VALCRI

D.1.0

Evaluation of DOTS: Data as Objects in Transposable Space

VERSION 1.0

Date submitted: 20 Apr. 16

Dissemination Level: PU

WORK PACKAGE:

SUPERVISORS: B.L. William Wong and Neesha Kodagoda

Report Prepared by: Varshita Sher

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no FP7-IP-608142.

SEVENTH FRAMEWORK PROGRAMME

Programme:	FP7 Theme 10: Security
Grant Agreement No.	FP7-IP-608142
Project Acronym:	VALCRI – Visual Analytics for sense-making in Criminal Intelligence Analysis
Project Co-ordinator:	Middlesex University – Professor B.L. William Wong

Deliverable No.	D 1.0
Document Title:	Evaluation of DOTS: Data as Objects in Transposable Space
Version:	1.0
Status:	For review
Document Manager:	B.L. William Wong
Organisation:	Middlesex University
Date:	20 Apr. 16
Filename:	VALCRI Report.doc

Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium participants (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Keywords:

Template, document, deliverable, report

EXECUTIVE SUMMARY

In criminal intelligence analysis, analysts grapple with voluminous data on routine basis in order to comprehend the uncertainties surrounding a crime. By means of sense-making and pattern recognition from this data, analysts can ascertain for themselves, the significance of the crime and take necessary measures. This research aims to uncover strategies that support investigators and analysts in the exploration and assembly of evidence and inferences. We aim to investigate how they make sense of the high dimensionality complex dataset by means of story-telling and sense-making. We focus here on real life problems such as those encountered by analysts on daily basis such as *"What time of the day needs more patrolling for Osterley"*. To better understand this problem and its related cognitive and interaction needs, we conducted a user study using a tool called DOTS (Data as Objects in Transposable Space) loaded with real but anonymised crime records. The tool is designed to aid analysts in organizing, structuring and assembling information to draw valid inferences from it. DOTS supports interaction with individual crime record as each crime is represented by a circular dot in space. It also provides the choice of five inbuilt views (Row, Column, Bar, Segment, Table) that represent the same dataset in multiple ways. Further, In-place transposition between these views allows the analysts to change perspective and thus gather new insights. The study focuses on the strategies user create and their reasoning process. We enforce some design guidelines while creation of DOTS and evaluate them against the strategies observed during the analysis. Finally, DOTS must not be considered as the replacement of human work-force from the decision making process but rather as a tool that assists the analysts in making informed and unbiased decisions. We envision the future applicability of the tool to diverse dataset as well.

Contents

1 INTRODUCTION

In criminal intelligence analysis, the analysts encounter a large amount of data on a daily basis which must be sifted, not just to find information but rather to constantly and iteratively assess, learn, draw conclusions and make judgments from it [1]. In addition, one of the most labor-intensive tasks is the creation of a matrix that enable crime pattern analysis using the Comparative Case Analysis i.e. the comparison of crimes across many different dimensions.

Most of the previously used analysis tools have been unable to cater to these requirements. Graphical visualizations currently used in criminal intelligence such as histograms and bubble plots focus on presenting the data aggregate rather than individual data points. However, this poses an inability for the analysts to perceive the “complete” data set altogether and retrieve patterns and connections within.

In this paper, we propose a novel criminal analysis tool- DOTS (Data as Objects in Transposable Space) which allows analysts to individually pay focussed attention to each crime report in addition to showing the aggregates. Aggregating data, yet being able to see the specific data items that make up the aggregate, helps in analysing crimes in context of other crimes to draw relevant connections and associations between them. DOTS assist the working of criminal intelligence analysts, who work on their hunches and intuition mostly, by providing cues for leap of faith and assumptions in a more organized and systematic manner.

DOTS also facilitate sense-making and pattern recognition, activities that are the first order business, as analysts grapple with the principle uncertainties surrounding a crime, specifically the who, what, why, when, where and so on. The search for patterns is inherently incessant regardless of whether the data is binned into meaningful clusters or categorizations. This is deemed important as Card, Mackinlay and Shneiderman declare that the "purpose of visualization is insight, not pictures" [10].

2 RELATED WORK AND RESEARCH REVIEW

We enumerate in this section several theoretical foundation that facilitate the development of the DOTS software along with specifying few of the previous software tools that aim to achieve and enhance insights gained during analysis of an information visualization.

2.1 Sense-making and insight generation

Yi *e*t *a*l [9] established the importance of the five procedural aspect of sense-making – Provide Overview, Adjust, Detect Pattern and Match Mental Model – in order to promote insight acquisition, regardless of the complexity and sophistication of the data transformations and interaction techniques of the software [9]. We incorporate them during design of DOTS so as to identify patterns, trends and unusual occurrence in dataset.

Drawing parallels from the pre-established theories in [9], Kodagoda *et al* [11] developed INVISQUE (INteractive Visual Search and Query Environment). They conducted a user study to evaluate INVISQUE, with the question of *"identifying the candidate influencers within an academic community"* to form design guidelines that support investigators and analysts in exploration and assembly of evidence and inference in order to understand cognitive and interaction needs of users [11]. The searches were submitted on an 'infinite' canvas and results presented as horizontal and vertical set of indexed cards, laid out according to user-selectable meta-data such as year of publication and citation count. The four resultant design implication – Better support for working-set creation, Playful Exploration, Enabling focus to selected candidate influencers and Comparing candidates – overlapped coherently with the framework described in [9].

2.2 2.2 Visualization tools for enhancing insight acquisition from visualizations

Animation has been recently gained new prominence for visual representations of trends in multi-dimensional data. With the introduction of Gapminder Trendanalyzer, Hans Rosling has used it on several occasions to represent trend animations for several data-set including breast and cervical cancer statistics and child mortality reduction. However, the disadvantage of such a visualization is the requirement of an engaging presenter who keeps the audience engaged by directing their attention towards interesting peaks and troughs. As an alternative, Robertson *et al* [12] proposed two trend visualizations that use static depiction of trends – Trace and Small-multiple visualization. The former shows traces of all trends overlaid simultaneously in one display with the progress direction represented by using transparency of bubbles, fading from opaque to transparent and the latter uses small multiple displays to depict potentially interesting areas.

Both the techniques have scalability issues (beyond 200 points), but we take the positive aspects of each while incorporating a novel trend visualization technique in DOTS. We introduce abstraction and depict the *x-y* trend using a line chart with a dot on the line representing the current intersection value of the two dimensions. This enables the analysts to preview the trend for the entire dataset along with highlighting the current value therein. Also, instead of presenting multiple small displays for trends of multiple *x-y* dimensions, we introduce user-selectable dimensions.

While the aforementioned techniques work well when data density is low (100-200 data points), it is inadequate to cater to voluminous data. Even more so, it becomes difficult to pin-point any particular record without browsing through a series of drop down menus and additional filters. There has been an inability to inspect a "complete" dataset on a visual timeline [13] due to factors such as visual clutter and screen resolution restrictions. Various studies have been done to efficiently organize data for visual display that doesn't hamper the analysis process at grass-root level by circumventing the need to summarize records and thereby allowing examination of individual records. Krautli and Davis [13] developed a timeline visualization tool that visualized the large cultural collections contained at Museum of Modern Art, horizontally along their acquisition date. They instill the use of zoom and pan technique, in accordance with Shneiderman's policy "overview first, details on demand", using which the analyst can concentrate on a particular cluster and examine a specific record therein. Each record is a disk of the same size and color-coded them by the type of collection (photographs, books, paintings, etc).

Similar visualization strategy was adopted by Stramp and Wilkerson, who developed Legislative Explorer [14]. It is an interactive animated data visualization tool that visualizes the process of bill submission by Congress houses, Senate and House of Representatives. The software is aimed at facilitating observation of large scale trends in congressional law making by people without any advanced methodological training. Each icon represents a specific member of congress, color coded by its sponsor, details of whom could be obtained by hovering the mouse over it. The software animates the visual spray of bills introduced by each party and their journey through Sign, Veto, Override and President's Desk until it finally becomes a law.

Previously some visualization tools for criminal intelligence analysis have been specifically developed. One such tool combines Chicago crime data with Census data for new insights. It is a form of scatter plot where the x and y axis can correspond to varying features such as arrest percentage, hardship level & crime count and each data point on the graph represents a community area. The size of the bubble is equivalent to the per-capita income of the community. A major disadvantage of this kind of visual representation is the overlap among different bubbles thereby blocking or hiding smaller bubbles underneath. The color corresponds to the percentage of unemployed people in the age group 16+. The tool is made 1) *interactive* by hovering over the bubbles for more detailed description (hardship index, per capita income) of the community area and 2) *animated* by using concept of motion charts, explained previously, where bubbles move over time to signify trend.

3 DOTS: A Brief Description

DOTS (Data as Objects in Transposable Space) builds upon the functionalities of the scatter plots, in that each data point, representing a crime record, is represented as a circular dot on screen. Akin to scatter plots, they show how the data distribution in terms of relevant *x* and *y* dimensions. However, in contrast to the scatter plots the spatial layout of the dots do not correspond to the distribution in the x-y plane. The dots are merely present as aggregates based on dimensions in the x and y axes.

Instead of conforming to a single visual representation of data distribution, DOTS allows the extended ability to smoothly transpose between a variety of visual forms such as contingency table and bar charts. It also allows data to be presented at different levels of abstraction depending on the requirements of the analyst, thereby following Shneiderman's policy of 'Detail on Demand' [10].

4 Sense-making using DOTS

In terms of sense-making, DOTS enhances the user experience of sense-making and inference generation by re-structuring (bar charts to contingency tables) and re-organizing (highlighting, sorting) data seamlessly from one visual representation to other. While the data presented remains the same, the amount of *information* conveyed differs as each individual has a unique process of sense-making and insight generation [5] conveyed through varying perspectives.

Data preparation constitutes a major chunk of the workload of an analyst who must organize the data first in order to derive conclusions from it. This has been the primary motivation behind development of DOTS so that analysts could save time and effort. DOTS aim to understand the data set, the patterns and anomalies resident therein, which might be time-consuming, if not impossible, when using conventional analysis tools such as Excel. The outcomes from it can be used to make future predictions and decisions about a situation.

5 Design Protocols for creation of DOTS

5.1 Exploration of Data:

The key goal is to help the user to see what is in the dataset and therefore what might be asked of it. To do this, DOTS present the data in a way to reveal the structure of the dataset, in particular how the data is distributed, organized, and related. The organizations may be done in a variety of formats to help the user understand the variations in the parameters of the data. The higher order goal is to assist in developing situation awareness.

5.2 Developing Situation Awareness:

One of the most common task performed when data is freshly made available is to generate an overview summary of it. This helps us assess the overall picture conveyed by the data in hand. The explanation that one provides at this time, is a good indicator of one's understanding of the situation. DOTS facilitate such a description by representing each crime report as an individual dot on screen which are all simultaneously available for scrutiny of any underlying patterns or anomalies.

5.3 Support for multiple views/visual representation/visualizations (Multiple Coordinated Views systems):

Each of the visual form in DOTS is intended to be a visual representation of the data organized in meaningful ways. The key goal of DOTS is to assist the analysts in understanding their dataset so that they can then query the data accordingly by asking informative questions relevant to data set. To do this, DOTS provide the functionality to reveal the structure of the data set or any unobvious patterns therein, using five main views/organizations/representation namely, Block view, Column view, Bar view, Segmentation view and Contingency Table view. All of these provide information on how the data is distributed, organized and related to achieve the higher order goal i.e. to create situation awareness. By indexing, filtering, organizing and classifying the data, to support analytical reasoning process the analysts can engage in the process of evidential reasoning by constructing explanations from any of the four presented views. These different views allow the analysts to look at the same data from different perspectives to create fresh observation from them. Thus, support for multiple visualization of same data with same parameters are the core strength of DOTS which will help "overcome possible misinterpretations and perform interactive investigative visualization through correlating the information between views" [15].

It is necessary to mention that although there isn't any optimum representation/view, per se, the choice of a view is largely dependent on the analyst's preference and the analysis task at hand. The five views differ mainly by the number of dimensions/features of the data set available for scrutiny. We now briefly explain each of them and leave it at analyst's discretion to select the one most relevant to their line of enquiry.

5.3.1 Block View

Figure 1 depicts a block view in DOTS where only one of the dimensions (year, month, district, etc.) of the dataset can be visualized. The entire dataset is presented using a 1-dimensional block of data. This big chunk can help provide the overview of the data in respect to the dimensions. For example, visualization in Figure 1 presents all the crimes that occurred during the *year* 2006.

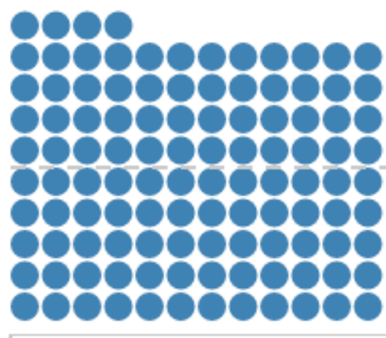


Fig 1: Each dot is a crime report.

5.3.2 Column View

This view allows two dimensions of the dataset to be visualized at once. Figure 2 exemplifies the crime data which are arranged in a 2-dimensional manner with x-axis as the *months* and y-axis as the *year*.

5.3.3 Bar View

This view allows three of the dimensions of the entire data to be visually represented simultaneously. DOTS does so by further allowing the dissection of each of the columns from Column view so that the crime records in each column are organized into bars on basis of third additional dimension. Figure 3 shows the Bar view with each bar representing day of the week, within each column which corresponds to the months and row set as the year 2006.

5.3.4 Contingency Table View

This view is similar to the Bar View in that it allows for the representation of three dimensions simultaneously. However, in Contingency Table View, the dimension representing columns in Bar view is analogous to the rows of the table and the bar correspond to the columns of the table. While the Bar view facilitates facile comparison of bars within one column, the crimes in a tabular format allow ease of comparison of a bar across multiple columns. For example, Figure 4 depicts the contingency table view with the fields of comparison highlighted in Red. As is evident, it is much easier to compare them in this manner compared to its equivalent Bar view in Figure 3.

5.3.5 Segmentation View

This view allows for the maximum number of dimensions to be displayed i.e. four. The first three correspond to those obtained from the Bar view. In addition to it, each bar is further broken down or segmented based on the fourth terminal dimension. Figure 5 depicts the Segmentation view which is an extension to the bar view in Figure 3 with crimes in each bar broken down in segments corresponding to the period of the day when the crime occurred.

Note: In all the five aforementioned views, in addition to the number of dimensions presented using each, there is *always* a possibility to represent an additional dimension using the color of the dot. DOTS provides a *Color By* attribute using which all the dots can be color coded by a dimension such as *district* or *year* of the crime offence.

5.4 In-place transposition between multiple views:

Attribute Dots support the functionality for in-place transposition. In-place transposition, is the feature where the dots are re-arranged from one visual form or view (e.g. dots stacked together in a Bar View), into another visual form (e.g. a Contingency Table View) in the same field of view. This is done dynamically using animations to create smooth transitions, and interactively as new questions arise as a result of the previous transpositions. This particular feature of DOTS is the key essence of any tool designed to assist in the analytic processes performed by crime intelligence professionals. This is because, although the computing power of tools is a crucial factor in the analytic process, most of the conventional tools used by the analysts are memory intensive. Thus, the time delay between querying the tool and generating a response hinders time-efficient analysis. However, DOTS allows the uninterrupted fluidity of one's work as it processes the data or performs operation on the go.

The ability to switch between multiple data organizations (views) on the screen has dual benefits – 1) accommodating data in a concise manner when one of the views become visually overwhelming, and 2) to follow up individual or a group of crimes (dots) across different views as and when they are categorized by the offence type, district type, day, month, year and period of the day.

5.5 Representing data as individual objects rather than aggregates:

DOTS is based on the concepts of representing each crime report as a single dot on screen. This kind of representation ensures tangibility as the analyst can now interact more directly with each crime record separately or in a group, if the need be. There is complete freedom to select, move and carry out actions on dots in addition to organizing them based on certain characteristics.

5.6 Highlighting data as Objects:

The highlighting of the crimes (under scrutiny) will enable the analysts to see which specific crime report (presented by the dot) may have unusual characteristics and viewed in the context of other crime reports.

5.7 Tangibility:

In the context of data as objects, we have mentioned that each dot is an object such as a crime report. Each dot can also be interacted directly with – by selecting it, moving it and carrying out actions directly on that dot. For example, the report directly associated with the dot can be called upon by clicking on the dot.

5.8 Slicing and dicing data as per visualization needs:

Depending on the analysis requirements, the analyst may choose to either generate an overview of the crime dataset (using low level representations such as Block View or Column View) or acquire in-depth details (using high level representations such as Segmentation view) such that dots are arranged together to ensure dimension-proximity. Nonetheless, the provision of the designs for the five different views are tailored to suit the task or challenges at hand and guarantee that all of the analysts' experience based routines such as searching and sorting are easily performed. This complements the Principle of Emergent Features as, depending on the level of granularity, specific emergent properties become visible as data is aggregated differently [7]. As prescribed by Bargiela and Pedrycz, each level in DOTS build upon the levels below it and contributes to the qualitative different emergent properties of the level above [8].

5.9 Laddering:

In context of DOTS, we refer to *elaboration* as the process of finding cues that trigger new questions and queries about the current information presented such that digging deeper into them, helps develop a better understanding of our dataset alongside revealing trends and patterns.

6 METHOD

Describe We use a carefully designed a within-subject repeated measure study to analyze the working of the police criminal intelligence analysts to understand how DOTS facilitate their sense-making and inference making processes from a given set of crime records.

Participants: The pilot study was conducted using university students as participants to ensure smooth working of tool (DOTS) and data collection techniques. Since using domain exerts such as the crime analysts themselves can perform the tasks most realistically to provide genuine results, for the main study, (n=10) senior police analyst from the Police Academy in UK were recruited as participants. All participants recruited reported normal colour vision. Although, they have previously been using tools such as MS Excel for data analysis, none of the participants had any previous experience of working with a tool similar to DOTS, per se.

Procedure: Prior to the study, the participants were presented with five self-paced training video demonstrating the functionality of DOTS, after each of which was a 2 minute gap to practice the concepts learned on the DOTS software. Its main aim was to familiarize the participants with the usage of DOTS software. The main experiment consisted of three scenarios comprising four, three and two tasks, respectively, relevant to the everyday analysis of crime data such as *"Describe the distribution of Burglary crimes during night time in Osterley".* The scenarios were of varying difficulty levels with Scenario 1 being the easiest and hence presented at the beginning of the *within-subject design* experiment to all participants. Thereon, we presented the remaining two scenarios in complete counterbalanced order in order to reduce any learning effect in participants gained from the previous scenarios. At the end of each task, the participants were subjected to a semi-structured interview which enquired about questions based on their performance in the task including, but not limited to, *"please describe the cues you used to answer this task"* and *"did you notice any unusual patterns while analyzing the data in this task?"*

Throughout the experiment, the participants were encouraged to make use of the think-aloud protocol as means of gathering data for qualitative analysis so that it's easier to analyze their thought process and trace their decision-making process. In addition to it, methods such as user observation (video capture and field notes taken), semi-structured interviews and questionnaires focusing on the system were used as data collection methods. Participants were informed of the study procedure, and gave consent for the recording of software screen and audio. Multiple Cognitive Task Analysis (CTA) methods were used to extract and understand the participants' decision making process during the tasks. The sessions lasted between 2 to 3 hours with no time limit imposed.

Setup: The experiment took place at the usability lab in police academy. Two back-to-back monitors were placed for each participant, one for running the DOTS software in full screen mode and the other one for experimenter to observe the participant strategy for working through the tasks and simultaneously produce self-notes for later analysis. All the participant interaction with the software were saved in form of screen capture and textual notes. Responses from sense-making were gathered in word files. Participants were provided with A4 sheets and access to MS Office in order to assist their performance throughout the tasks.

7 RESULTS

In this section, we first present the resultant of the user study conducted. We present the overall performance summary of participants in the tasks and the problems encountered by them. We then describe the strategies that were observed as employed by participants during task analysis and trace them to the design protocols established during designing of DOTS which helps the analyst in applying cognitive strategies successfully and thus expedite insight generation.

7.1 Task Performance Summary

All perception and cognition task from all scenarios had an associated ground truth and solution, respectively, that was validated by the study designers. Accuracy of each task was assessed by grading responses with four possible outcomes:

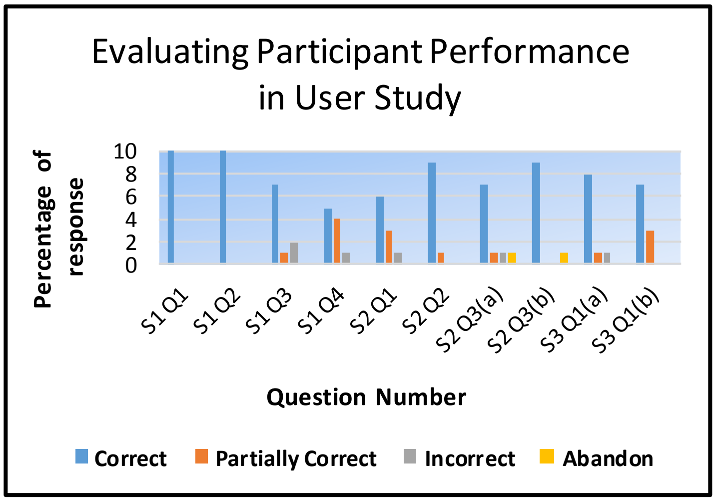
a) Correct – all relevant information found;

b) Partially correct – some relevant information found;

c) Incorrect – no relevant information found;

d) Abandon – unable to find any relevant information;

Figure 7.1 below shows a statistical graph for each of the tasks' performance.



Summary – As is evident from the graph above, each of the questions were answered correctly by more than half the participants with maximum correct responses given for S1Q1 and S2Q2 followed by S2Q2 and S2Q3 (b). Only 1 participant (P4) abandoned two of the questions due to paucity of time. The number of incorrect or partial responses are greater in the early scenarios as the participants had just started to familiarize themselves with the software. The incorrect responses have been mostly due to failure on participant's end to look at the entire dataset and instead focus only on a subset of crime records. The partially correct responses were those where participants struggled with the software usage and hence refrained from going into further depths to gain insight. Instead, they focussed only on the task in hand and provided bare minimum details such as Who, What, Why and Where.

7.2 Strategies Identified

A summary of 21 cognitive strategies identified from the data analysis is shown in Table 1.

We wanted to address the following research questions:

RQ1: How does DOTS enable comprehension of large dataset of varying features without compromising on data individuality?

RQ2: How does DOTS support analysts with their unique and varied style of analysis?

RQ3: How does DOTS motivate further probing into data?

RQ4: How does DOTS help contradict false intuition?

RQ5: How does DOTS promote profiling?

RQ1: DOTS allow representation and interaction with each data point as individual objects

A significant feature of DOTS which separates it from other crime visualizations available is the ability to present each crime record as an individual object. In DOTS, each crime record is represented using a unique dot, each of which represents a full crime report in the form of INVISQUE cards [6] which can be viewed by pulling out the dots into the space on the screen. Each card contains detailed information about the crime such as date of occurrence, offence category, district along with the complete address and postcode of the crime and a brief description of the offence.

The analysis of participant data from the study revealed that the ability to interact with each crime record individually in a visualization allows the analysts to compare them on-the-go, analyze them to gain insight and make preliminary decisions. Following this, they could later delve deeply into connecting crimes and finding patterns across the entire dataset, if the need be, to learn about the nature of crime and its occurrence over a period of time as more than often, understanding of each piece is necessary to the understanding of the whole picture.

Participant P4 was presented with the task of analyzing the NFIB crimes in other years, given a visualization showing 2 of them in 2006. P6 began with the analysis of those 2 crimes first by pulling out the records and found both of them to be from the district Sandlebar but varying addresses. The description of the MO in the pull-out cards suggested to the participant that *"…. it could be the same fraud (gang of offenders) in different place",* as a result of which, they were intrigued to check other years "*in case they (offenders) must have continued in other areas as well".* This further helped them ponder upon an intriguing question as to *"why there are five [NFIB] crimes in a year when there are only 1 or 2 on average?".* For the same task, participant P8 followed the similar analysis pathway of pulling out records and was able to determine that *"all the crime in March occurred on the same date"* which was an unexpected result triggering deeper investigation into the area. The subsequent question that followed was:

“Are these crimes linked or is it just an unlikely co-incidence? Were the frauds committed by the same group of offenders?”

RQ2: DOTS program is compatible with both: analytic and holistic cognitive style preference.

It is perceived that people differ in their individual preference for approaching analysis. They engage in either context-independent analytic perceptual processes by focusing on a salient object independently of its context or attend to relationship between the object and the context in which the object is. Thus, they have either more analytical or holistic approach, respectively. The former cognitive style allows to approach data by processing information in an ordered, linear sequence. The latter enables viewers to view the whole situation at once, which than helps to synthesize available information [18].

From the analysis of the participant data, we observed that DOTS program caters to both of the aforementioned approaches and helps analysts to draw accurate conclusions regardless of the preference for analytic or holistic approach. For instance, in S1 Q4, participant (P10) preferred to start analysis from a detail to search for a particular crime:

"P10: First I want to do is to choose the ‘violence against the person'".

When the participant progressed analysis, he made an inference; primarily by defining each piece of data, followed by summarizing them into a final conclusion.

"P10: I can see in February we have 2, in March 1, lots more in April, and in May and less in June. So, what I can see is that the most busy months with the violence against the person in April and May. That’s what I want to look."

Another participant (P9) wanted to initiate analysis by looking at the big picture of a given problem – to see how crimes are distributed throughout the whole year:

7 "P9: SO, I need to move to 2006? Because it is 2009 now."

After further analysis, the participant defined appearing trend just in general terms.

28 "P9: Ok, so in terms of trends it peaks in May and in June time.".

RQ3: DOTS provide cues and directions to further probe the data.

In most criminal intelligence investigations, it isn't apparent for the insights be to tacit/obvious. Thus, they work their way through the analysis in small increments. DOTS facilitate such an analysis for allowing room for multiple viewpoints using the five visualization views supported by it.

At the beginning of each task, participants playfully sift through the data to explore and identify the view that would best suit their requirements. In one of the tasks, Scenario 2 Q1, Participant P9 reported that before starting an analysis, she has a complete picture in her head, of how the final visualization (that would help her answer the task at hand) would look like.

"In my head I want to go to Bretton Fields and then compare the column of April and Nov."

Thus, DOTS acts as a medium to present the participant's vision of the kind of data and how they would want to have it displayed.
15 "P9: How to change the rest of the bars? I need to change it to the type of offence [pointing the bar with displayed days of the week]"
51 "P9: I want to see what kind of area the crimes appears and time [06:03]"

In addition to this, DOTS also help visual cues to surface that guide the analyst's attention towards relevant aspects of the crime. For instance, task 3 in Scenario 1 required the participants to *"describe the patterns of the Violence Against Person (VAP) crimes in 2006 over the months"*. Participant P2 started with the rudimentary organization of data where each row corresponds to the year and the columns reflected the months of the year. Highlighting of the VAP crimes in the contingency table view (month X offence) revealed that maximum number of crimes occurred in April. Following this observation, P6 further wanted to investigate: *"which of these districts in April were most affected".* P2 then switched to the segmentation view with segments corresponding to Districts to comprehend that in April *"… (there are) majority (of the crimes) in Bretton Fields but Sandlebar has a consistent level of offending"*. P2 was then curious as to what period of the day these crimes took place in Bretton Fields so that vigilance can be increased. Again the participant opted for the segmentation view with each segment as POD (Period Of Day) to reveal that night was the peak time for VAP offence.

The above example depicts how DOTS assist the analysts to further investigate and reason about the observations in-hand. It helps them better explain what is happening by means of narrative story-telling and can question their own thinking so that the analysis is not biased. This technique is also backed up with the feedback from P2 which states:

“.. it [DOTS] makes it easier to break down into finer levels of analysis. Moreover one can change focus quickly to what you really want to find and identify it quickly”.

It is well known that humans perceive different insights from the same visualization and criminal intelligence analysis is no different. Since a number of analysts work on the same data, it becomes necessary and relevant to address data from multiple viewpoint to reduce any bias or error.

For participant P9, changing the view of data caused a significant change in the perceived data.

31 "P9: So it doesn’t peak in May, don’t it or April? OMG I am looking at the wrong peak!"

RQ4: DOTS paves way for analysis of intuitive hunches

It is rarely ever the case that the insights from crime analysis are presented in a platter for analysts to observe. It has to be deeply dug into and this is difficult since there is no predefined direction to work in. Hence, in criminal intelligence analysis, analysts usually indulge in intuitive hunches [3]. While these may not directly lead us to a concrete or correct solution [19], these are considered “diagnostic tools” in the investigation process, particularly in “decision making and in identifying ‘strangeness’, that is, identifying anything that stands out as unexpected, and that does not quit fit into the ‘normal’ pattern.”[17].

57 „P6: Which is what we are trying to get away use gut instinct, is the worse one. Gut instinct ... it is obviously that .. he is doing it because he must be doing it, and the intelligence does not exist. So this is a problem. It does work, in occasions it does work, but most times it does not“

Despite that, DOTS do not condemn the use of intuition altogether, but instead, uses it for progressing data analysis [20]. The design of DOTS is such that it allows analysts to progress towards intuitive hunch but also alert them accordingly when intuition indicates a wrong path.

51 “P6: A lot of it comes from hunches. if you don’t follow your feeling of where it is going you can tie yourself down with tons and tons of data in one dataset.“

The result analysis from DOTS showed that participants engaged in explorative searching of the data in beginning of each task when it wasn’t easily evident what they must be looking for, to begin with.

"P9: I need to play it a little bit just to figure out. [Participant switch ON and OFF: split column to bars]"

Following this, they gathered new insights and made deductions based on their analysis for further detailed investigation.

P8 S3Q1b "are these crimes linked? Were the fraud committed by the same group of offenders? We need to look at it."

RQ5: DOTS promotes profiling

Analysts engage in descriptive and predictive offender profiling to identify likely suspects and analyze patterns that may predict future offenses and victims. The analysis depicted that participants could detect deviation, if any, from the general trend of criminal offender based on their prior knowledge and thus detect anomalies.

For instance, Participant 5 described the observed pattern with reference to the *Profiling* that is appropriate for the *Crime Type* i.e. *Violence Against The Person* as it is associated with *unorderly behaviour* which is more common with high levels of alcohol consumption over the festive periods such as December/January or large social events such as a World Cup. This indicates understanding/knowledge of external influences/factors and how that could impact the observed output of the dataset. The participant explains the peak in April/May with reference to the knowledge of these external factors, but caveats it as a possibility which requires further investigation.

(P5:06.56) "…Okay, so what I can then see is that I've got definitely a grouping of offencing going on in April and May… which makes no sense at all… unless you've got something like the world cup happening… can't understand why we haven't gotten a peak going on in December when most people are going out and getting drunk and that's when we always have a peak going up… that is sort of challenging of what I'd normally expect to see in terms of that."

Another participant, P7, tried to reason about the increasing number of crimes in January as follows.

"Do the maximum amount of crime in January credited to the fact that it was a festive season in December say Christmas?.. With the proper knowledge of the area i.e. rich vs. poor or working class vs. jobless; help detect the reason of the high amount of crime".

The task 4 in Scenario 1 enquired about the *"time of the day that requires additional patrolling for VAP crimes in the month of September"*. Participant P6 made use of the Bar view, with each bar corresponding to POD and columns representing the months. Looking at September, P6 (and P4) utilized his previous expertize to deduce that the crimes aren't always distributed in a similar manner on weekdays as they are on weekends and vice versa. Thus, it triggered him to make use of Segmentation view (segment corresponding to days) to reveal the POD requiring additional patrolling, for each day individually (rather than an aggregate for the entire month). This helped the participant pinpoint Friday afternoon as the prime target of offenders in month of September.

8 Discussion and Conclusion

Our observations suggest that the more analysts "play" with their data, the richer the picture they develop, the more nuanced their insights, and the more confident they are in their analytical conclusions.

"Interaction is the key to exploratory visualization" (Remco Chang, Finding Waldo).

Overall, DOTS was well received by the participants who unanimously concluded that they would benefit from introduction of DOTS in their daily crime investigations and analyses. The subjective feedback reinforces that DOTS as a sense-making tool will provide strong interactive and visual cues which further help the analysts in familiarizing themselves with the dataset in hand.

DOTS help in implementing the Fluidity and Rigour Model of thinking and reasoning processes when engaged in the intelligence analysis task, proposed in Wong (2014). DOTS allow room for creative and critical thinking to form part of rich sense-making. The former is achieved via navigating through different views at the very beginning of investigation analysis to create a story as the analysts are constantly trying to explain the situation by its re-construction using pieces of data and from inferential claims. The objective of the latter helps in solidifying the claims made in the story and this is where DOTS can provide cues as starting point to begin an investigation.

9 Acknowledgement

We would like to thank Phong H. Nguyen, Middlesex University for creation of the DOTS software. This study has been approved from Middlesex University Research Ethics Committee.

10 References

[1] Wong, B. W., & Varga, M. (2012, September). Black holes, keyholes and brown worms: Challenges in sense making. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 56, No. 1, pp. 287-291). SAGE Publications.

[2] Brown, E. T., Ottley, A., Zhao, H., Lin, Q., Souvenir, R., Endert, A., & Chang, R. (2014). Finding waldo: Learning about users from their interactions. Visualization and Computer Graphics, IEEE Transactions on,20(12), 1663-1672.

[3] Wong, B. L. (2014, September). How Analysts Think (?): Early Observations. In Intelligence and Security Informatics Conference (JISIC), 2014 IEEE Joint(pp. 296-299). IEEE.

[4] Plaisant, C. (2004, May). The challenge of information visualization evaluation. In Proceedings of the working conference on Advanced visual interfaces (pp. 109-116). ACM.

[5] Klein, G. (2013). Seeing what others don't: The remarkable ways we gain insights. PublicAffairs.

[6] Wong, B. W., Choudhury, S. T., Rooney, C., Chen, R., & Xu, K. (2011). Invisque: technology and methodologies for interactive information visualization and analytics in large library collections. In Research and Advanced Technology for Digital Libraries (pp. 227-235). Springer Berlin Heidelberg.

[7] Crnkovic, G. D., Duran, J. M., & Slutej, D. (2010). Content Aggregation; Visualization and Emergent Properties in Computer Simulations. In SIGRAD(pp. 77-83).

[8] Bargiela, A., & Pedrycz, W. (Eds.). (2008). Human-centric information processing through granular modelling (Vol. 182). Springer.

[9] Yi, J. S., Kang, Y. A., Stasko, J. T., & Jacko, J. A. (2008, April). Understanding and characterizing insights: how do people gain insights using information visualization?. In Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization (p. 4). ACM.

[10] Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). Readings in information visualization: using vision to think. Morgan Kaufmann.

[11] Kodagoda, N., Attfield, S., Wong, B. L., Rooney, C., & Choudhury, S. (2013). Using interactive visual reasoning to support sense-making: Implications for design. Visualization and Computer Graphics, IEEE Transactions on, 19(12), 2217-2226.

[12] Robertson, G., Fernandez, R., Fisher, D., Lee, B., & Stasko, J. (2008). Effectiveness of animation in trend visualization. Visualization and Computer Graphics, IEEE Transactions on, 14(6), 1325-1332.

[13] Krautli, F. and Davis, B. (2015). Revealing cultural collections over time. Computer Graphics and Visual Computing.

[14] Stramp, N., & Wilkerson, J. (2015). Legislative Explorer: Data-Driven Discovery of Lawmaking. PS: Political Science & Politics, 48(01), 115-119.

[15] Roberts, J. C. (1998, July). On encouraging multiple views for visualization. In Information Visualization, 1998. Proceedings. 1998 IEEE Conference on(pp. 8-14). IEEE.

[16] Tsetsos, K., Chater, N., & Usher, M. (2012). Salienuce driven value integration explains decision biases and preference reversal. Proceedings of the National Academy of Sciences, 109(24), 9659-9664.

[17] Rodseth, Ingerid (2007). The role of intuitive feelings as a diagnostic tool in the criminal investigation process. ISBN. Workshop proceedings of Ethnographies of Diagnostic Work April 17-18, 2007, Lancaster, UK.

[18] Hannond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. Systems, Man and Cybernetics, IEEE Transactions on, 17(5), 753-770.

[19] Kahneman, D. (2011). Thinking, fast and slow. Macmillan.

[20] Informatics Conference (pp. 296–299). IEEE. Gerber, M., Wong, W., Kodagoda, N. (2016) How Analysts Think: Intuition, Leap of Faith, Insight. Unpublished

11 Appendix

	Strategy identified	Description of Strategy	Study the participant did		
			S1 Q3	S2 Q2	S1 Q3a
1	approach to the problem - generalization to specialization		P09	P5,P10	…
2	approach to the problem - specialization to generalization		P10	…	…
3	Pattern recognition	Find patterns related to crime such as district, year or time of occurrence	P09	…	…
4	Exploration	navigation through the data to learn more about it	P10, P09	…	…
5	Second chance - Return to the initial exploration point	If participant stuck in the analysis, they can return to initial point of enquiry to start afresh, much like real-life example of trekking through an unknown forest	P10	…	…
6	Understanding the crime to understand motives, MOs, patterns and WHYs (hovering and opening record)	Participant reads/examines the crime records and the information associated with it to learn about nature of crime.		…	…
7	broadening view		P10, P09	…	…
8	define each piece to comprehend the whole picture		P10	…	…
9	Get understanding of the screen		P10, P09		
10	Unwanted data elimination from screen	Participant prefers to maintain only the intentional data on screen to avoid any distraction while analysis or to focus their attention to a particular scenario.	P10, P09		
11	Insight as a result of unexpected outcome		P10, P09		
12	insight as a result of inplace transposition		P10, P09		
13	pre-defining final view/visualization	Participants have a mental picture which aptly describes their preferred layout (view) of the dots in the final visualization which will help them answer their questions about the data.	P09		
14	profiling		P09		
15	observe significance of patterns in context of entire dataset	Participants look at the graph atop each column or selected number of crimes atop each row to compare them with rest of the data set to see whether they are significant and requires analyst attention	P09		
16	data comparison by Color grouping	Participants color sort the crime records in a column with respect to the year, district, offence,etc such that all the crimes belonging to a particular year, district, offence, etc are grouped together in the column and provide visual clarity.	P09		
17	distinguish relevant data to facilitate further analysis	Participants highlight the relevant crimes by placing a cross over them to ensure focussed attention to them and segregate them from others.			
18	defining data with respect to general statistics terms	DOTS allows easier identificated of statistical metrics such as maximum, minimum, mode and mean number of crime records in any category.	P09		
19	Contradiction of Intuition	Participants often have a pre-conceived notions regarding the dataset and many a times, interaction with DOTS help in surfacing the otherwise hidden facts. For ex. categorizing Osterley as a safe district because of observing a decrease in NFIB crime rate whereas analysis reveal an increase in all other crimes in the district			
20	Laddering	DOTS provide participants the ability to find cues during their analysis which triggers new questions and queries about the current information presented and helps them dig deeper for better understanding of the data.			
21	Facilitate the five Ws questions - What, When, Where, Who, Why and How	It is very convinient to answer these questions using five views of DOTS and the coloring, filtering and selection features.			