MetricsVis: A Visual Analytics Framework for Performance Evaluation of Law Enforcement Officers

Jieqiong Zhao, Abish Malik, Hanye Xu, Guizhen Wang, Jiawei Zhang, Chittayong Surakitbanharn and David S. Ebert
Purdue University, 500 Central Drive, West Lafayette, IN 47907,
Phone: 765-496-3747, Fax: 765-494-1028
Email: {jieqiongzhao|amalik|xu193|wang1908|zhan1486|csurakit|ebertd}@purdue.edu

***Abstract*—Limited resources and increasing costs require law enforcement agencies to develop effective methods for measuring and evaluating officer performance. The methods enable law enforcement to be more effective in their event planning, resource allocation, decision making, and community policing efforts. The paper introduces a visual analytics framework for efficient measurement and evaluation of officers' performance through interactive and coordinated visual dialogs. Through collaboration with our partner law enforcement agency, we have developed a comprehensive categorization of offense types utilizing a crowd-sourcing approach. Our system allows end-users to interactively specify the offense types and customize the performance metric based on their domain knowledge and policing priorities. The performance scores for each officer are then visualized based on a matrix representation. The representation supports a rich set of interactions including selection, filtering, ranking and correlation to allow end-users to supervise and refine the performance evaluation process. With our system, end-users can explore the activity patterns and performance trends for either a large group or an individual, and identify critical factors that help to improve the operational decision making process. To demonstrate the proposed approach, we present two case studies and provide domain expert feedback.**

I. Introduction

Resource allocation for law enforcement agencies is a critical problem due to limited resources and the time sensitivity of emergency cases. Typically, supervisors have sufficient knowledge about the officers with whom they frequently interact. However, they may neglect subtle characteristics of officer performance or only be familiar with certain aspects of an officer's effectiveness on certain types of cases. Instead of spending time to read every individual crime case report, our system shows the performance of selected officers using a common tabular visualization, providing an overview of the entire workforce and individuals' performance for each category of crime. Dynamic selection of officers, multiple modes of sorting and scrolling through time, provides an interactive performance assessment and scheduling tool. Understanding employee performance at a micro-scale can better facilitate the process of resource allocation and cooperation between police officers and supervisors. Officers may develop particular professional affinities or skills in dealing with specific types of events through experience or their personalities. Understanding these specialties is critical to further improve their assignment, performance across all cases, and community collaboration.

Ranking for sports team and players is popular due to its clarity and easy comparison nature [1]. However, as far as we know there is no existing ranking or comparison tool for law enforcement agencies to assess the performance of officers. When supervisors need to write a report about their teams, they have to go through all crime case reports and recall the performance of officers. To leverage the evaluation process and improve resource allocation, we designed metrics to quantify the amount of work accomplished by officers. With our law enforcement partner, we developed a comprehensive tool that can dynamic query crime events from a database and demonstrate with intuitive visualization.

We have developed a visual analytics framework called MetricsVis to provide insight for the working schema of police officers, as shown in Figure 1. MetricVis allows users to measure, evaluate, and compare officer performance through interactive and coordinated visual dialogs. Our system provides a comparison and evaluation environment for the entire dataset using a holistic matrix view, as shown in Figure 1(b). The coherent matrix view supports multiple sorting interactions for comparison between a group of officers and different offense categories. We worked intensively with our partner law enforcement agency to develop the multivariate attributes and their weights that contribute to the assessment criteria. We applied 27 offense types as attributes and adopted a survey involving both citizens and police officers to categorize the weights for each offense type. Alternatively, end-users can develop a set of customized weights in order to adaptively tailor the analysis based on domain knowledge, or recent events in a particular region. Through the interactive exploration of the actual crime events handled by law enforcement agencies, end-uses can discover the activity patterns and performance trends and further identity the critical factors that impact the operational decision making.

In the rest of the paper, we describe the related work in Section II. We explain the system design choices in Section III. In Section IV describes the implementation of our visual analytics system interface and highlights the supported

978-1-5090-6356-7/17/$31.00

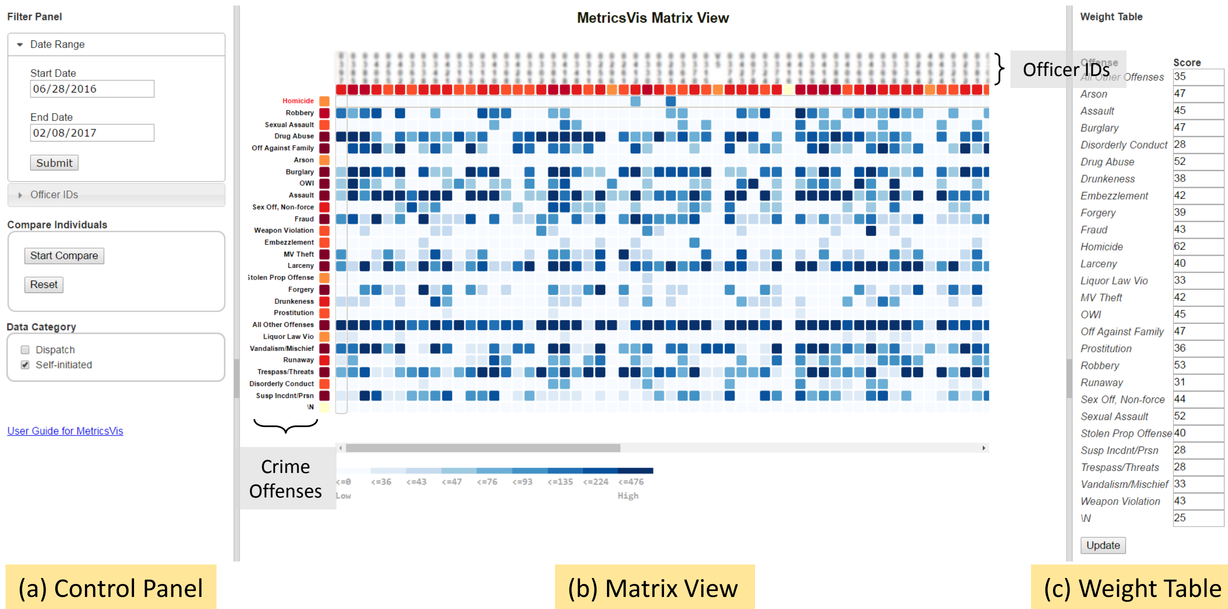


Fig. 1. MetricsVis visual system overview. (a) Control panel with selection of time range, interested officers, and behavior type. (b) Matrix view shows the multivariate attributes (27 offenses) contribute to the overall performance of officers in a holistic tabular view. (c) Weight table for dynamically adjustment on weights of 27 offenses.

functionality and interactions. Case studies are presented to show the capability of our system in Section V. Section VI present the feedback from our law enforcement partner and addresses the strengths and limitations of the work. Finally, we conclude our work and point our future research directions.

II. Related Work

A law enforcement incident may involve multiple offenses as classified by NIBRS [2] and UCR [3]. Patrol officers come across various criminal incidents and play an important role to improve the safety of community. Oettmeier and Wycoff [4] introduced an evaluation model for the behavior of community policing. It emphasizes the contribution difference between dispatched crime cases versus the self-initialized proactive cases. For dispatched scenario, police officers rapidly respond to 911 calls and provide services to citizens who initialed their requests. For self-initialized scenario, patrol officers discover the incidents during their shift and proactively respond to incidents. As previously mentioned, we utilize 27 offense types as the attributes to contribute to performance measurements based on our discussion with law enforcement partner. Furthermore, we distinguish the motivation (dispatched vs. self-initialized) of dealing with a crime case. For visualizing multidimensional data [5], parallel coordinates [6] and matrix views are familiar to end-users. Parallel sets [7] present a particular design for categorical dimensions. Tufte [8] studied the representations of quantitative data and Mühlbacher and Piringer [9] also use discrete regions to present continuous quantitative data.

Tabular visualization [10] possesses the properties of compactness and straightforwardness while showing all multidimensional data in one view. Each dimension appears as a small equalized region in a matrix. Besides sequential colors, glyphs [11]–[13] are often used to represent categorical data. Loorak and colleagues [14], [15] proposed the concept of integrating a tabular visualization for multidimensional data with online popular visualizations. Our system, in contrast, displays the whole dataset using tabular visualization with the additional tuning factors of dynamically updated weight table and fluid interaction that provides end-users more freedom to manipulate the matrix and easily gain insights into the individual and squad level performance.

III. Design Rational

The primary goal of our system is to provide a straightforward and intuitive visual analytics tool to help our law enforcement partner understand and evaluate the activity patterns and performance trends of their team members. To fulfill this goal we summarize and rationalize our design choices as follows:

- **Ranking of all officers** To support the overall evaluation of performance, we show the ranking of all officers. A tabular visualization is applied to show the entire dataset in one view. To speed up the comparison of a large number of quantitative values, we map the continuous scores into sequential colors. Apart from this, we support simple ranking interactions and the interaction methods are explained in Section IV-A.
- **Manipulating of evaluation metrics** For each multidimensional data entry, the contribution of each attribute to the outcome varies. The weight can decide the importance level of an attribute and has a significant impact on the assessment process. We therefore support the end-users to manipulate and customize the weights on a weight table. The changes in the weight table are instantly reflected on the tabular visualization.
- **Comparing data entries of interest** Despite powerful ranking of all data entries, we still need to compare

some data items that stand out. We want to compare these data items under the context of the entire dataset. A comparison mode is designed to facilitate comparison of a small group of data entries. We aligned the selected data entries to the left of the tabular visualization. More details are explained in Section IV-C

In the next section, we describe the implementation of our system and how we address our design requirements in more detail.

IV. Visual Analytics Interface

Figure 1 shows the multiple linked visual dialogs, including a Control Panel, a Matrix View, and a Weight Table of the system. The visual dialogs aim to provide a holistic matrix view with sequential colored cells to show an overview of the whole dataset. An interactive weight table allows dynamic updating of the metrics by end-users. The control panel allows filtering by time range, officer IDs, the comparison between selected officers, and behavior type.

Through law enforcement feedback, we recognized the need for providing more weight for self-initiated calls as a depiction of proactive behavior. Accordingly, we allow users to filter their data by self-initiated and dispatch-initiated calls. The different dialogs are highly coordinated and provide a rich set of interaction means for analysts to explore the dataset.

The system is developed based on a server-client architecture. The client side is mainly developed based on JavaScript, HTML, and visualization libraries such as D3JS [16]. The system retrieves the data in the back-end from the record management system so that up-to-date information is available.

A. Matrix View

The Matrix View serves as the main workspace for exploration and analysis. In the Matrix View, each row represents a specific offense type, while each column represents an officer, denoted by his/her officer ID. Each matrix cell shows the product of the frequency of the corresponding offense type responded to by an officer and the weight of that offense type (specified in the Weight Table). A sequential color scheme [17] based on blue has been applied to encode the magnitude of the value.

Besides the cells in the matrix, we also encode the volume of the aggregated offense types and officers in the left and upper side of the matrix that are aligned to the corresponding row or column, respectively. A sequential red color scheme has been applied to encode the magnitude of the aggregated value. The visual encoding provides an intuitive visual summary for different offense types or officers, where darker color corresponds to higher values. Users can hover over an individual cell to obtain more details regarding the value.

Two effective types of sorting interactions are supported directly in the matrix view in order to discover performance trends. The first is sorting based on aggregated value of officers or offenses, and another is sorting based on an individual officer or one particular offense. A mouse click on any red cell in the first row that is adjacent to officer IDs will generate the sorted result of the total score assigned to each officer in descending order from left to right. Similarly, a mouse click on any red cell on the first column that adjacent to offense types will produce the sorted result of offenses in descending order from top to bottom. Manipulating the sorting functions for both aggregated values leads to dark cells on the upper left corner. Secondly, sorting for individual officer or one particular offense is a simple click on the text label.

B. Weight Table

TABLE I
EIGHT PARAMETERS IN EVALUATION OF EACH OFFENSE.

Economic Loss to Victim
Economic Loss to Group
Economic Loss to Government
Economic Loss to Private Organization
Impact on Culture
Impact on Victim's Mental Well-being
Impact on Victim's Physical Well-being
Risk to Officer's Life

The value of each matrix cell equals the product of the frequency of offense and its weight. The weights of offense plays an important role when evaluating the performance. We assign an initial weight value for each offense with respect to the survey results provided by citizens (30 participants) and law enforcement officers (59 participants). For each offense type, participants were asked to grade on a scale from 0 to 10 for eight parameters that are shown in table I. Participants skipped some survey questions that they found difficult. For each offense, we averaged each parameter on the number of the responders, and then added up the eight parameters. The weight of each offense ranged from 0 to 80. The result of weights of the 27 offenses is shown in Table II. For simplicity, we only used integer weight values in our system. The survey results provided by citizens and law enforcement officers exhibit some notable difference, since each person has different judgment of an offense. Yet, they all agreed that homicide was the top severe offense. End-users can tune the metrics for weight according to their domain knowledge. The visual output of the matrix view changes accordingly with the updated weight. Moreover, the sorting function generates different results.

C. Control Panel

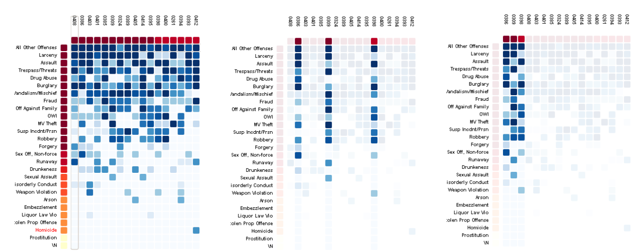
Basic filtering functions are provided in the control panel, such as filtering by time range, officer ID, and self-initiated vs. dispatch-initiated. The system is initialized with the crime records for the last month. The database updates the crime records every six hours by data feeds from our law enforcement partner. The role of police officers is different when they are dispatched to an emergency call or proactively self-initiate an activity. Both dispatched and self-initiated activities are critical and not interchangeable. Therefore, the tool allows the selection of dispatched and self-initiated data individually or as a union. Our partner suggested to assign more weight on self-initiated activities when union operation is applied.

978-1-5090-6356-7/17/$31.00 ©2017 IEEE

TABLE II
SAMPLE SURVEY RESULT FOR WEIGHTS OF 27 OFFENSE CATEGORIES.

Offense Category	Weight Score
Homicide	61.97
Robbery	53.47
Drug Abuse	51.94
Offense against family	47.39
Arson	46.93
Burglary	46.62
Operating while intoxicated (OWI)	45.38
Assault	45.18
Fraud	43.43
Weapon Violation	43.20
Embezzlement	42.43
(Motor Vehicle)MV Theft	41.66
Larceny	39.73
Stolen Property Offense	39.51
Forgery	39.41
Drunkenness	37.87
All other offenses	35.00
Liquor Law Violation	33.26
Vandalism/Mischief	32.59
Runaway	30.96
Trespass/Threats	28.08
Disorderly Conduct	28.03
Suspicious Incident/Person	27.81
Curfew Violation	24.60

Our system also provides comparisons between particular groups of officers. For instance, we can compare the performance of the officers that have the same patrol time slots but serve in the different areas. After applying the sorting function, the officers of interest may not appear close to each other. We can freeze the current sorting result (Figure 2(a)) and start a comparison mode shown in Figure 2(b). We can then select a group of officers and align them together (Figure 2(c)) to further compare them next to each other. Meanwhile, we can adjust the order of officers by our preference.



(a) Sorting on aggregated values. (b) Comparison mode. (c) Aligned selected officers

Fig. 2. Example of comparing groups of officers. (a) shows them sorted based on the total value of each officer's performance and the total of each offense. In (b), comparison mode is active and a few officers are selected for study. The respective officers are shown in (c).

V. CASE STUDIES

In this Section, we present two case studies to demonstrate the capability of our system in terms of more effective measurement, evaluation, and comparison of the officers' performance. Figure 3 demonstrates the analysis process and the corresponding interactions. Both case studies use a sample dataset of self-initiated crime records from Sept. 1st, 2016 to Feb. 1st, 2017 (Figure 3(a)). As an exploratory example, we apply the sorting function to compare the performance of individual officers in the first case study. In the second case study, we apply the k-means clustering algorithm to group officers by skill proficiency as a way to identify and assign officers of particular skills to the same shifts.

For the first case study, we select all patrol officers for the time period using the filter panel (Figure 3(b)). Second, we select all officers on patrol duty who responded to self-initiated case (Figure 3(c)). Third, we sort all the officers by their total performance score by clicking one of the red cells on the top row (Figure 3(d)). As a result, the officer IDs are sorted in descending order from left to right. Similarly, we can sort the aggregated scores of 27 offenses by clicking on one red cell on the far left column (Figure 3(e)), which would result in offense types sorted in descending order from top to bottom.

Based on the sorting result, we observer that officer "5818" has the best performance score of 4779, while the second ranked officer "5873" has a score of 3130. The score of the best officer is almost 50% better than the score of second ranked officer. We compare the performance of the first and second ranked officers (Figure 3(f)) by sorting based on individual officer (Figure 3(g)).

We click the ID label text "5818" to sort by the officer's performance in 27 crime categories. The top six crime offenses that officer "5818" came across were 1) all other offenses, 2) trespass/threats, 3) larceny, 4) vandalism/mischief, 5) burglary, and 6) drug abuse. The second ranked officer "5873" spent the most time on these offenses: 1) all other offenses, 2) burglary, 3) larceny, 4) assault, 5) trespass/threats, and 6) drug abuse. Both officers spent the most time on a generalized category "all other offenses" that includes violations of traffic regulations (e.g. speeding, ignore of stop signs, improper parking), animal related violations, loud and unnecessary noise, and others. Comparing the score of "all other offenses" for both officers, we find that officer "5818" has a score of 2100; almost twice the score of officer "5873"(1155).

Since a large amount of work for patrol officers is categorized as "all other offenses", the weight of "all other offenses" is reduced to zero to amplify the differences in the remaining 26 categories. After the removal, the total score difference between the two officers drops from 50% to around 35%. In comparing the other crimes, both officers dealt with roughly the same total number of larceny and burglary cases. Officer "5818" handled more cases of motor vehicle theft, drunkenness, and runaways. In contrast, officer "5873" handled more assault cases. Through inspection of raw crime records (Figure 3(h)), we notice that officer "5818" worked more night shifts (6pm to 6am) while officer "5873" worked more day shifts (6am to 6pm), which may explain the differences in handled cases. In general, the basic sorting functions can lead to effective comparisons of officer performance due to the tabular layout and sequential color scheme of the matrix view.

The second case study demonstrates the tool's potential use to assign shifts to patrol officers based on their skill

978-1-5090-6356-7/17/$31.00 ©2017 IEEE

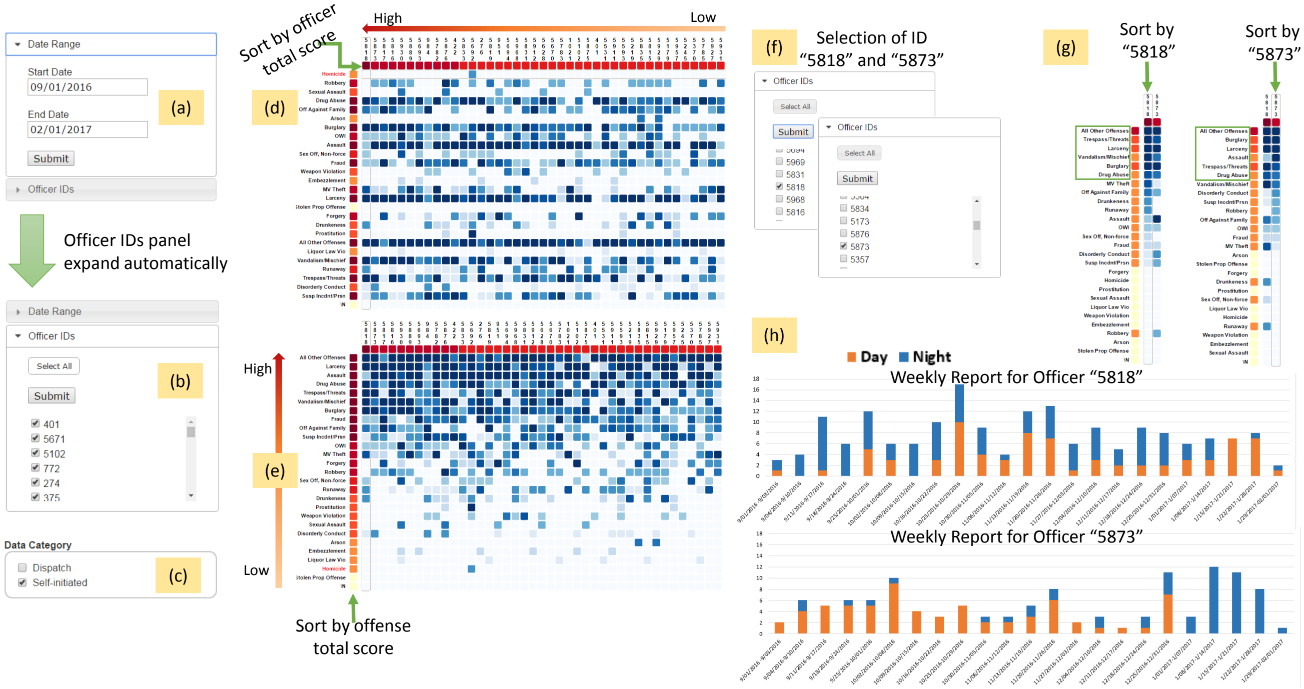


Fig. 3. Demonstration for first case study. (a) Input for time period. (b) Selection of all officers involved in crime records during given time period. (c) Selection of behavior type. (d) Sorting based on total performance score of officers. (e) Sorting based on total score of 27 offenses. (f) Selection of first ranked officer “5818” and second ranked officer “5873”. (g) Sorting based on performance of individual officer. (h) Visualize the weekly frequency of crime handled by individual officer using stacked bar char to compare the day and night shifts.

proficiencies. We apply the k-means clustering algorithm to the 128 officers using the same dataset, period, and self-initiated case filter. For the clustering method, each officer is regarded as one instance of input data. The numeric performance score (the product of frequency of offense by its weight) for 27 offense categories are applied as 27 attributes of that instance. As an example, we clustered all the instances into six clusters and then visualized the six clusters in the matrix view separately (Figure 4).

First, we can use the tool to assign officers from a diversity of experiences and skills to a certain patrol area. If we focus on the top three crime categories of each Cluster, we find that officers in Cluster 1 mostly dealt with larceny, all other offenses, and assault (Figure 4(a)) and officers in Cluster 6 mostly dealt with all other offense, drug abuse, and larceny (Figure 4(f)). Since officers in Cluster 1 and Cluster 6 have different case experiences, we can choose officers from these two clusters to cover the same patrol area to provide the needed spectrum of skills and experiences.

Second, patrol areas with the highest crime incidences may be distributed across a larger pool of officers rather than always assigned to the same officers. For example, there are officers in Cluster 6 that report little to no exposure to self-initiated crime incidents (Figure 4(f), middle to far right columns), while officers in Cluster 1-5 all have exposure to an array of crime incidents. Assigning officers from Cluster 6 with officers in Cluster 1-5 to the same patrol areas may result in improved workload balance across the officer pool.

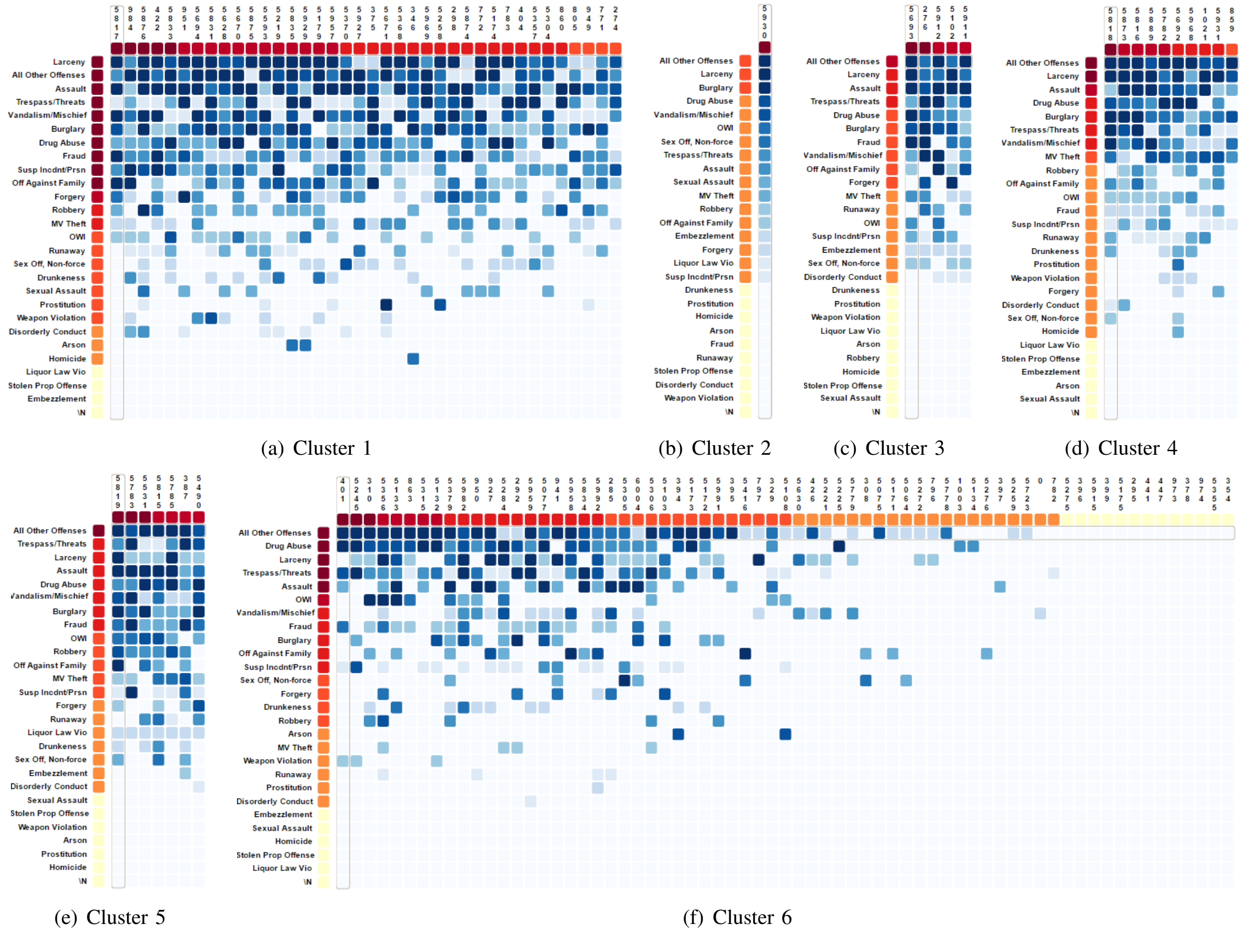
Third, we can concentrate the specialties of officers to particular areas of need. For example, Cluster 1 demonstrates exceptional experience in dealing with larceny incidents. Thus, officers from Cluster 1 can be assigned to areas with high incidences of larceny based on their experience in dealing with this particular crime.

Finally, the previous examples are only hypothetical scenarios to demonstrate how the tool can be used. Actual implementation requires more in depth understanding of the agencies’ operations, which we have gained through domain expert feedback sessions.

VI. Domain Expert Feedback

In this Section, we interviewed several domain experts from our partner law enforcement agencies and present their feedback in terms of the strengths and limitations of our approach. In general, our partner law enforcement officers liked the capability of our visualization system due to its simple sorting functions to compare the performance of all officers. It is easy to find the types of cases an officer spent the most time on and how many cases an officer handled. The sequential color scheme makes it simple to compare the performance score of officers, and the adjustment of weight table makes the system more flexible. However, our current system did not include geographical information, which may reveal why an officer handled a higher frequency of particular cases. We quantified the performance of an officer using the frequency of cases and its weight, but this evaluation method does not show the quality of dealing with the cases. The officers suggested using an extra parameter to define the quality of how an officer dealt with offenses.

978-1-5090-6356-7/17/$31.00 ©2017 IEEE



(a) Cluster 1 (b) Cluster 2 (c) Cluster 3 (d) Cluster 4

(e) Cluster 5 (f) Cluster 6

Fig. 4. Demonstration of 6 clusters.

VII. Conclusion

MetricsVis is a visual analytics framework to evaluate and compare the performance of law enforcement officers to improve resource allocation and community policing. A holistic matrix view shows the overview of the entire dataset. To compare the performance of officers, analysts can manipulate the sorting operations for either aggregated total or individual score of officers and offense types in the matrix view. Experts are able to investigate the raw data as well as adjust the weight parameters to tune the contribution of each attributes.

In our current system, we evaluated the performance of officers using the product of frequency of each offense type and its corresponding weight. We can consider other factors (e.g., division, shifts, portal locations) that also impact performance and extend our system to take in such information. In the future, we plan to apply more sophisticated visualization to combine hierarchical design with current flat tabular design. Also, we can provide users operations to merge and split the fundamental attributes to create customized attributes.

Acknowledgment

This work was funded by the U.S. Department of Homeland Security VACCINE Center under Award Number 2009-ST-061-CI0006.

References

[1] A. Y. Govan, A. N. Langville, and C. D. Meyer, "Offense-defense approach to ranking team sports," *Journal of Quantitative Analysis in Sports*, vol. 5, no. 1, p. 1151, 2009.

[2] L. Addington, "National incident-based reporting system (nibrs)," *Encyclopedia of victimization and crime prevention. Thousand Oaks, CA: Sage*, 2010.

[3] U. C. R. Handbook, "Federal bureau of investigation," *Last accessed on April*, vol. 30, p. 2012, 2004.

[4] T. N. Oettmeier and M. A. Wycoff, "Personnel performance evaluations in the community policing context," *Police and policing*, pp. 57–77, 1999.

[5] T. Jirka, "Multidimensional data visualization," Technical Report (DCSE/TR-2003-03), Tech. Rep., 2003.

[6] A. Inselberg and B. Dimsdale, "Parallel coordinates: a tool for visualizing multi-dimensional geometry," *San Francisco CA*, pp. 361–375, 1990.

[7] R. Kosara, F. Bendix, and H. Hauser, "Parallel sets: Interactive exploration and visual analysis of categorical data," *IEEE transactions on visualization and computer graphics*, vol. 12, no. 4, pp. 558–568, 2006.

[8] E. R. Tufte and P. Graves-Morris, *The visual display of quantitative information*. Graphics press Cheshire, CT, 1983, vol. 2, no. 9.

[9] T. Mühlbacher and H. Piringer, "A partition-based framework for building and validating regression models," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 1962–1971, 2013.

[10] C. Perin, P. Dragicevic, and J.-D. Fekete, "Revisiting bertin matrices: New interactions for crafting tabular visualizations," *IEEE transactions on visualization and computer graphics*, vol. 20, no. 12, pp. 2082–2091, 2014.

[11] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen, "Glyph-based visualization: Foundations, design

guidelines, techniques and applications," *Eurographics State of the Art Reports*, pp. 39–63, 2013.
[12] J. Bertin, "La graphique et le traitement graphique de linformation," Nouvelle bibliothèque scientifique. Flammarion, Tech. Rep., 1975.
[13] M. O. Ward, "A taxonomy of glyph placement strategies for multidimensional data visualization," *Information Visualization*, vol. 1, no. 3-4, pp. 194–210, 2002.
[14] M. H. Loorak, C. Perin, N. Kamal, M. Hill, and S. Carpendale, "Timespan: Using visualization to explore temporal multi-dimensional data of stroke patients," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 409–418, 2016.
[15] M. H. Loorak, C. Perin, C. Collins, and S. Carpendale, "Exploring the possibilities of embedding heterogeneous data attributes in familiar visualizations," *IEEE Transactions on Visualization and Computer Graphics*, 2016.
[16] M. Bostock, V. Ogievetsky, and J. Heer, "D^3 data-driven documents," *IEEE transactions on visualization and computer graphics*, vol. 17, no. 12, pp. 2301–2309, 2011.
[17] C. Brewer and M. Harrower, "Color brewer," 2016. [Online]. Available: http://colorbrewer2.com

978-1-5090-6356-7/17/$31.00 ©2017 IEEE