UWE Bristol | University of the West of England

Presentation by

Dr. Phil Legg

Associate Professor in Cyber Security

Security Data Analytics and Visualisation

7: Visual Analytics

Date: Autumn 2019

Module Roadmap

Week 6 (last week): Pause and Coursework Discussions

Week 7 (this week): Visual Analytics

Week 8 (next week): Research on Visualisation for Cyber Security

Week 9: Case Studies – Text Analytics

Week 10: Case Studies – Video Analytics

Week 11: Future Directions in Security Analytics, and Revision Guidance

Week 12: Drop-in Q & A, and Revision Guidance

What is Visual Analytics?

Creation of tools and techniques to enable people to:

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data
- Detect the expected and discover the unexpected
- Provide timely, defensible, and understandable assessments
- Communicate these assessment effectively for action

What is Visual Analytics?

- **"Visual Analytics is the science of analytical reasoning supported by a highly interactive visual interface." [Wong and Thomas 2004]**

- "Visual Analytics combines **automated analysis** techniques with **interactive visualisation**s for an effective **understanding, reasoning and decision making** on the basis of **very large** and **complex datasets**" [Keim 2010]

- **Detect the expected and discover the unexpected**

Knowledge Generation using Visual Analytics

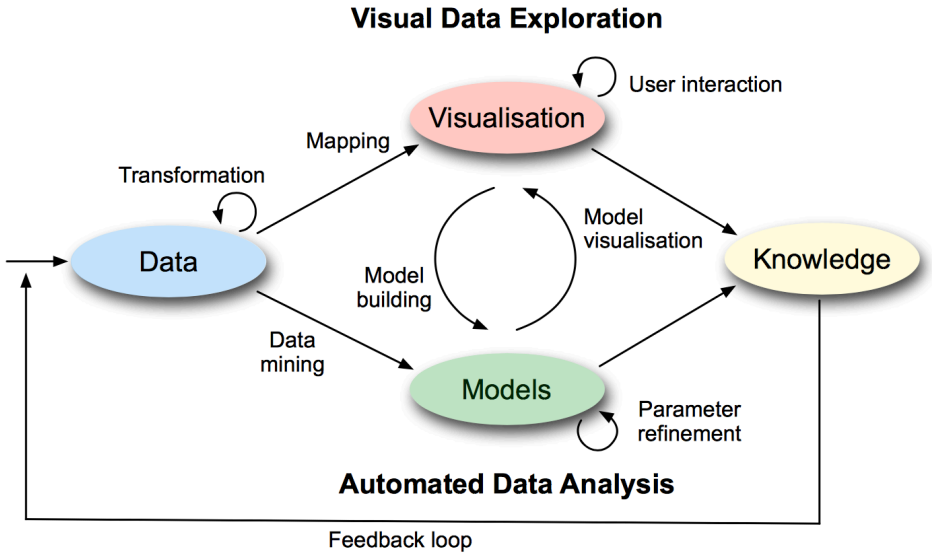


Figure 2.3: The visual analytics process is characterised through interaction between data, visualisations, models about the data, and the users in order to discover knowledge

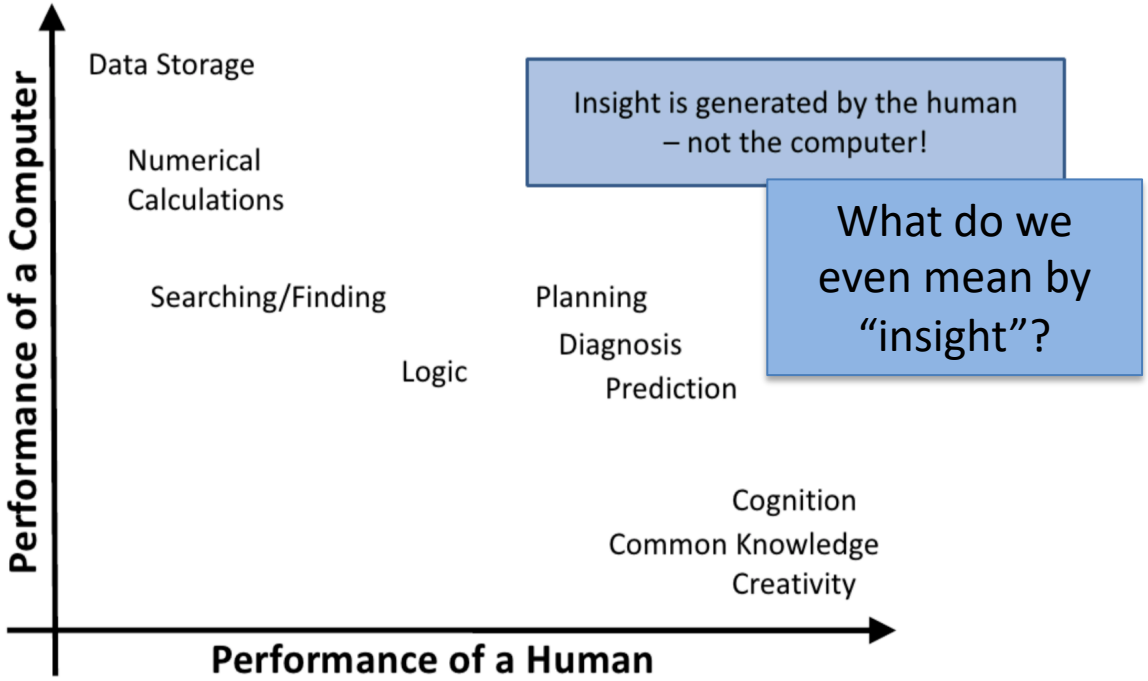
D. Keim. Mastering the Information Age-Solving Problems with Visual Analytics, 2010.

What is Visual Analytics?

- Automated methods
  - + Scale well
  - - Get stuck in local optima
  - - Run in a "black box" fashion
- Visualisation methods
  - + Interactive data analysis
  - - Scalability

- **Visual Analytics** integrates both
  - Tied together by the user
  - Alternating between visual and automatic methods
  - Collaborative approach for problem-solving / story-telling between user and machine

The ability matrix

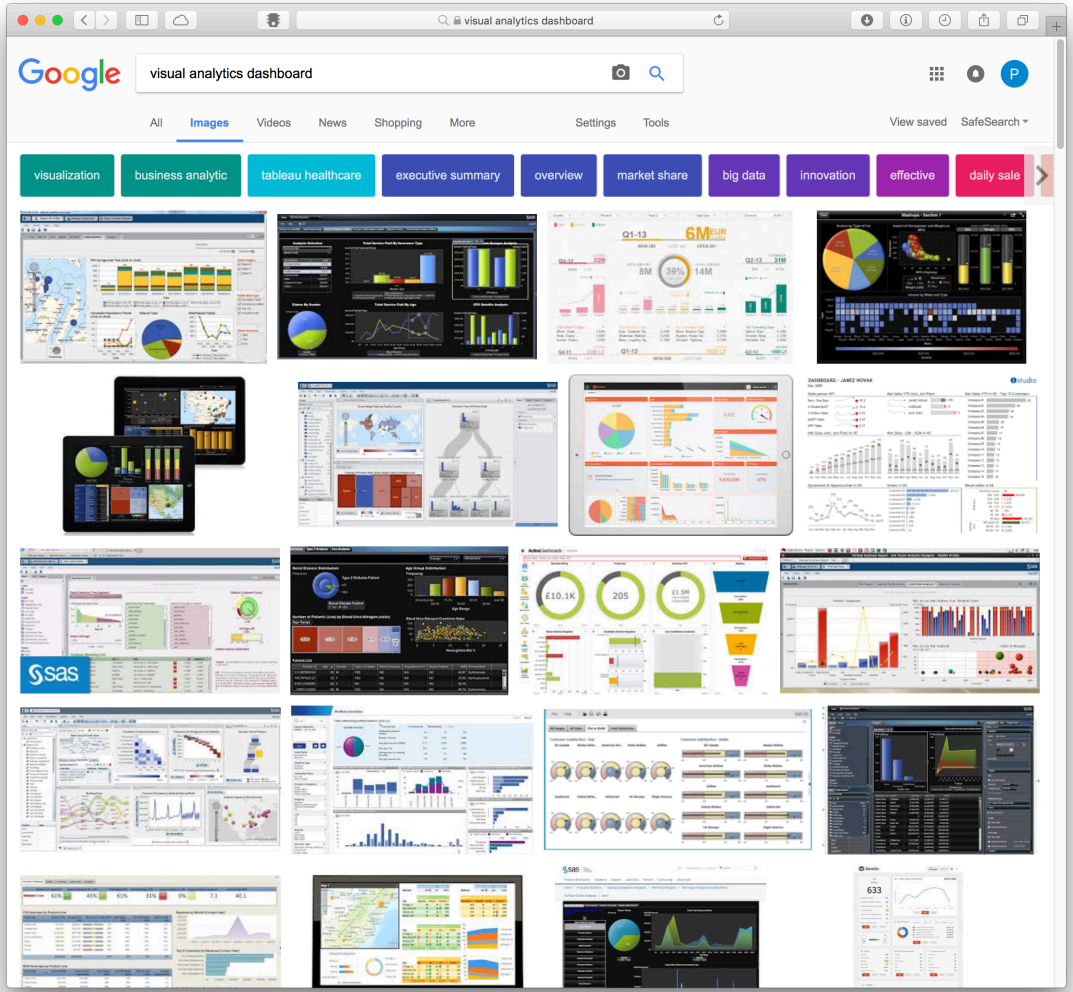


adapted from Daniel Keim, Uni. Konstanz

Visual Analytics Mantra

Analyse first,
show the important,
zoom/filter,
analyse further,
details on demand

[Keim 2006]



- Many examples exist of visual analytics "dashboards" – just see Google Images!
- "Dashboards" aim to summarise data to provide information – to what extent do they support both overview and detail?
- "Linked" visualisations that update through interaction – be careful of appropriate design choices (pie charts!)

User Interaction

- Humans learn by "doing" – same in visualisation
  - If we can interact with it, we can see how it works much clearer
  - Allows us to also identify strengths and weaknesses of our analysis

- Parameter selection (e.g., select a date range)
- Filtering and selection (e.g., select a particular group of users)
- Brushing (e.g., select a particular region on an axis)
- Reordering (e.g., changing the data on each axis)
- Zoom (e.g., increase detail for a given region on a map)

User Interaction

- How can we interact with such systems?
  - Mouse / Keyboard
  - Stylus / Pen
  - Touchscreen
  - Hand gestures (e.g., Leap Motion / Kinect)
  - Voice

- What are the costs associated with interactions?
  - Tiredness / Fatigue, Complexity of interaction
- Can we make use of multi-modal interactions?
  - E.g., Combination of voice and touch?

There is good reason why mouse and keyboard interactions are still the most common…!

Need to account for user expectation / familiarity

Exploratory Data Analysis

- As mentioned, **Detect the expected and discover the unexpected**
- It may be that we don't know what we are looking for initial (e.g., think of a malware infection – we want to understand why it happened, but we don't know where to start)
- Exploratory Data Analysis is – as the name suggests – exploration of parameters to see how they impact on our analysis. We may not know what it is we are after until we see it.
- This process means we are continually updating our **mental model** of the data, the relationship of attributes, and what may be informative from the analysis.

UWE Bristol | University of the West of England

Challenges in Visual Analytics

Cognitive and Perceptual Load

- "Big Data" – humans can not comprehend the full dataset, and machines can not "process" the full dataset (what do we mean by process anyway?!)
- Visual Analytics can support **memory externalisation** such that the system can serve to "remind" the user of some detail when needed, rather than the user be expected to recall all previous observations of data.
- Similarly, users can not visualise all data at once. As system designers, we need to make decisions about the level of perception that is appropriate when data is being conveyed to the user (or at least, offer the user a means of adjusting this level).
- **Scalability** will continue to dominate as a challenge

Semantics of the data

- How do we extract the **semantics** of our data?
  - i.e. the real meaning of our data.

- Think of a large document collection (e.g., e-mail conversations) – a system may be able to count the occurrence of words across different documents, but does the computer *understand* what the words mean?

- This re-enforces the need for the combination of automated methods and user interactions – but we sometimes need to consider how we can convert our knowledge into something the machine can understand and visa versa.

Uncertainty in data

- All data is collected from some **sensor**
  - E.g., packet capturer, video camera, e-mail log file, user log file
- How do we know that our sensor is reliable?
  - Is the data incorrect, or is the sensor reporting incorrectly?
- Are there ways that we can extend traditional VA techniques to illustrate the confidence / uncertainty that we have about our data.
- How do humans learn to deal with such additional knowledge?
  - How will a human interpret the knowledge that a sensor is 65% confidence? (e.g., Frequentist vs Bayesian probabilities)

Uncertainty in users

- How do we know that a user will interact as we (designers) have intended?
  - What if they make incorrect assumptions because of poor representation (either from poor analysis or from poor visual mapping)?

- How do we know that users will provide the "correct" input?
  - Consider a supervised machine learning application – it requires the human to provide accurate input.

- How can we trust the user?
  - They may make mistakes, they may act deliberately (think insider threat!)
  - Can we use VA / ML to recognise user error to learn to discard this?

Examples of VA

- "Linked views"
  - Interactions in one pane may update information elsewhere
- Multiple views support combined analysis
  - Do multiple views help support our understanding better?
- How would we expect a user to use this system?
  - "Intuitive" designs where possible

Visualizing the Insider Threat:
Challenges and tools for identifying malicious user activity

Philip A. Legg

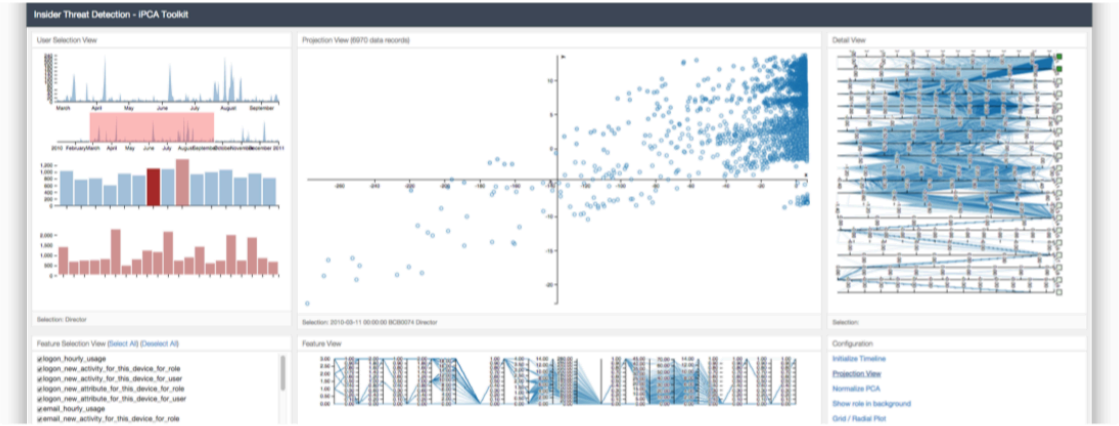


Fig. 3. Layout of the visual analytics dashboard. The dashboard consists of six panes (from top-left to bottom-right): User Selection View, Projection View, Detail View, Feature Selection View, Feature View, and Configuration View.

Examples of VA

- "Linked views"
  - Interactions in one pane may update information elsewhere
- Multiple views support combined analysis
  - Do multiple views help support our understanding better?
- How would we expect a user to use this system?
  - "Intuitive" designs where possible

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS VOL. 23, NO. 1, JANUARY 2017

A Visual Analytics Approach for Understanding Reasons behind Snowballing and Comeback in MOBA Games

Quan Li, Peng Xu, Yeuk Yin Chan, Yun Wang, Zhipeng Wang, Huamin Qu, *Member, IEEE*, and Xiaojuan Ma

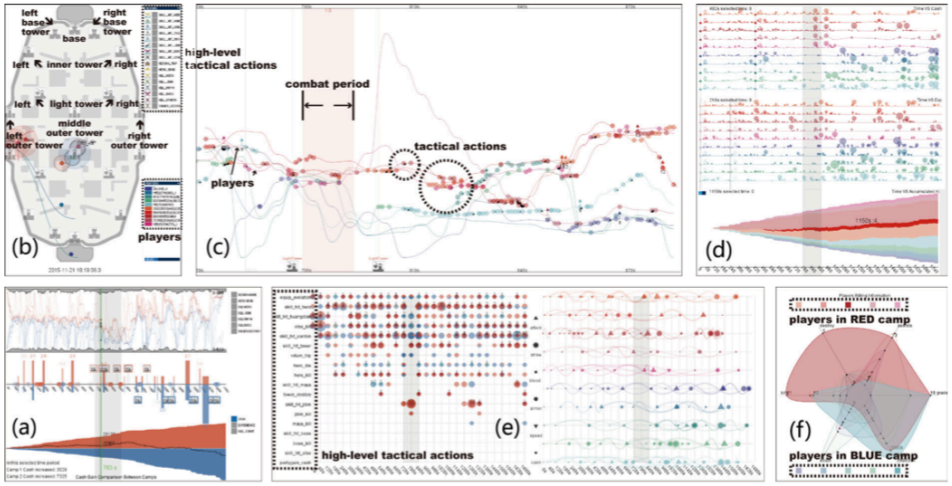


Figure 1. A match with *comeback* occurrence. (a) Trend View discloses the trend of game play during a match. (b) Trajectory View simulates the game replay. (c) Tactic Geographical Timeline View presents details of players' behavior in the time period of interest. (d) Resource Time Sequence View displays the accumulated resources and changes in resources of each player. (e) Tactic Comparison View (Left) unfolds the temporal dynamics of all the tactical actions in two camps while Equipment Evolution View (Right) shows the equipment evolution hierarchies. (f) Player Billing Radar View represents the statistical information of each player.

Examples of VA

- "Linked views"
  - Interactions in one pane may update information elsewhere
- Multiple views support combined analysis
  - Do multiple views help support our understanding better?
- How would we expect a user to use this system?
  - "Intuitive" designs where possible

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS VOL. 23, NO. 1, JANUARY 2017

SmartAdP: Visual Analytics of Large-scale Taxi Trajectories for Selecting Billboard Locations

Dongyu Liu, Di Weng, Yuhong Li, Jie Bao, Yu Zheng, Huamin Qu, and Yingcai Wu

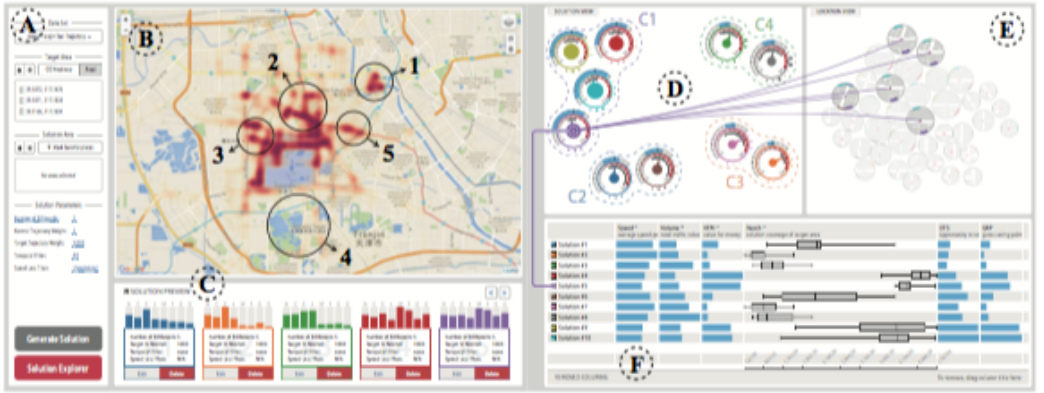


Fig. 1. SmartAdP system. (A) Dashboard View shows the information of the current solution for billboard placements. (B) Map View provides a visual summary of the geospatial environment. (C) Solution Preview lists the parameters and statistics of the candidate solutions. (D) Solution View lays out all the solutions as glyphs to reveal the relationships among the solutions. (E) Location View supports in-depth analysis at the fine-grained location level. (F) Ranking View displays multi-typed ranks of the solutions.

Examples of VA

- "Linked views"
  - Interactions in one pane may update information elsewhere
- Multiple views support combined analysis
  - Do multiple views help support our understanding better?
- How would we expect a user to use this system?
  - "Intuitive" designs where possible

TextTile: An Interactive Visualization Tool for Seamless Exploratory Analysis of Structured Data and Unstructured Text

Cristian Felix, Anshul Vikram Pandey, and Enrico Bertini, *Member, IEEE*

Fig. 1. *TextTile* interface showing data from the *Yelp-Heathcare* reviews dataset with: a) *fields* panel showing all the fields present in the data; b) *filter* panel with filter specification to select only reviews from New York; c) *split* panel with three segments generated using the *business category* field; d) *summarize* panel having three segments (*Medical Centers*, *Chiropractors*, and *General Dentistry*) with *keywords charts* to show relevant words, bar charts to show rating distribution and maps for location distribution by zip code.

Examples of VA

- "Linked views"
  - Interactions in one pane may update information elsewhere
- Multiple views support combined analysis
  - Do multiple views help support our understanding better?
- How would we expect a user to use this system?
  - "Intuitive" designs where possible

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS VOL. 23, NO. 1, JANUARY 2017

Visual Analysis of MOOC Forums with iForum

Siwei Fu, Jian Zhao, Weiwei Cui, and Huamin Qu

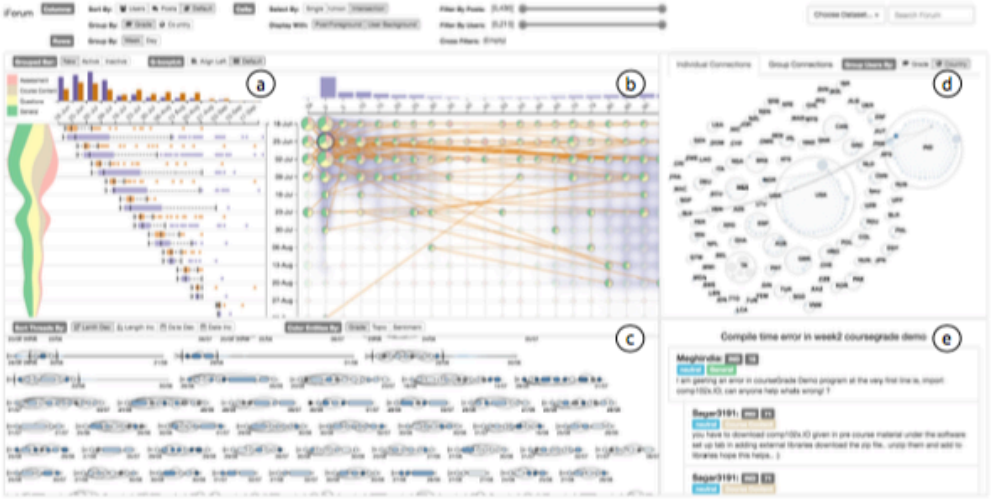


Fig. 1. Using iForum to explore the MOOC forum of a JAVA programming course that has attracted more than ten thousand students during a ten-week course period. (a) The Overview shows the overall changes of posts, threads, and users on the forum. (b) The Matrix View further enables the comparison of dynamic patterns of different user groups along time. After a cell of interest is selected, orange lines are shown on top of the matrix to indicate the threads passing through that cell. (c) Meanwhile, the Thread View presents all selected threads in a compact layout, and (d) the Social Network View reveals the interactions among corresponding users based on their replying relationships. (e) When a specific thread is selected, the Text View displays discussions in traditional indented form.

Evaluation of VA

- What makes a good Visual Analytics system?
- How can we scientifically evaluate design choices?
- What can we do to develop reproducibility for scientific rigour – rather than VA research just resulting in bespoke software development?
- Evaluation of Visual Analytics systems is very difficult (and often debating in the research community). We know when a system is appropriate for a given use case, but expressing ***why it is effective*** can often be challenging.
  - Very much an on-going discussion – relies heavily on user studies.
  - What else can we do beyond user studies?

VAST and VisSec

Summary

- What is Visual Analytics and why is it useful?
- Why VA and not just automated methods or visualisation alone?
- "A Visual Analytics Mantra"
- Dashboards as a form of VA
- User Interaction
- On-going Challenges in VA

Next Week:

Look at the Visualisation for Cyber Security (VizSec) webpage and Proceedings Browser – explore publications from previous years https://vizsec.org

Look at (read!) the papers available from this year (I've put some on Blackboard for you) – *we shall discuss next week!*