deemed correct by higher level training and ingrained behaviors.

There were strong associations between Unsafe Supervision and Violations, but these two categories seem intuitively related, and no novel associations are uncovered in this particular category web.

The Violations category had strong associations with Unsafe Conditions and Rule-Based Errors. These associations were reinforced because Violations usually cause unsafe conditions, and also require a decision to commit them.

Aircraft Issues were associated with Unsafe Conditions, as equipment malfunctions often lead to unsafe conditions. The other associations were too weak and trivial to mention.

Weather was associated with Unsafe Conditions and Rule-Based Errors. Often in general aviation, a conscious decision is made to proceed into inclement weather, despite training and repeated instructions and notices to pilots to avoid such conditions. There is often pressure to maintain the flight plan or to make a landing to keep a schedule. Rapidly changing weather conditions can cause a sudden loss of visibility, decrease aircraft performance, or otherwise lead to Unsafe Conditions.

Dimensionality Reduction findings in Identifying Important Concepts

The diffusion maps operation was in agreement with pre-established categories of Anomaly reports 23.8 percent of the time. This classification scheme operated on completely unsupervised data.

Unsupervised diffusion map operation on 100 words

Dimensionality reduction was performed using the Diffusion Maps algorithm within code freely distributed by Ann Lee (2010). The algorithm demonstrated a limited ability to classify records simply based by word frequencies. This operation can be greatly improved with appropriate selection of words, as their meanings were not considered for the extraction process. This method was completely unsupervised.

Study Limitations

The study limited the number data sources to a single large repository, the NASA Aviation Safety Reporting System. This repository contained only voluntary reports submitted by pilots flying both private and commercial operations. No military data was included in this dataset. The database was chosen because the literature suggested that, aside from NASA's internal metrics and monthly publication summarizing relevant, current anomalies with commentary, no large-scale analysis had been conducted on the data.

The encoded nature of the ASRS database, with its many abbreviations, excluded the possibility of a context-link analysis using a traditional dictionary. The data would have had to be decoded; this was deemed unfeasible due to the number of encoded terms. In addition, much of aviation terminology is rife with acronyms, abbreviations, and non-standard technical terms. A context analysis, though very powerful, is usually limited to full-text sources such as web pages and interview or survey data.

In order to provide the widest breadth of data for the present study, all available anomaly data was analyzed. The data was not partitioned by any means, such as by date, type of operation, or type of aircraft. It is possible and very likely that there are underlying patterns within sections of the data. Anomaly types and frequencies for large airline operations, for example, will be different than they are for student pilots or recreational flights. These types of underlying patterns were not discovered. Future studies might investigate whether anomaly types and frequencies changed over time, as this was deemed beyond the scope of the present study. The overarching research objective was to have as much data as possible for model creation, text mining, clustering, and dimensionality reduction activities.

To reduce noise by eliminating infrequently occurring anomalies, not all anomaly types were analyzed. The top seven most frequently occurring anomalies were studied. The other anomaly data were discarded. There may have been additional insights or categorizations made based on these additional anomaly types. The decision to discard these data was due to the fact that these additional anomaly types added noise, because few records were associated with these less frequently occurring anomaly types, and because the ability to build predictive models was thereby drastically reduced. To account for this, no factor data was discarded. All factors were kept, no matter how low their frequency of occurrence was.

The study relied heavily on the use of SPSS r17 and IBM PASW Modeler 13. These two software packages are made to work with each other. However, any inherent flaws or biases in these two applications would be sustained throughout most of the analysis. Results were not compared to those from another package.

The diffusion maps aspect of the study did not have an official, well-developed software application. Diffusion maps for document clustering are a relatively novel idea at the time of this study; most other studies that classified documents had fewer numbers of larger documents containing full text. Of those studies, the classifications were usually binary in nature. Other studies also reported limited success in classifying documents with diffusion maps; e.g., Underhill (2007) reported a 27 percent accuracy rate. The rate reported in this study was 23.8 percent using a completely unsupervised input dataset. Records from the ASRS dataset usually had multiple classifications, and it is entirely possible that the unsupervised classification algorithm focused on a different feature or identified a feature that the human classifier did not.

The techniques used to create the document difference matrices were not exceedingly sophisticated. Dimensionality reduction performs best with large samples of data, while the method used to handle the vectors of word presences was insufficient. Microsoft Excel was used for the task, and the software's limits were quickly reached when attempting to create a matrix sufficient to encompass the 100 columns of words and 127,776 rows of text records. A reduced dataset of 1000 records was used in its stead.

Future Studies

Future studies in this area might investigate partitioned data to identify additional opportunities for insight that could be gleaned from focusing on specific areas or eras of aviation. Other databases, ones not limited to aviation, could also be used. Many organizations employ feedback systems and surveys, and these domain-specific text records could be analyzed in a similar fashion.

Other dimensionality reduction techniques could also be employed. There are many techniques, both linear and nonlinear, that can be applied once data is properly encoded. One technique often cited in the literature was Principle Component Analysis PCA; it was not employed in this study. Additionally, the method of encoding anomaly reports could be altered to generate better results, or else a more sophisticated or better-equipped tool for dimensionality reduction could be employed. A software package that could encode raw text, apply the appropriate dimensionality reduction, and suggest or perform appropriate analyses on the output would be highly beneficial to this and other domains employing Literature Based Discovery (LBD).

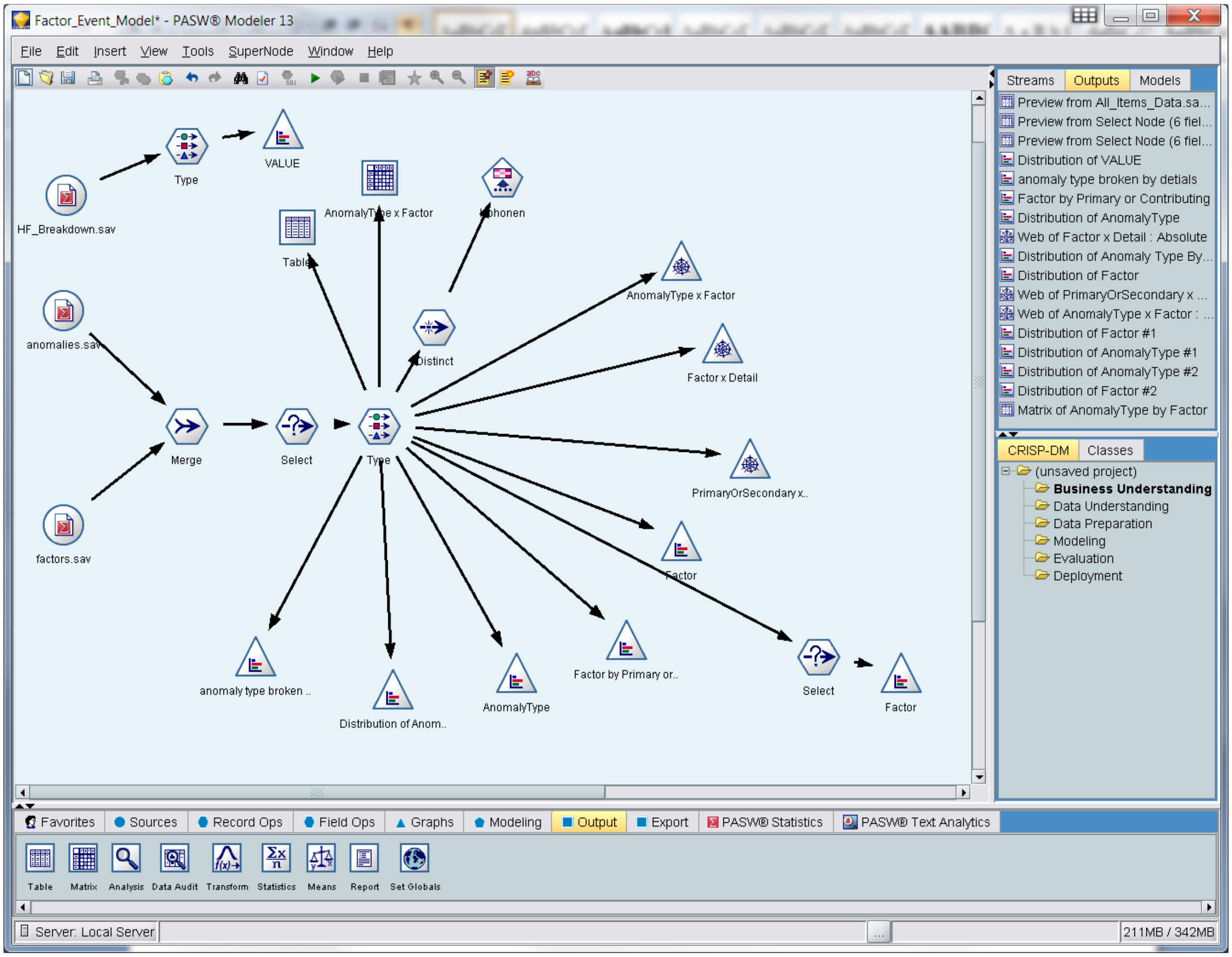
This study dealt with highly encoded data. To perform contextual analysis, the data would have to be decoded or an appropriate dictionary created. Thoroughly decoding the data is likely to be a better alternative. Contextual link analysis could explore relationships between meanings within anomaly reports, thereby revealing additional insight not uncovered by database administrators or maintainers.

Finally, this study employed unsupervised, semi-supervised, and supervised approaches. The approaches that provided the most meaningful results were those that were supervised by a human analyst. Future studies could develop wholly unsupervised methods and techniques. The method of performing keyword extraction was highly rudimentary, especially for the dimensionality reduction element of the study. Future studies may perform more in-depth and comprehensive keyword extraction, in order to create the most accurate possible relationships and associations, as well as providing a richer feature set for the dimensionality reduction activity to act upon.

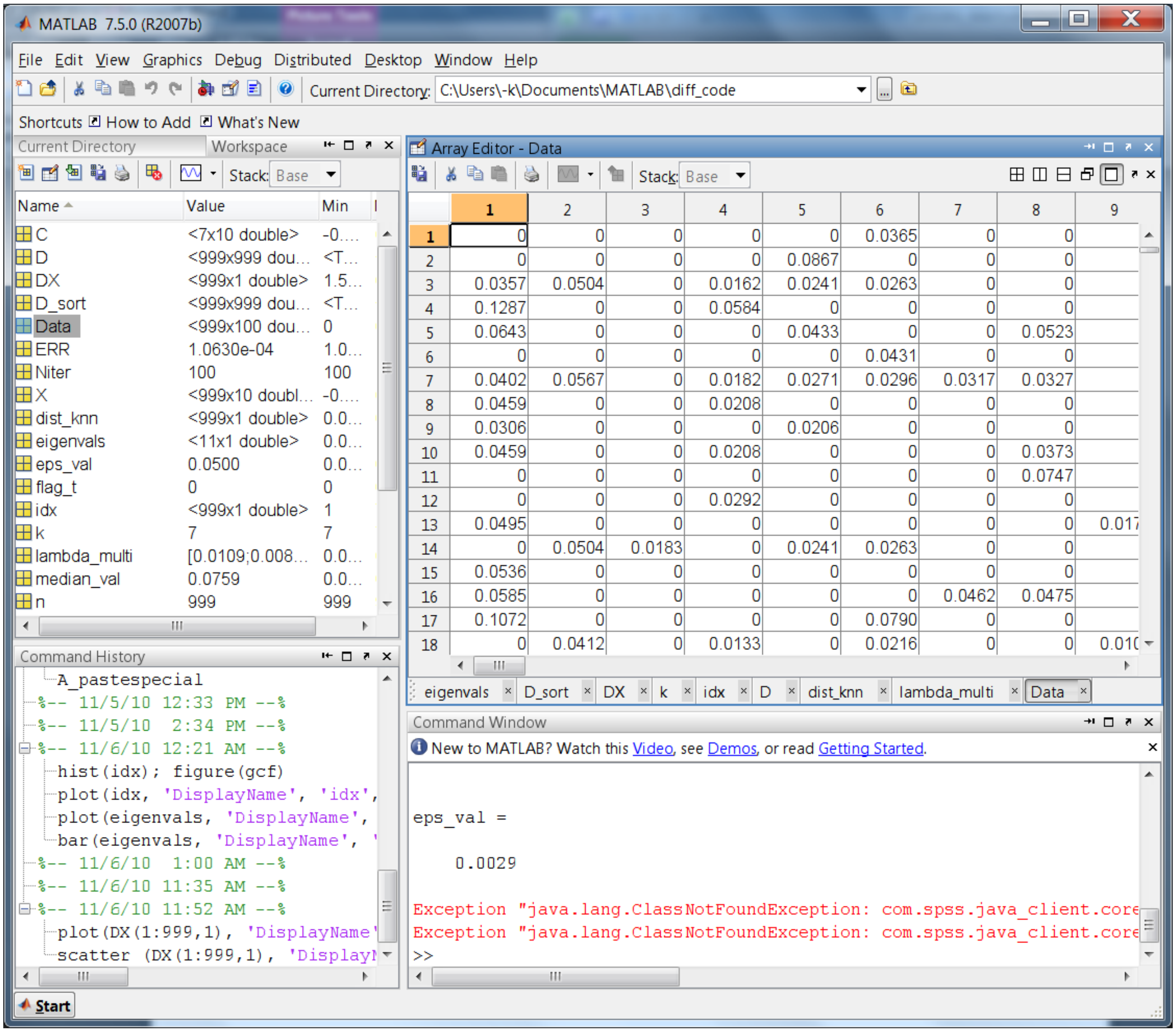
Final Thoughts

Although the possibility of achieving a perfect safety record and zero percent accident rates for the aviation industry is highly unlikely, it is necessary to reduce the current accident rate to accommodate the up and coming drastic increases in worldwide air traffic. The aviation market demands steady increases in performance and safety, and only through diverse, multidisciplinary, and systematic employment of findings from studies such as these will this demand be realized. Sheridan (2010) describes that changes to engineering itself are necessary; claiming that engineers are often focused on designing *things*, when really their focus should be designing *relationships to people*. It is this kind of thinking that often results in the "wild, unpredictable, *organic*" human to be blamed for incidents rather than the orderly and digital technology. Understanding that the human can also be the savior of the system and designing with the intention of exploiting these "saving features" should be the focus of design, training, and research. Only through collaboration across disciplines and integration of meaningful findings can powerful, self-correcting, and sustainable practices emerge that will guide the aviation field into its exciting future.

APPENDIX A: MISCELLANEOUS FIGURES



Modeling Software Screenshot



MATLAB Software Screenshot

ACN: 868116 *(1 of 43)*

Time / Day

Date : 201001
Local Time Of Day : 0001-0600

Place

Locale Reference.ATC Facility : ZZZ.TRACON
State Reference : US
Altitude.MSL.Single Value : 10000

Environment

Flight Conditions : IMC
Weather Elements / Visibility : Icing
Light : Daylight

Aircraft

Reference : X
ATC / Advisory.TRACON : ZZZ
Aircraft Operator : Air Carrier
Make Model Name : PC-12
Crew Size.Number Of Crew : 2
Flight Plan : IFR
Mission : Passenger
Flight Phase : Descent
Airspace.Class E : ZZZ

Component

Aircraft Component : Cockpit Window
Aircraft Reference : X
Problem : Failed

Person : 1

Reference : 1
Location Of Person.Aircraft : X
Location In Aircraft : Flight Deck
Function.Flight Crew : Pilot Flying
Function.Flight Crew : Captain
Qualification.Flight Crew : Commercial
ASRS Report Number.Accession Number : 868116
Human Factors : Training / Qualification
Human Factors : Distraction
Human Factors : Human-Machine Interface

Person : 2

Reference : 2
Location Of Person.Aircraft : X
Location In Aircraft : Flight Deck
Reporter Organization : Air Carrier
Function.Flight Crew : Pilot Not Flying
Function.Flight Crew : First Officer
Qualification.Flight Crew : Commercial
ASRS Report Number.Accession Number : 868117
Human Factors : Human-Machine Interface
Human Factors : Distraction
Human Factors : Training / Qualification

Events

Anomaly.Aircraft Equipment Problem : Critical
Anomaly.Inflight Event / Encounter : Object
Detector.Person : Flight Crew
Were Passengers Involved In Event : N
When Detected : In-flight
Result.General : Declared Emergency
Result.General : Maintenance Action
Result.General : Police / Security Involved
Result.Flight Crew : Overcame Equipment Problem
Result.Aircraft : Aircraft Damaged

Assessments

Contributing Factors / Situations : Weather
Contributing Factors / Situations : Human Factors
Contributing Factors / Situations : Aircraft
Primary Problem : Aircraft

Typical ASRS Record

Narrative: 1

On descent, leveling off at 10,000 feet there was a loud thunk/crack sound. The left hand windscreen was instantly spider webbed with cracks emanating from a central impact point in the lower left center, spreading to all sides of the screen. I immediately disconnected the autopilot and began to slow and level the aircraft. After resuming manual control of the aircraft, I transferred control of the aircraft to the First Officer, using a positive exchange of flight controls. Once established, I donned my oxygen mask as a precautionary measure affording me both oxygen in the event of depressurization and additional physical protection for my face. I also zipped up my **jacket** and lowered my seat. When I had accomplished these tasks, I resumed flying the aircraft by hand, and asked the First Officer to make a PA to the passengers, stating that we had sustained damage to the windscreen, there was no need to don their oxygen masks and we would be making a normal descent and landing. I then declared an emergency. I requested a lower altitude and notified them we would be slowing to a slower airspeed. They repeated the restriction of 10,000 Ft and promised lower soon, speed was our discretion. During the descent the First Officer and I discussed contingency plans if the screen were to fail (as it was continuing to crack), primarily focusing on the fact that he would take control of the aircraft and make the landing. We closed the curtain partially to remove the damage from passenger view and partially to provide protection to the cabin occupants in the event of total structural failure of the glass/plastic. We also notified operations, requesting they advise dispatch and notify maintenance. The remainder of the descent was conducted via normal procedures at a reduced airspeed. When we were unable to acquire the airport visually, we were climbed to 2200 FT and cleared for the ILS. We landed without further incident and taxied into parking, followed by both airport operations and fire/rescue as per their procedures (we did not request additional services upon landing). Upon closer inspection of the damage on the ground it appears as though the outer layer of the windscreen cracked, but no moisture or damage penetrated to the inner layer. No loss of pressurization occurred. No passenger injury or damage to persons/property on the ground occurred to our knowledge. No residue of a bird, nor additional damage was found during a thorough post flight of the aircraft. Due to lack of remains/residue, I suspect ice or foreign object rather than a bird.

Synopsis

When the outer pane of the Captain's windshield on their PC-12 shattered the flight crew declared an emergency and landed at their destination airport.

Date : 200902
Local Time Of Day : 0601-1200

Place

Locale Reference.Airport : MIA.Airport
State Reference : FL
Relative Position.Distance.Nautical Miles : 0
Altitude.AGL.Single Value : 0

Person

Reference : 1
Location Of Person : Company
Reporter Organization : Air Carrier
Qualification.Flight Attendant : Current
ASRS Report Number.Accession Number : 823441

Events

Anomaly.Other
Detector.Person : Other Person
Result.General : None Reported / Taken

Assessments

Contributing Factors / Situations : Human Factors
Contributing Factors / Situations : Environment - Non Weather Related
Primary Problem : Ambiguous

Narrative: 1

In December 2008, I had my uniform pants and a navy blue vest and 2 white shirts to be dry-cleaned. In January, when I returned from my vacation to pick-up my uniform, I gave the clerk the two pink receipts, but the clerk only returned my shirts and informed me that the pants and the vest were returned and discarded the receipts. I advised him that was not possible, but just to be sure I returned home to check my closet, my car trunk and any other location that my uniform could be, but for the last 15 years that I do business with them my dry-cleaned uniform is picked up on the return of my working trips. The owner of the establishment did show me that the pants and the vest were retrieved on the same day according to the computer entries that he did check out to me, so it's his word against my word and the thing that is not normal in this situation is that to pick up any order without the pink receipt I have to sign a release book. Today my pants, tomorrow could be my **jacket** and soon someone could have an entire uniform...please someone could you call this establishment to check this situation? I truly believe that this is not an isolated occurrence.

Synopsis

Cabin Attendant discovers that uniform pants and vest left at dry cleaners were picked up, according to the owner, but not by the Reporter.

APPENDIX B: CATEGORY RESOURCES

Term	Match	Type
intxn	Entire Term	Violation
deviated	Entire Term	Violation
drunk	Entire Term	Violation
exceeded	Entire Term	Violation
inebriated	Entire Term	Violation
over the legal limit	Entire Term	Violation
over the limit	Entire Term	Violation
overspeed	Entire Term	Violation
too fast	Entire Term	Violation
too low	Entire Term	Violation
violated	Entire Term	Violation
violation	Entire Term	Violation
error	Entire Term	Skill-Based
forgot	Entire Term	Skill-Based
oversight	Entire Term	Skill-Based
pulled the wrong	Entire Term	Skill-Based
pushed the wrong	Entire Term	Skill-Based
slip	Entire Term	Skill-Based
the incorrect button	Entire Term	Skill-Based
the incorrect switch	Entire Term	Skill-Based
turn	Entire Term	Skill-Based
wrong button	Entire Term	Skill-Based
wrong switch	Entire Term	Skill-Based
crm	Entire Term	Rule-Based
indications	Entire Term	Rule-Based
clred	Entire Term	Rule-Based
confusion	Entire Term	Rule-Based
decision	Entire Term	Rule-Based
i was sure	Entire Term	Rule-Based
mistake	Entire Term	Rule-Based
under the impression	Entire Term	Rule-Based
we were sure	Entire Term	Rule-Based

wrong idea	Entire Term	Rule-Based
wrong impression	Entire Term	Rule-Based
wrong plan	Entire Term	Rule-Based
manual	Entire Term	Knowledge-Based
info	Entire Term	Knowledge-Based
supplemental info	Entire Term	Knowledge-Based
attempt	Entire Term	Knowledge-Based
question	Entire Term	Knowledge-Based
student	Entire Term	Knowledge-Based
did not know	Entire Term	Knowledge-Based
no knowledge of	Entire Term	Knowledge-Based
was not familiar	Entire Term	Knowledge-Based
was unfamiliar	Entire Term	Knowledge-Based
were not familiar	Entire Term	Knowledge-Based
were unfamiliar	Entire Term	Knowledge-Based
were unsure	Entire Term	Knowledge-Based
cruising	Entire Term	Routine Behaviors
moderate	Entire Term	Routine Behaviors
instructions	Entire Term	Routine Behaviors
vfr conditions	Entire Term	Routine Behaviors
taxi	Entire Term	Routine Behaviors
clr of class	Entire Term	Routine Behaviors
level	Entire Term	Routine Behaviors
freq change	Entire Term	Routine Behaviors
procs	Entire Term	Routine Behaviors
contact	Entire Term	Routine Behaviors
response	Entire Term	Routine Behaviors
request	Entire Term	Routine Behaviors
uneventful	Entire Term	Routine Behaviors
inspection	Entire Term	Routine Behaviors
clbing	Entire Term	Routine Behaviors
talking	Entire Term	Routine Behaviors
event	Entire Term	Routine Behaviors
visual apch callback	Entire Term	Routine Behaviors
conversation	Entire Term	Routine Behaviors
apch	Entire Term	Routine Behaviors
appropriate freq	Entire Term	Routine Behaviors
arpt tfc pattern	Entire Term	Routine Behaviors

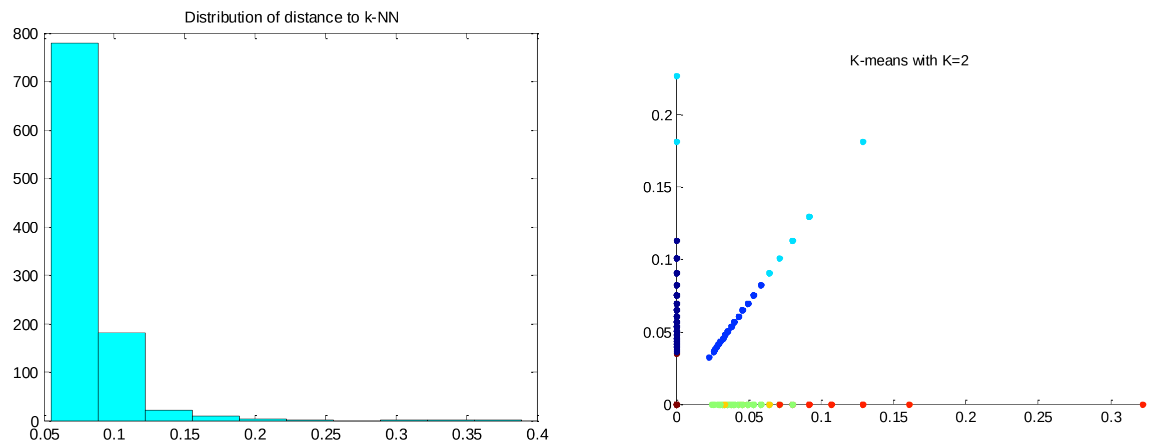
assigned rte	Entire Term	Routine Behaviors
chked	Entire Term	Routine Behaviors
clb	Entire Term	Routine Behaviors
clbed	Entire Term	Routine Behaviors
dsnded	Entire Term	Routine Behaviors
dsnding	Entire Term	Routine Behaviors
flying	Entire Term	Routine Behaviors
lndg	Entire Term	Routine Behaviors
missed apch	Entire Term	Non-Routine Behaviors
assistance	Entire Term	Non-Routine Behaviors
strong odor	Entire Term	Non-Routine Behaviors
hit	Entire Term	Non-Routine Behaviors
burning	Entire Term	Non-Routine Behaviors
injuries	Entire Term	Non-Routine Behaviors
medical	Entire Term	Non-Routine Behaviors
smoking	Entire Term	Non-Routine Behaviors
evac	Entire Term	Non-Routine Behaviors
paramedics	Entire Term	Non-Routine Behaviors
prob	Entire Term	Non-Routine Behaviors
odor	Entire Term	Non-Routine Behaviors
smell	Entire Term	Non-Routine Behaviors
smoke	Entire Term	Non-Routine Behaviors
conflict alert	Entire Term	Non-Routine Behaviors
dev	Entire Term	Non-Routine Behaviors
low fuel	Entire Term	Non-Routine Behaviors
oxygen masks	Entire Term	Non-Routine Behaviors
warning	Entire Term	Non-Routine Behaviors
return	Entire Term	Non-Routine Behaviors
damage	Entire Term	Non-Routine Behaviors
alert horn	Entire Term	Non-Routine Behaviors
alternate fuel	Entire Term	Non-Routine Behaviors
emer	Entire Term	Non-Routine Behaviors
evasive action	Entire Term	Non-Routine Behaviors
incident	Entire Term	Non-Routine Behaviors
problem	Entire Term	Non-Routine Behaviors
snow	Entire Term	Weather
icing	Entire Term	Weather
imc	Entire Term	Weather
low visibility	Entire Term	Weather

wx	Entire Term	Weather
turb	Entire Term	Weather
ifr	Entire Term	Weather
sun	Entire Term	Weather
visibility	Entire Term	Weather
ice	Entire Term	Weather
rain	Entire Term	Weather
wind	Entire Term	Weather
cloud	Entire Term	Weather

APPENDIX C: DIFFUSION MAPS OUTPUT DATA

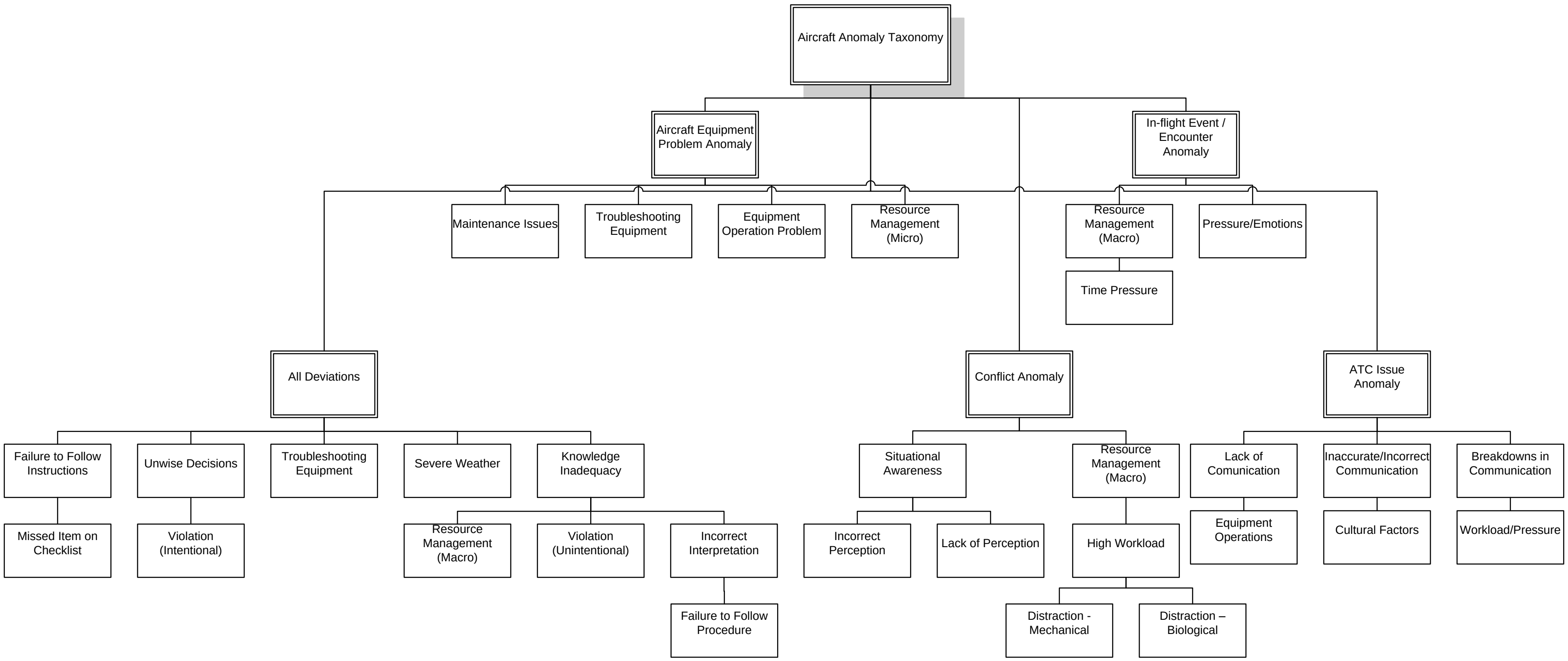
1
0.010735
0.008455
0.005728
0.004123
0.003394
0.003138
0.002874
0.002692
0.002481
0.002262

Eigenvalues from DR reduction



Other Dimensionality Reduction Outputs – distances to *k* and clustering plot

APPENDIX D: COMPLETE TAXONOMY



REFERENCES

Ahmad, A., and Dey, L. (2007). A k-mean clustering algorithm for mixed numeric and categorical data. Data & Knowledge Engineering. Vol 63. 2007. pp 503-527.

Alpaydin, E. (2004). Introduction to machine learning. The MIT Press. Cambridge, MA.

Aviation Safety Reporting System. (2002). An Analysis of ASRS Maintenance Incidents. Mountain View, CA.

Bah, B. (2008). Diffusion Maps: Analysis and Applications. Master's Thesis submitted to University of Oxford.

Bird, F.E., and Germain, G.L. (1996). Practical loss control leadership. Revised Ed. Calgary, Alberta: Det Norske Veritas.

Boquet, A., Detwiler, C., Roberts, C., Jack, D., Shappell, S. A., & Wiegmann, D. A. (2004). General aviation maintenance accidents: An analysis using HFACS and focus groups. Office of the Chief Scientist for Human Factors Aviation Maintenance Human Factors 4-8.

Borooah, P. V. K. (2001). Logit and Probit: Ordered and Multinomial Models (1st ed.). Sage Publications, Inc.

Clausen, S. (1998). Applied correspondence analysis. Sage. Thousand Oaks, CA.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement Vol.20, No.1, pp. 37–46.

Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F., & Zucker, S. W. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. Proceedings of the National Academy of Sciences of the United States of America, 102(21), 7426 -7431. doi:10.1073/pnas.0500334102

Concept map - Wikipedia, the free encyclopedia. (2010). Retrieved September 28, 2010, from http://en.wikipedia.org/wiki/Concept_map

Cox D. (1957). Note on grouping. Journal of American Statistical Association. Vol. 52, pp. 543–547.

Dekker, S.W.A. (2005). Ten questions about human error. CRC Press, Taylor and Francis Group. Boca Raton, FL.

Dennett, D. (1987). Cognitive wheels: the frame problem of AI. In Z. Pylyshyn (ed.), The Robot's Dilemma: the Frame Problem in Artificial Intelligence. Ablex. Norwood, NJ.

Durso, F.T., & Alexander, A.L. (2010). Managing workload, performance, and situational awareness in aviation systems. In E. Sala and D. Maurino (Eds.), Human factors in aviation. Academic Press. San Diego, CA.

Edwards, E.E. (1988). Human factors in aviation. Academic Press, Inc, San Diego, CA.

Eisen, L. A., & Savel, R. H. (2009). What went right: lessons for the intensivist from the crew of US Airways Flight 1549. Chest, 136(3), 910-917. doi:10.1378/chest.09-0377

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic situations. Human Factors, 37, 32-64.

Engle. E., & Lott, A.S. (1979). Man in flight: biomedical achievements in aerospace. Annapolis, MD: Leeward.

Fleishman, E. A., & Quaintance, M. K. (1984). Taxonomies of human performance. San Diego, CA: Academic Press.

Franke, J., Nakhaeizedeh, G., and Renz, I. (2003) Text mining: theoretical aspects and applications. Physica-Verlag Hedelberg, New York, NY.

Fukunaga, K. (1990). Introduction to statistical pattern recognition. Academic Press Professional, Inc., San Diego, CA, USA, 1990.

Lafon, S., & Lee, A. B. (2006). Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(9), 1393–1403.

Lautman, L.G. and Gallimore, P.L. (1987). Control of the crew caused accident. Boeing Airliner, April-June, 1-6.

Garland, D.J., Wise, J.A., Hopkin, D.V. (1999). Handbook of Aviation Human Factors. Mahwah, NJ: Lawrence Erlbaum Associates.

Gibson, D., Kleinberg, J., Raghavan P. (1998) Clustering Categorical Data: An Approach Based on Dynamical Systems. In Proceedings of the 24th International Conference on Very Large Data Bases, (VLDB), New York, NY, USA, August 1998, Morgan Kaufmann, pp. 311-322.

Guo, B., Perron, B.E., Gillespie, D.F. (2009). A systematic review of structural equation modeling in social work research. British Journal of Social Work 39, 1556-1574. 2009.

Hartigan, J.A., and Wong, M.A. (1979). A k-means clustering algorithm. Journal of the Royal Statistical Society: Vol. 28

Hawkins, S.A., and Hastie, R. (1990). Hindsight: biased judgments of past events after the outcomes are known. Psychological Bulletin 107: 311-327.

Hessburg, J. (2000). Air Carrier MRO Handbook (1st ed.). McGraw-Hill Professional.

Higgs, B., Weller, J., & Solka, J. (2006). Spectral embedding finds meaningful (relevant) structure in image and microarray data. BMC Bioinformatics, 7(1), 74. doi:10.1186/1471-2105-7-74

Hobbs, A., & Williamson, A. (2003). Associations between Errors and Contributing Factors in Aircraft Maintenance. Human Factors: The Journal of the Human Factors and Ergonomics Society, 45(2), 186-201. doi:10.1518/hfes.45.2.186.27244

Hosmer, D.W., and Lemeshow, S. (2000). Applied logistic regression 2nd edition. Chichester, Wiley, New York, NY

Huang, Z. (1997). Clustering large data sets with mixed numeric and categorical values. Proceedings of the First Pacific Asia Knowledge Discovery and Data Mining Conference, Singapore: World Scientific, pp. 21–34.

Huang Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining Knowledge Discovery. Vol. 2, No. 2, pp. 283–304.

Jarvis, S., and Harris, D. (2010). Development of a bespoke human factors taxonomy for gliding accident analysis and its revelations about highly inexperienced UK glider pilots. Ergonomics Vol. 53, No. 2, February 2010, 294–303

Kantardzie, M. (2003). Data Mining: Concepts, Models, Methods, and Algorithms, by Mehmed. IEEE Press, New York, NY, 2003

Kapperud, G., Skjerve, E., Vik, L., Hauge, K., Lysaker, A., Allmen, I., Ostroff, S.M., Potter, M. (1993). Epidemiological investigation of risk factors for campylobacter colonization in Norwegian broiler flocks. Epidemiological Infection 1993. Vol 111, pp 245-255.

Keim, D.A. (2002). Information visualization and visual data mining. IEEE Transactions On Visualization And Computer Graphics, Vol. 8, No. 1, January-March 2002

Khalid, H.M., Rahda, J.K, Yang, X., Helander, M.G. (2010) Analyzing ABC of disaster experience using text mining. In Proceedings of the Applied Human Factors and Ergonomic (AHFE) Conference, 2010.

Kohonen, T. (1998). The self-organizing map. Neurocomputing, 21, 1-6. 1998

Kohonen, T., and Honkela, T. (2007). Kohonen network. Scholarpedia, 2(1):1568, 2007.

Koonce, J.M. (1999). A historical overview of human factors in aviation. Handbook of Aviation Human Factors. Mahwah, New Jersey, 1999.

Lee, A.B. (2010). MATLAB Code Page. Retrieved November 6, 2010, from http://www.stat.cmu.edu/~annlee/software.htm

Leibowitz, H.W. (1988). Human senses in flight. In: Human factors in aviation. pp 83-110. Academic Press, San Diego, CA

MacQueen J.B. (1967). Some methods for classification and analysis of multivariate observations.—Proceedings of the Fifth Symposium on Mathematical Statistics and Probability, Berkelely, CA, Vol. 1, pp. 281–297.

Martinez, A. R., and Wegman, E. J., (2002). Text stream transformation for semantic-based clustering. Proceedings of Computing Science and Statistics, 34, 2002.

Menard, S. (1995). Applied logistic regression analysis. Sage Publications. Thousand Oaks, CA.

Mosier, K.L. (2010). The human in flight: from kinesthetic sense to cognitive sensibility. In E. Sala and D. Maurino (Eds.), Human factors in aviation. Academic Press. San Diego, CA.

Nadler, B., Lafon, S., Coifman, R.R., Kevrekidis, I. (2006). Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. Applied Computational Harmonics Analysis. Vol 21, 2006, pp 113-127.

Nagel, D. (1988). Human error in aviation operations. In E. Weiner and D. Nagel (Eds.), Human factors in aviation. Academic Press: San Diego, CA.

Novak, J.D. & Cañas, A.J. (2006). The theory underlying concept maps and how to use them. Technical Report IHMC CmapTools 2006-01

Novak, J. D., & Musonda, D. (1991). A twelve-year longitudinal study of science concept learning. American Educational Research Journal, 28(1), 117-153.

Perrow, C. (1984). Normal accidents: living with high-risk technologies. New York: Basic Books.

Phimister J.R., Bier V.M., Kunreuther H.C. (2004). Accident precursor analysis and management: reducing technological risk through diligence. National Academy of Engineering. Washington, D.C.: The National Academies Press.

Primavera, L. Allison, D., Alfonso, V. (1997). Measurement of dependent variables. In R. Franklin, D. Allison, and B. Gorman (Eds.), Design and analysis of single-case research. Erlbaum: Mahwah, NJ.

Reason, J. (1990). Human error. Cambridge University Press, New York, NY 1990.

Reason, J., Hollnagel, E., & Paries, J. (2006). Revisiting the 'swiss cheese' Model of accidents. EEC Note No. 13/06. France: EuroControl Experimental Center.

Rankin, B., & Allen, J. (1996). Boeing introduces MEDA, maintenance error decision aid. Airliner, April-June, pp. 20-27.

Rasmussen, J. (1983). Skills, rules, knowledge: signals, signs, and symbols and other distinctions in human performance models. IEEE Transactions: Systems, Man, & Cybernetics, 1983, SMC-12, 257-267

San, O.M., Huynh, V, Nakamori, Y. An alternative extension of the k-means algorithm for clustering categorical data. International Journal of Applied Mathematics and Computer Science. Vol. 14, No. 2, 241-247.

Sheridan, T.B. (1988). Human factors in aviation. Academic Press, Inc, San Diego, CA.

Sheridan, T.B. (2010). The system perspective on human factors in aviation. In E. Sala and D. Maurino (Eds.), Human factors in aviation. Academic Press. San Diego, CA.

Skinner, B.F. (1974). About behaviorism. Vintage Books, New York, NY.

Steinbach, M., Karypis, G., & Kumar, V. (2000). A comparison of document clustering techniques. In KDD workshop on text mining (Vol. 400, pp. 525–526).

Stephenson, J. (1991). System Safety 2000: A Practical Guide for Planning, Managing, and Conducting System Safety Programs: Van Nostrand Reinhold.

Sträter, Oliver. (2005). Cognition and safety: an integrated approach to systems design and assessment. TJ International, Cornwall, Great Britian.

Swain A., and Guttman H. (1983). Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications, (NUREG/CR-1278). Washington, DC: Nuclear Regulatory Commission, 1983.

Swanson, D.R. (1986). Fish oil, raynaud's syndrome, and undiscovered public knowledge. Perspectives in biology and medicine, 30(1):7–18, 1986.

Thoraval, F.R., Bastie, A., Pawlotsky, J.M., Dhumeaux, D. (1997) Epidemiological Factors Affecting the Severity of Hepatitis C Virus-Related Liver Disease: A French Survey of 6,664 Patients. Hepatology. Vol 26, No 2. August 1997 pp 485-490

Underhill, D.G. (2007). Exploring dimensionality reduction for text mining. Trident Scholar project report no 362. United States Naval Academy, Annapolis, Maryland

van der Maaten, L. (2010). Matlab toolbox for dimensionality reduction. Retrieved October 2, 2010, from http://homepage.tudelft.nl/19j49/Matlab_Toolbox_for_Dimensionality_Reduction.html

Vaughan, D. (1998). Rational Choice, Situated Action, and the Social Control of Organizations. Law & Society Review, 32(1), 23. doi:10.2307/827748

Vidulich, M.A., Wickens, C.D., Tsang, P.S., Flach, J.M. (2010). Information processing in aviation. In E. Sala and D. Maurino (Eds.), Human factors in aviation. Academic Press. San Diego, CA.

Wallace, B., and Ross, A. (2006). Beyond human error: taxonomies and safety science. Taylor and Francis Group. Boca Raton, FL.

Weiner, E.L. (1987). Management of human error by design. Human Error Avoidance Techniques Conference, Washington, D.C. December 1-3.

Wiggins, J.S. (1973). Personality and prediction: principles of personality assessment. Reading, MA. Addison-Wesley

Wiegmann, D., & Rantanen, E. (2003). Defining the relationship between human error classes and technology intervention strategies (Final Tech. Rep. AHFD-03–15/NASA-03–5). Savoy: University of Illinois, Aviation Human Factors Division.

Wiegmann, D. A., & Shappell, S. A. (2001). Human error analysis of commercial aviation accidents: Application of the human factors analysis and classification system (HFACS). Aviation, Space, and Environmental Medicine, 72(11), 1006-1016.

Wiegmann, D. A., & Shappell, S. A. (2003). A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System (1st ed.). Ashgate Publishing.

Willits, P., Abbot, M., and Kailey, L. (2004). Private pilot. Englewood, CO: Jeppesen Sanderson, Inc.

Yang, Z.R. and Chou, K.C. (2003) Mining biological data using self-organizing map. J. Chem. Inf. Comput. Sci., 43, 1748–1753

Zhao, Y., Karypis, G., & Fayyad, U. (2005). Hierarchical clustering algorithms for document datasets. Data Mining and Knowledge Discovery, 10(2), 141–168.