



Visual Analytics: *Visually Exploring Masses of Data*

David Darvill, PhD

Research Director

Vancouver Institute for Visual Analytics

Contact: ddarvill@sfu.ca

Visual analytics (VA) was initially proposed to help US intelligence analysts deal with the masses of security-related information made available to them following the World Trade Center attacks on September 11, 2001.

The amount of data stored in digital databases is rapidly growing beyond the capabilities of even expert analysts to manage and use in a coherent manner (see *World Data Deluge* sidebar). Many organizations are trying to deal with this growing data tsunami.

Visual analytics is “the science of analytical reasoning facilitated by interactive visual interfaces”¹. VA is a multidisciplinary field intended to help people understand how to synthesize information in order to derive insights from massive, dynamic, ambiguous, and often conflicting data. VA proposes to take advantage of the visual intelligence and cognitive capabilities of human analysts using interactive, exploratory, visualisation tools that also include analyst-guided statistical and machine learning capabilities.

VA was originally intended to solve data analysis problems in the security and intelligence domains. Tools and techniques that have been developed in the past 6-8 years have become of great interest to analysts in many other data-rich domains, e.g., aerospace safety, manufacturing and maintenance; financial risk analysis and fraud detection; business

process analysis; health care and medical research; and environmental health and safety.

World Data Deluge

Global digitally stored data was estimated to be 195 exabytes (10^{18}) in 2007 and is expected to grow to 1.8 zettabytes (10^{21}) by 2011 (the US Library of Congress collection contains approximately 10 petabytes (10^{15}) of data).

Raw data has little intrinsic value.

Data mining helps find expected patterns, e.g., look for and find gold in the data.

Visual analytics will help analysts find the expected, but also to discover the unexpected in their data, e.g., look for and find gold, but also possibly find silver or copper in the same data.

Humans have some very impressive visual and cognitive capabilities, but humans change very slowly, e.g., brain volume has doubled in approximately 2.5×10^6 years.

Technology, however, has changed very quickly, e.g., Moore’s law shows integrated circuit capacity doubles in approximately 2 years.

One *goal of visual analytics* is to build better tools (technology) to take advantage of human visual and cognitive problem solving capabilities.

In 2007, the Boeing Company provided a \$1.30M research grant over 5 years to the University of British Columbia and Simon

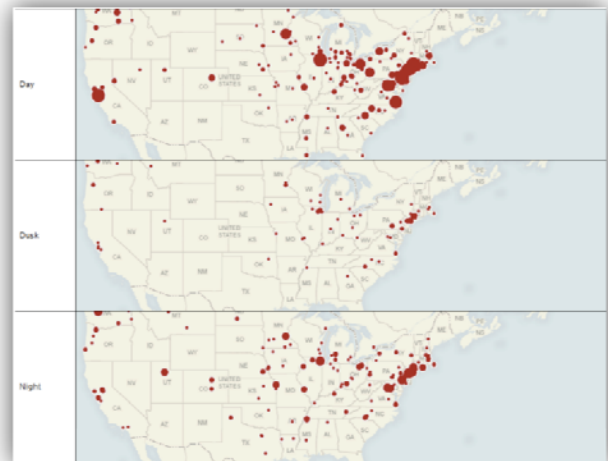
Fraser University to jointly study VA, with the general goal of discovering how VA might be used within Boeing. In 2010, a similar gift was given to both universities to establish the Vancouver Institute for Visual Analytics (VIVA). VIVA will advance VA in Canada by further promoting the excellent research already being conducted in the context of real world problems and by building a capacity to train VA human resources at all levels.

The work of the initial Boeing grant was focused on four areas:

1. Basic research on human visual capabilities relevant to computer based visualisations, e.g., seeing correlations in scatterplots, visual chunking of information for sequential assembly, representation of causation, and studying the effects of stress on visual processing.
2. Analysis and modeling of the analysis process itself using information from field studies where a VA tool expert is paired with a relevant subject matter expert to study a specific problem domain, e.g., aircraft safety engineering issues related to bird strike incidents and environmental health and safety issues related to shop floor incidents.
3. Development of a challenging learning environment where participants, e.g., university students and corporate employees, can learn about VA by mentoring newer participants to use available VA tools to explore real problems using real data sets. Participants keep journals in order to capture their learning process so that researchers can better understand the learning of VA procedures and tools.
4. Design, development and evaluation of a next-generation VA tool, CZSaw, that will address procedural and capability gaps identified by users of other VA

tools, e.g., analysis of document and entity networks and capture and support of the analysis process itself, allowing analysts to better understand, control and re-use procedures.

In 2010, as part of the VA research program, two students participated in VA internships at the Boeing Company. One intern worked with a Boeing Safety Engineer on the problem of bird strikes on commercial aircraft. All airframe manufacturers collaborate to study safety incidents to assure that air travel stays the safest mode of mass transit. In this way, manufacturers can change their designs and improve pilot training to minimize the impact of incidents like the Hudson River landing of US Airways Flight 1549 on January 15, 2009. Both engines shut down after hitting a flock of Canada Geese.



Birdstrikes during day, dusk and night.

The VA work done by the Boeing bird strike expert and the SFU VA intern led to four design change recommendations and improvements to pilot training material.

The second VA intern worked with Boeing Environmental Health and Safety (EHS) ergonomists at the Boeing manufacturing facility in Everett, Washington². This facility was opened in 1967 to build the

then-new 747 aircraft. Since then, the facility has been extended to manufacture the 767 and 777 models, and is now also producing the latest 787 Dreamliner. All of these aircraft are built in the world's largest, by volume, manufacturing facility.



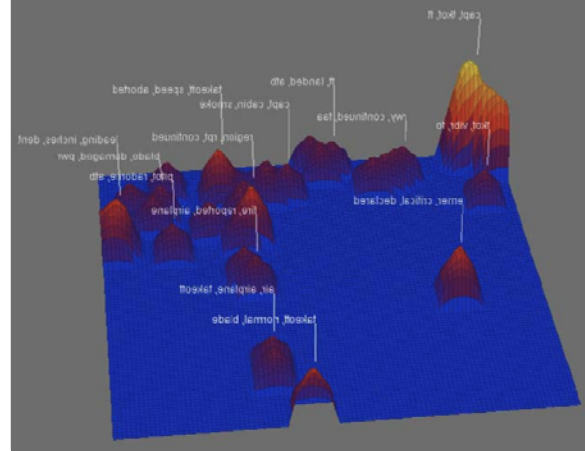
787 production line (Gail Hanusa/Boeing)

The value of VA in an EHS context was demonstrated by the analysis of injury incident reports for one of the aircraft production lines at Everett. Each year, the EHS ergonomists review all of the incident reports for their manufacturing line to establish what types of safety problems were most prevalent in specific areas. Using this information, they then prepare worker safety programs for the upcoming period, e.g., if there were frequent ladder incidents in an area, they might prepare a campaign related to ladder safety.

This process required the ergonomics team to manually review the incident reports to determine the frequency and impact of incidents in order to select the problems to concentrate on. This procedure was very time consuming.

After overcoming data import issues related to the not uncommon truth that necessary data is often in different datasets, the VA intern was able to use one of the existing VA tools, INSPIRE, to analyse the injury incident text records. She was able to display the results in a

view that allowed the ergonomists to see clusters of related incident types as well as the quantity of incidents in each cluster in a single 3D visualisation (see example image).



INSPIRE 3D cluster visualisation example.

A comparison of the VA visualisation results to the actual safety programs chosen for the same year indicated that there was a high correspondence between the peaks of the 3D visualisation and the topics of the worker safety initiatives chosen by the ergonomists using their traditional method. There are obvious cost savings to be made when comparing the multiple person-days needed for the traditional method compared to the day or so required using the VA tool. This potential saving has generated a project to extend the VA analysis to other aircraft lines at Everett and eventually to other manufacturing locations within Boeing.

Notes:

1. Thomas, J.J. & Cook, K.A. (Eds.). Illuminating the path: The research and development agenda for visual analytics. IEEE Computer Society, 2005. (See pdf at <http://nvac.pnl.gov/agenda.stm>.)
2. <http://www.boeing.com/commercial/facilities/index.html>.