Security Data Visualization

GIAC (GSLC) Gold Certification

Author: Balaji Balakrishnan <pingbalaji@gmail.com>

Advisor: Chris Walker <chriswwalker@hotmail.com>

Accepted: December 16th 2014

Abstract

The objective of this paper is to provide guidelines on information security data visualization and insights with repeatable process and examples on visualizing (communicating) information security data. Security data visualization can be used in many areas in information security. Security metrics, Security monitoring, anomaly detection, forensics, and malware analysis are examples where security data visualization can play a vital role and make us better security professionals. Security data visualization also plays key role in emerging fields such as data science, machine learning, and exploratory data analytics. There are many uses for security data visualization; so, in order to cover key aspects the paper is categorized in to two parts.

The first category is communicating value. There is a well-known proverb "a picture is worth a thousand words" (Piqua Leader-Dispatch, One Look Is Worth A Thousand Words, 1913, p. 2) which explains this. The problem with traditional metrics is numbers and tables can be daunting and details can be missed easily. Visualizing it will enable the security team to highlight the salient points in the data. Security data visualization enables you to tell a story with the data. Information security is becoming a common topic in boardroom discussions and it is becoming more and more important that the value of information security is communicated to business leaders.

The second category is finding anomalies using security data visualization. One of the key strengths of security teams is access to enterprise log data, meta-data, network traffic data, and netflow data. The challenge is finding and isolating the bad actors from legitimate traffic. The human mind, by evolution, is trained to identify patterns and anomalies using visualization. Security professionals can benefit by visualizing enterprise data to find anomalies and identify patterns which will be helpful in isolating events which might indicate compromise.

Hopefully some of the examples will be useful to generate more ideas in this space and will be a valuable resource for all Information Security practitioners. Once security professionals get an understanding of using security data visualization it will open a whole new world and there is a possibility that this knowledge of security data science will have significant improvement on information security tasks.

1.0 Introduction

Security data visualization can be used in many areas in information security. Security metrics, Security monitoring, anomaly detection, forensics, and malware analysis are examples where security data visualization can play a vital role and make us better security professionals. Till now security professionals were able to survive with Microsoft Excel and similar tools without in-depth knowledge in security data visualization. But security data visualization is becoming extremely important due to big data, machine learning and exploratory data analytics. Due to the volume of data in big data it is extremely impossible to find anomalies using traditional methods. First thing to do after a statistical computation is to understand the data visually. Recent generations of SIEM log collection and correlation solutions use big data analytics. Security data visualization plays a very vital part in analyzing the big data. Data science field is evolving at a rapid pace. Data visualization is important component of data science.

Botnet Visualization

Microsoft's Digital Crimes Unit tapped The Office for Creative Research, a multidisciplinary digital design group based in New York, to come up with new ways of looking at one particular threat: botnets, the global networks of infected computers that cyber criminals enlist to do their bidding. OCR came up with a prototype tool called Specimen Box. Specimen Box offers many views including live display of botnet activity "which can be used to analyze botnet data" ("#005: The Sight and Sound of CyberCrime", o-c-r.org, 2014, para. 3).

Reverse Engineering

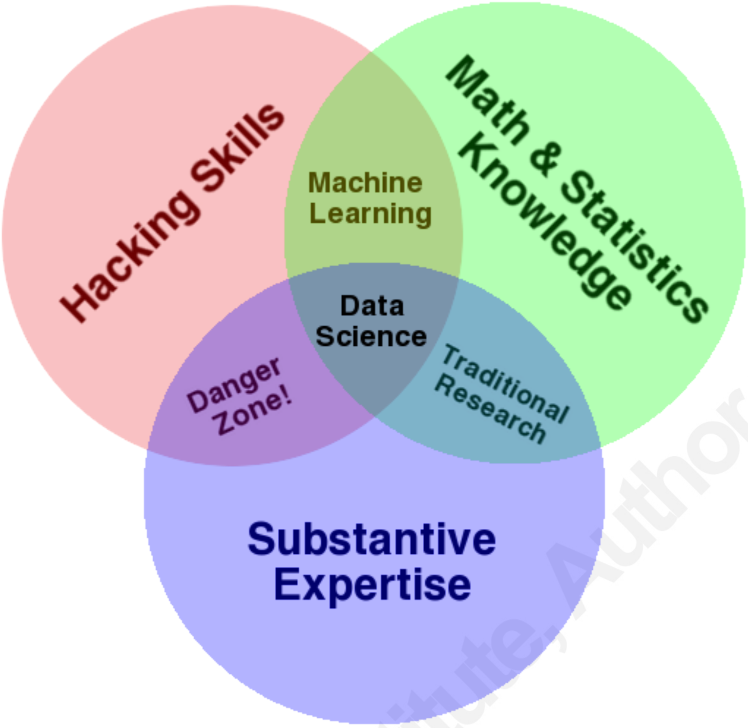
Security data visualization is used more and more in reverse engineering. “In this engaging TED(TED is a platform for ideas worth spreading - http://www.ted.com/) talk, Chris Domas shows how researchers use pattern recognition and reverse engineering (and pull a few all-nighters) using visualization to understand a chunk of binary code whose purpose and contents they don't know.”(Domas, C. (n.d.). The 1s and 0s behind cyber warfare. Retrieved December 15, 2014, from http://www.ted.com/talks/chris_domas_the_1s_and_0s_behind_cyber_warfare, para. 1)

Currently the information security practitioners are just scratching the surface in this area, additional security data visualization magic is captured in Appendix A for inspiring and invoking the curiosity and awe in security practitioners for utilizing the full potential of security data visualization in information security day-to-day jobs.

Hopefully some of the examples will be useful to generate more ideas in this space and will be a valuable skill for all Information Security practitioners. Once security practitioners get an understanding of using security data visualization it will open a whole new world and there is a possibility that this knowledge of security data science will have significant improvement on information security tasks.

2.0 Security Data Visualization Skills

Data science and security visualization require the skills described in the Venn diagram. It is the space where the hacking skills, statistical knowledge and domain knowledge meet.



(Conway, D. (n.d.). The Data Science Venn Diagram. Retrieved November 29, 2014, from http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram)

Substantive Expertise – This is the security domain knowledge, which will enable the security practitioner to understand the data, determine what is expected and find anomalies or metrics from visualization.

Hacking Skills – Hacking skills are the skills from a data scientist language required for working with massive amount of data that should be acquired, cleaned and sanitized.

Math & Statistics Knowledge – This knowledge is critical to understand which tools to use, understand the spread and other characteristics to derive insight from the data.

Security practitioners will be comfortable with domain knowledge and hacking skills. Statistics knowledge is one aspect that security practitioners have to understand to gain insight from data and also to ask the right questions to derive the right security visualization.

One resource for statistical knowledge is an online free course “Data to Insight: An Introduction to Data Analysis” ("Data to Insight: An Introduction to Data Analysis - The University of Auckland - FutureLearn", 2014). This course is a hands-on introduction to statistical data analysis that emphasizes fundamental concepts and practical skills. This course also introduces the tool iNZight.

One aspect google looks at while recruiting engineers is their knowledge on statistics and probability. The reason might be that they need people who understand the basics in deriving value from data. Using statistics, probability in combination with machine learning/artificial intelligence there are lot of predictions based on the data in various fields. Data science field is evolving at a rapid pace. Data visualization is important component of data science. These techniques will soon be applied to information security field for better identification of bad actors.

One of the most important advantages of data visualization is that all the resources on data visualization are publicly available for learning the key concepts.

There are numerous Coursera and eDX courses available for free about data visualization. There is extensive material about R project with numerous examples from various experts. If enough time is dedicated, data visualization tools and R can be learned easily by security analysts.

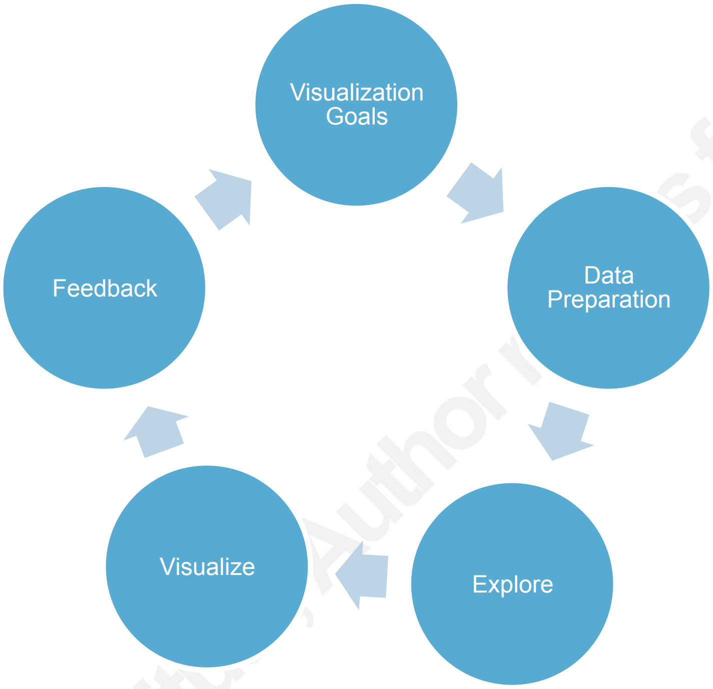
The key advantage for security analysts is that security analysts have access to security data like security metrics data, network traffic data, malware indicators of compromise data, and many more.

Security domain expertise is very important before starting data visualization, starting with the right question and the domain expertise will enable to get good output using data visualization.

By using data visualization techniques on security data, security analysts can gain be valuable insights on metrics and anomaly detection. Hopefully these insights can make security practitioners jobs easier.

3.0 Security Data Visualization Process

At a very high level the security visualization process consists of below five steps:



(Security Data Visualization process)

The key steps involved in visualization are

Step 1 – Visualization Goals

Step 2 - Data Preparation phase

Step 3 - Exploration phase

Step 4 - Visualization phase

Step 5 - Feedback and fine-tune

Below section review the activities involved in each step:

3.1 Step 1 – Visualization Goals

It is important to understand our goals and what the team is trying to achieve before jumping into any security visualization. Visualization should be goal driven and use case driven and not data driven. By thinking and documenting the goals security analysts start with the end objective in mind which is very important to capture the right data and use the right tools for visualization.

3.2 Step 2 - Data Preparation

It starts with searching data and preparing the data for analysis. The next step is to explore the data with the right questions, then visualize the data to develop insights and act on it.

The most important step before starting visualization is data cleansing or making the data available in a usable format. In real life, this is the biggest challenge since the data might be in an incompatible format; some parts may be missing or other similar challenges. This means a good amount of time has to be spent on data cleaning.

3.3 Step 3 - Explore

Asking the right question will lead to further exploration and visualization using statistical/probabilistic models/algorithms and lead to useful insights/decisions.

The explore phase will look at some analytical activities that will enable security teams to ask the right questions and look at the data to see how security teams can achieve their goals.

3.3.1 Analytical activities of Security Data Visualization analysts

Finding the right way to view data is as much an art as a science. Below is a framework of analytical activities adapted to information security domain based on paper (Amar, "Low-Level Components of Analytic Activity in Information Visualization", 2005).

1. “Retrieve Value - Given a set of specific cases, find attributes of those cases –
    - What is number of security incidents per day due to malware?
    - How long does it take to resolve a security incident?
2. Filter - Given some concrete conditions on attribute values, find data cases satisfying those conditions.
    - Which types of security incidents did not meet the Service Level Agreement defined?
3. Compute Derived Value - Given a set of data cases, compute an aggregate numeric representation of those data cases.
    - What is the average time taken to resolve security incidents?
4. Find Extremum - Find data cases possessing an extreme value of an attribute over its range within the data set.
    - What is the office location which most security incidents?
5. Sort - Given a set of data cases, rank them according to some ordinal metric.
    - Order the security incidents by severity and impact.
6. Determine Range - Given a set of data cases and an attribute of interest, find the span of values within the set.
    - What is the time taken during various phases in the Cyber Kill chain during incident response?

7. Characterize Distribution - Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attribute's values over the set.
   - What is the distribution of phishing/malware/insider threat incidents?
8. Find Anomalies - Identify any anomalies within a given set of data cases with respect to a given relationship or expectation, e.g. statistical outliers.
   - Are there any outliers in type of incidents?
9. Cluster - Given a set of data cases, find clusters of similar attribute values.
   - Are there groups of incidents w/ similar TTPs?
   - Is there a cluster of incidents which take long times to resolve?
10. Correlate - Given a set of data cases and two attributes, determine useful relationships between the values of those attributes.
    - Is there a trend of increasing time to resolve security incidents?"

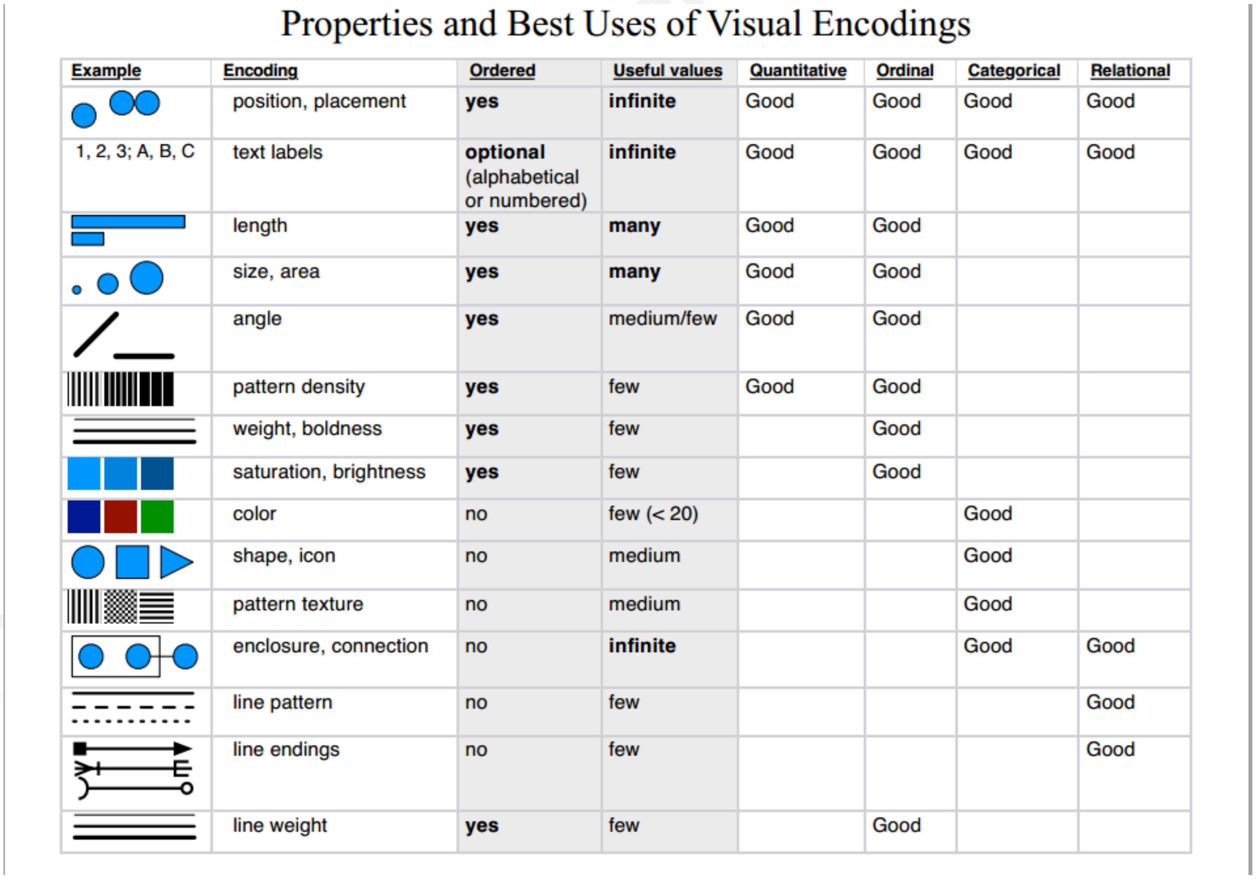
(Amar, "Low-Level Components of Analytic Activity in Information Visualization", 2005)

It is important to understand and use these tasks in different visualization techniques and allows security analyst to think through all the possibilities for coming up with right question and the results the organization is looking for. These analytical tasks/activities form foundation in understanding the statistical possibilities which can be used to explore the data.

3.4 Step 4 - Visualize

There are two aspects to visualization theory, one is the aesthetics. There is literature around how to use color, hue, thickness and other aspects to make visually pleasing images to intended audience. There is lot of design guidelines in the book “Tufte, E. (1983). The visual display of quantitative information. Cheshire, Conn. (Box 430, Cheshire 06410): Graphics Press”. There is a dedicated chapter in the book “Marty, R. (2009). Applied security visualization. Upper Saddle River, NJ: Addison-Wesley”.

Below table of properties and best uses of visual encoding by Noah Iliinsky (Iliinsky, "properties and best uses of visual encoding ")



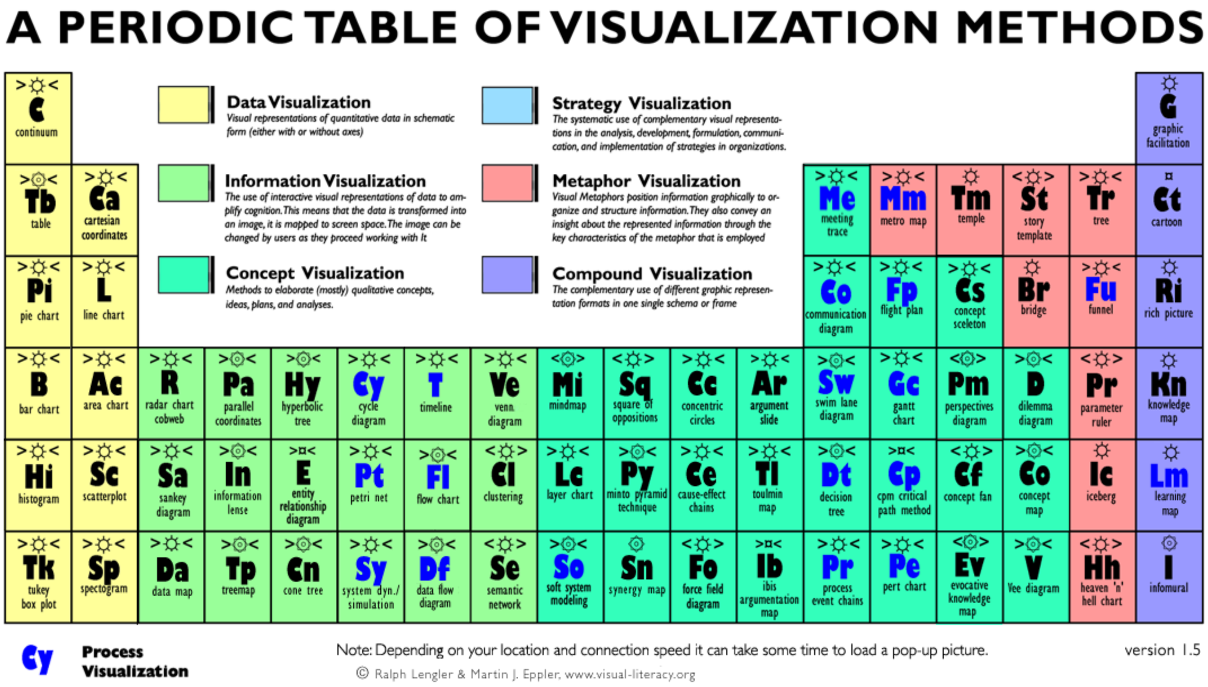
Properties and Best Uses of Visual Encodings

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
● ●●	position, placement	**yes**	**infinite**	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	**optional** (alphabetical or numbered)	**infinite**	Good	Good	Good	Good
▬	length	**yes**	**many**	Good	Good		
• ● ⬤	size, area	**yes**	**many**	Good	Good		
/ —	angle	**yes**	medium/few	Good	Good		
▥	pattern density	**yes**	few	Good	Good		
≡	weight, boldness	**yes**	few		Good		
■ ■ ■	saturation, brightness	**yes**	few		Good		
🟦 🟥 🟩	color	no	few (< 20)			Good	
● ■ ▶	shape, icon	no	medium			Good	
▥ ▩ ≡	pattern texture	no	medium			Good	
● ●–●	enclosure, connection	no	**infinite**			Good	Good
- - - - -	line pattern	no	few				Good
→	line endings	no	few				Good
≡	line weight	**yes**	few		Good		

(Properties and best uses of visual encoding (Iliinsky, N., 2014))

The other aspect is to understand the different visualization methods which are available. The section of this paper below covers some of the visualization methods.

KPI Library has developed the "A Periodic Table of Visualization Methods"(“A Periodic Table of Visualization Methods”, http://www.visual-literacy.org/periodic_table/periodic_table.html).



A PERIODIC TABLE OF VISUALIZATION METHODS

Data Visualization *Visual representations of quantitative data in schematic form (either with or without axes)*

Information Visualization *The use of interactive visual representations of data to amplify cognition. This means that the data is transformed into an image, it is mapped to screen space. The image can be changed by users as they proceed working with It*

Concept Visualization *Methods to elaborate (mostly) qualitative concepts, ideas, plans, and analyses.*

Strategy Visualization *The systematic use of complementary visual representations in the analysis, development, formulation, communication, and implementation of strategies in organizations.*

Metaphor Visualization *Visual Metaphors position information graphically to organize and structure information. They also convey an insight about the represented information through the key characteristics of the metaphor that is employed*

Compound Visualization *The complementary use of different graphic representation formats in one single schema or frame*

C continuum																	G graphic facilitation
Tb table	Ca cartesian coordinates											Me meeting trace	Mm metro map	Tm temple	St story template	Tr tree	Ct cartoon
Pi pie chart	L line chart											Co communication diagram	Fp flight plan	Cs concept sceleton	Br bridge	Fu funnel	Ri rich picture
B bar chart	Ac area chart	R radar chart cobweb	Pa parallel coordinates	Hy hyperbolic tree	Cy cycle diagram	T timeline	Ve venn diagram	Mi mindmap	Sq square of oppositions	Cc concentric circles	Ar argument slide	Sw swim lane diagram	Gc gantt chart	Pm perspectives diagram	D dilemma diagram	Pr parameter ruler	Kn knowledge map
Hi histogram	Sc scatterplot	Sa sankey diagram	In information lense	E entity relationship diagram	Pt petri net	Fl flow chart	Cl clustering	Lc layer chart	Py minto pyramid technique	Ce cause-effect chains	Tl toulmin map	Dt decision tree	Cp cpm critical path method	Cf concept fan	Co concept map	Ic iceberg	Lm learning map
Tk tukey box plot	Sp spectogram	Da data map	Tp treemap	Cn cone tree	Sy system dyn./ simulation	Df data flow diagram	Se semantic network	So soft system modeling	Sn synergy map	Fo force field diagram	Ib ibis argumentation map	Pr process event chains	Pe pert chart	Ev evocative knowledge map	V Vee diagram	Hh heaven 'n' hell chart	I infomural

Cy **Process Visualization**

Note: Depending on your location and connection speed it can take some time to load a pop-up picture.

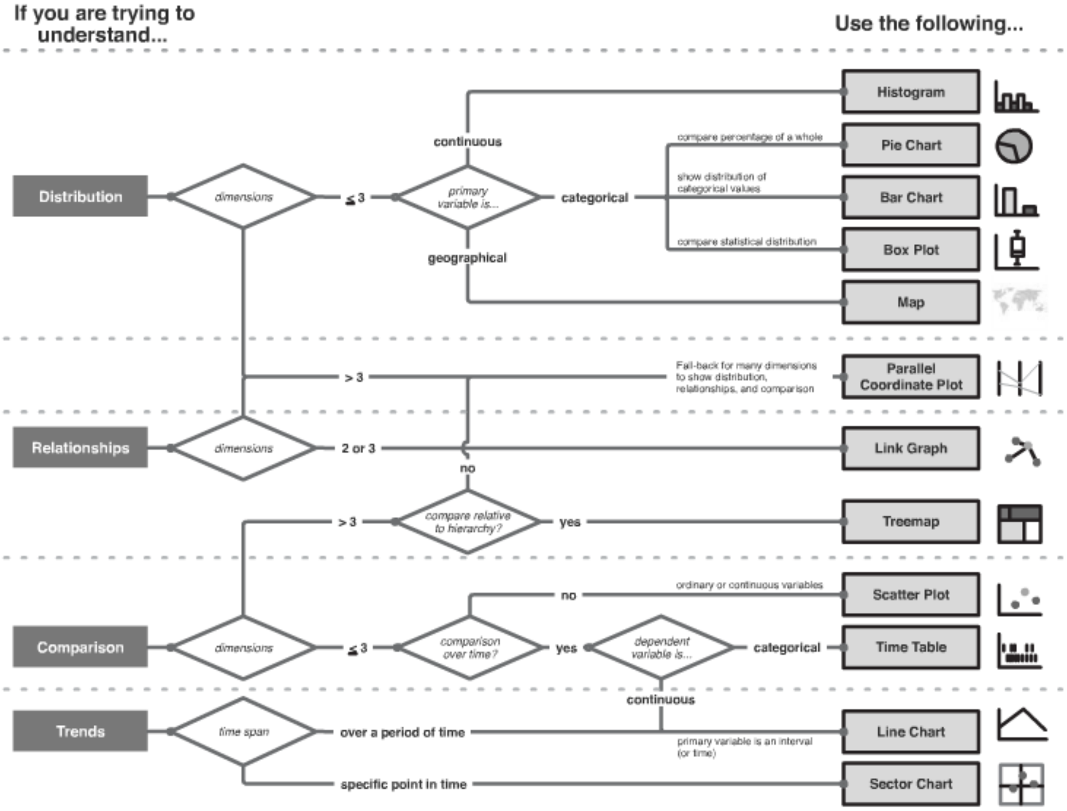
version 1.5

© Ralph Lengler & Martin J. Eppler, www.visual-literacy.org

(A Periodic Table of Visualization Methods(KPI library, 2014))

The flowchart below is from Marty, R. (2009). Applied security visualization helps explain which graphs can be used for which purpose.

pingbalaji@gmail.com



(Choosing Graph Types, "Marty, R. (2009). Applied security visualization. Upper Saddle River, NJ: Addison-Wesley")

Next section will review some examples of security data visualization methods.

Most of these visualizations can be developed using R. R is a free, open source language with access to powerful, cutting-edge analytics with more than 2000 packages. The installation of R and basic examples of R are explained in detail with screenshots in Appendix C. Appendix C further offers R examples on various basic graphs, plots, treemap, and models.

If you use Splunk the below webpage provides instructions on how to create Dashboards and Visualizations with time chart, bar chart, spark line, line chart, scatter chart and bubble chart

Dashboards and Visualizations. (n.d.). Retrieved December 15, 2014, from http://docs.splunk.com/Documentation/Splunk/6.2.0/Viz/Visualizationreference

If you need additional examples visit the below reference pages which catalogues different visualization techniques:

Mbostock/d3. (n.d.). Retrieved December 15, 2014, from

https://github.com/mbostock/d3/wiki/Gallery

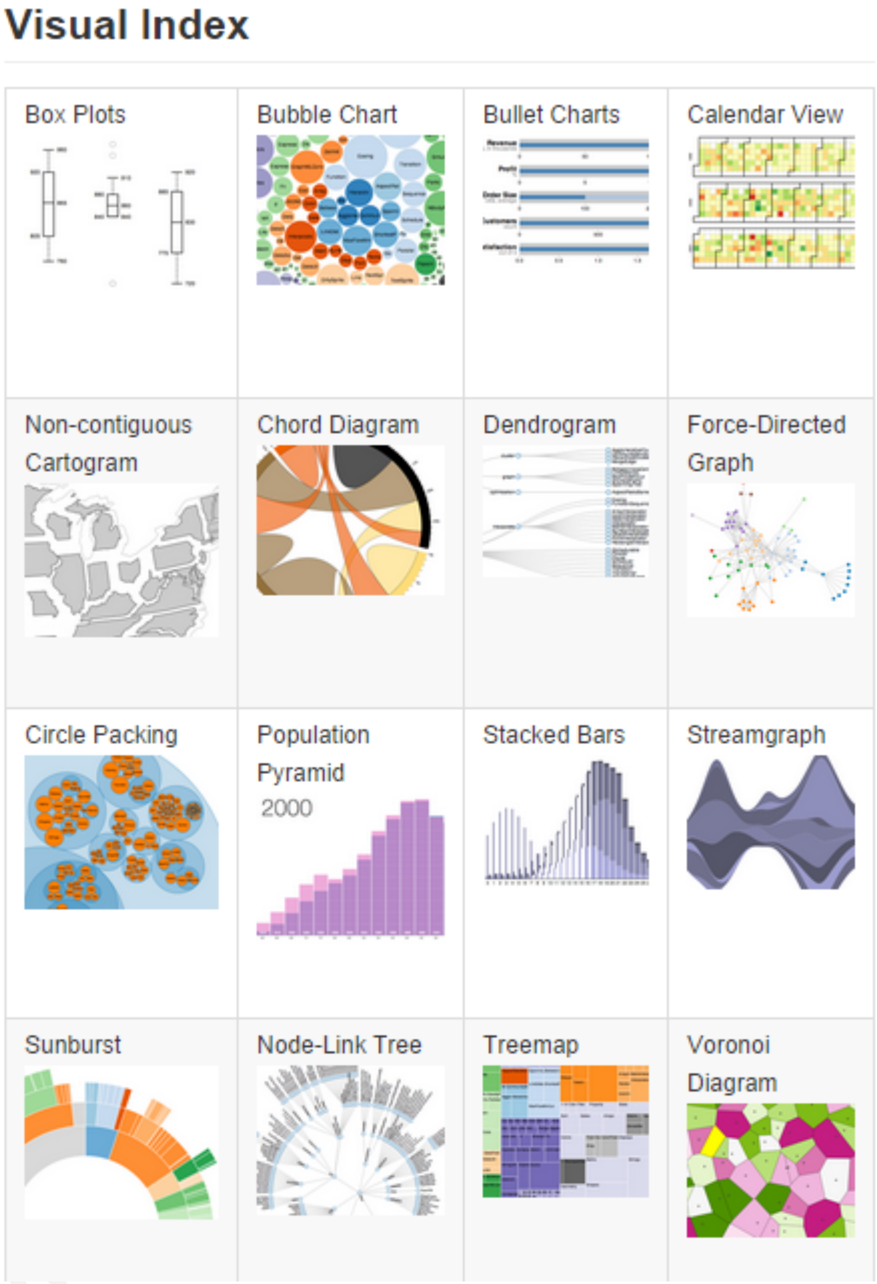
Visualization Methods. (n.d.). Retrieved December 15, 2014, from

http://kitwallace.co.uk/viz/method

Highcharts - Basic line. (n.d.). Retrieved December 15, 2014, from

http://www.highcharts.com/demo/

Below is the snapshot of the examples in the D3 data visualization Gallery:



Visual Index

Box Plots	Bubble Chart	Bullet Charts	Calendar View
Non-contiguous Cartogram	Chord Diagram	Dendrogram	Force-Directed Graph
Circle Packing	Population Pyramid	Stacked Bars	Streamgraph
Sunburst	Node-Link Tree	Treemap	Voronoi Diagram

(Mbostock/d3. (n.d.). Retrieved December 15, 2014, from https://github.com/mbostock/d3/wiki/Gallery)

3.5 Step 5 - Feedback and fine-tune

This step involves continuous improvement with feedback from the stakeholders and availability of new data.

4.0 Security Data Visualization Project Plans

4.1 Security Data Visualization Project 1 – Communicating value

There is a well-known proverb "a picture is worth a thousand words" (Piqua Leader-Dispatch, One Look Is Worth A Thousand Words, 1913, p. 2) which explains this. The problem with traditional metrics is numbers and tables can be daunting and details can be missed easily. Visualizing effectively will enable the security team to highlight the salient points in the data. Senior business leaders have very less time so by using security data visualization techniques give the capability to creatively represent the metrics data. Security data visualization enables you to tell a story with the data Information security is becoming a common topic in board room discussions and it is becoming more and more important that the value of information security is communicated to business leaders. This will enable the Information Security team in any organization to ensure it is managing the risks effectively and is aligned towards business needs and objectives.

Below are the high level phases in developing security data visualization project.

- Knowledge gathering phase
  - Statistical knowledge – 1 week - 2 weeks
  - Visualization theory – 1 week - 2 weeks
  - Information Security Domain Expertise – Ongoing
- Visualization Goals – 1 week
- Data Preparation phase – 1 week - 2 weeks
- Exploration phase – 1 week - 2 months
- Visualization phase – 1 week - 2 months
  - Experimentation

- Feedback and fine-tune – Ongoing

Depending upon the expertise level and time provided to the team within 1 month to 6 months effective metrics to communicate information security value to senior management can be developed in this project.

Let us see the activities involved in these steps in detail.

Knowledge gathering

Statistical knowledge – 1 week - 2 weeks

Statistics knowledge is very important from metrics point of view. Security metrics has a lot of numeric data, understanding the different terms like mean, median, mode, standard deviation, linear regression will enable the project team to create more insightful metrics. Since most of the security team major would be engineering based it will be easy for the team to revisit these topics within one or two weeks.

Visualization theory – 1 week - 2 weeks

There are two aspects to visualization theory, one is the aesthetics. There is literature around how to use color, hue, thickness and other aspects to make visually pleasing images to intended audience. Since this project is about communicating value to senior management and they usually only provide only small amount of real estate to capture their attention like a single slide it is very important the visualization is pleasing and tells story clearly. There is lot of design guidelines in the book "Tufte, E. (1983). The visual display of quantitative information. Cheshire, Conn. (Box 430, Cheshire 06410): Graphics Press". There is a dedicated chapter in the book "Marty, R. (2009). Applied security visualization. Upper Saddle River, NJ: Addison-Wesley".

The other aspect is to understand the different visualization methods which are available. The 3.3.4 Visualize section of this paper covers some of the visualization methods.

Information Security Domain Expertise – Ongoing

Information security domain expertise is the knowledge gathered as part of the day-to-day job activities. As information security practitioners it will be easy for gathering additional domain expertise if required in specific information security areas. This is ongoing activity where the team will keep updating their security metrics knowledge. In this case information security metrics is the domain expertise, Appendix B gives the background materials on information security metrics which might provide guidance on understanding the requirements and the goals.

Visualization Goals – 1 week

Security data visualization is only valuable if the security team has the right questions and the right data. It is important to understand our goals and what the organization is trying to achieve before jumping into any security visualization. Visualization should be goal driven and use case driven and not data driven. By thinking and documenting the goals the security team starts with the end objective in mind which is very important to capture the right data and use the right tools for visualization. For example the scope might be to improve the quarterly information security dashboard shared with senior management say CEO, CFO and their management team.

If the goal is to improve security metrics before jumping to the step of setting up security visualization toolbox or gathering data, it is important that the correct requirements are gathered for the information security metrics.

Data Preparation phase – 1 week - 2 weeks

Data preparation phase takes stock all the available metrics data and takes in to consideration any normalization that needs to happen before the data can be used. For security metrics it is mostly reports provided by various security tools which need to be normalized for reporting.

Exploration – 1 week - 2 months

Exploration phase is where the right questions are asked to understand and clarify the scope of visualization effort. This is where all the current metrics is looked to understand the gaps. The team may require some brainstorming session to come up with different options using statistical methods. Next step would be to develop use case with all possible options of displaying the data. The use cases might also take in to consideration the statistical methods for future trend predictions. Section 4.3.1 explains some analytical activities which might be useful. Below section highlights some techniques from Stephen Few for reference.

Data visualization techniques for Quantitative messages - Stephen Few

"Time-series: A single variable is captured over a period of time, such as the security incidents rate over a 1-year period. A line chart may be used to demonstrate the trend.

Ranking: Categorical subdivisions are ranked in ascending or descending order, such as a ranking of incident duration (the measure) by incident category during a single period. A bar chart may be used to show the comparison across the incident category.

Part-to-whole: Categorical subdivisions are measured as a ratio to the whole (i.e., a percentage out of 100%). A pie chart or bar chart can show the comparison of ratios, such as the phishing incidents share represented by total incidents.

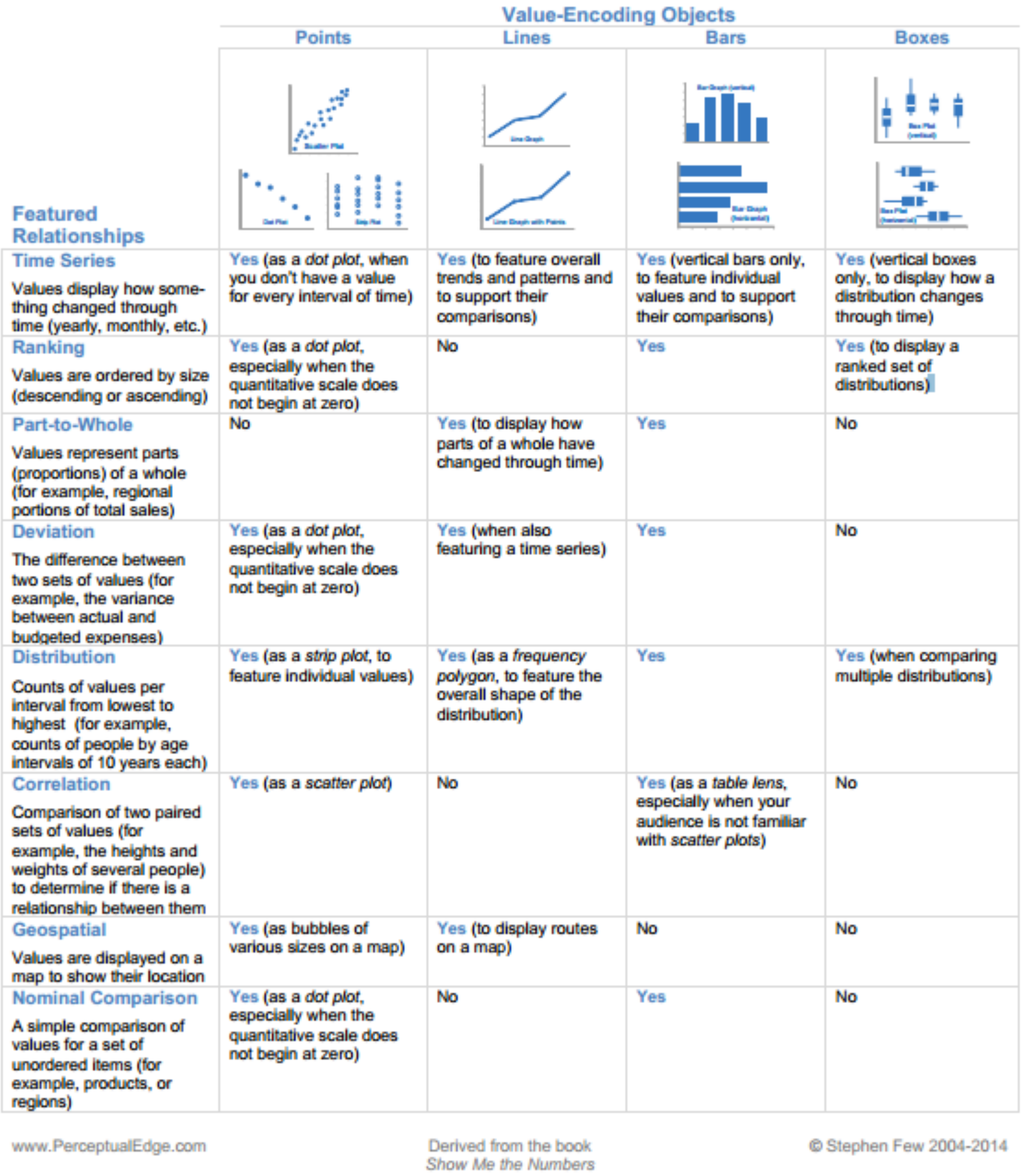
Deviation: Categorical subdivisions are compared again a reference, such as a comparison of actual time taken to resolve incidents vs. agreed time for resolving incidents for a given time period. A bar chart can show comparison of the actual versus the reference/agreed amount.

Frequency distribution: Shows the number of observations of a particular variable for given interval, such as the number of incidents in which the malware is involved is between intervals such as 0-10%, 11-20%, etc. A histogram, a type of bar chart, may be used for this analysis.

Correlation: Comparison between observations represented by two variables (X,Y) to determine if they tend to move in the same or opposite directions. For example, plotting incidents (X) and holiday (Y) for a sample of months. A scatter plot is typically used for this message.

Nominal comparison: Comparing categorical subdivisions in no particular order, such as the incidents by category. A bar chart may be used for this comparison.

Geographic or geospatial: Comparison of a variable across a map or layout, such as the incident rate by offices or the number of persons on the various floors of a building. A cartogram is a typical graphic used.” (Few, "Selecting the Right Graph for Your Message ")


Scatter Plot
Dot Plot
Strip Plot
Line Graph
Line Graph with Points
Bar Graph (vertical)
Bar Graph (horizontal)
Box Plot (vertical)
Box Plot (horizontal)

	Value-Encoding Objects			
Featured Relationships	Points	Lines	Bars	Boxes
Time Series Values display how something changed through time (yearly, monthly, etc.)	Yes (as a *dot plot*, when you don't have a value for every interval of time)	Yes (to feature overall trends and patterns and to support their comparisons)	Yes (vertical bars only, to feature individual values and to support their comparisons)	Yes (vertical boxes only, to display how a distribution changes through time)
Ranking Values are ordered by size (descending or ascending)	Yes (as a *dot plot*, especially when the quantitative scale does not begin at zero)	No	Yes	Yes (to display a ranked set of distributions)
Part-to-Whole Values represent parts (proportions) of a whole (for example, regional portions of total sales)	No	Yes (to display how parts of a whole have changed through time)	Yes	No
Deviation The difference between two sets of values (for example, the variance between actual and budgeted expenses)	Yes (as a *dot plot*, especially when the quantitative scale does not begin at zero)	Yes (when also featuring a time series)	Yes	No
Distribution Counts of values per interval from lowest to highest (for example, counts of people by age intervals of 10 years each)	Yes (as a *strip plot*, to feature individual values)	Yes (as a *frequency polygon*, to feature the overall shape of the distribution)	Yes	Yes (when comparing multiple distributions)
Correlation Comparison of two paired sets of values (for example, the heights and weights of several people) to determine if there is a relationship between them	Yes (as a *scatter plot*)	No	Yes (as a *table lens*, especially when your audience is not familiar with *scatter plots*)	No
Geospatial Values are displayed on a map to show their location	Yes (as bubbles of various sizes on a map)	Yes (to display routes on a map)	No	No
Nominal Comparison A simple comparison of values for a set of unordered items (for example, products, or regions)	Yes (as a *dot plot*, especially when the quantitative scale does not begin at zero)	No	Yes	No

www.PerceptualEdge.com

Derived from the book *Show Me the Numbers*

© Stephen Few 2004-2014

(Few, "Graph Selection Matrix")

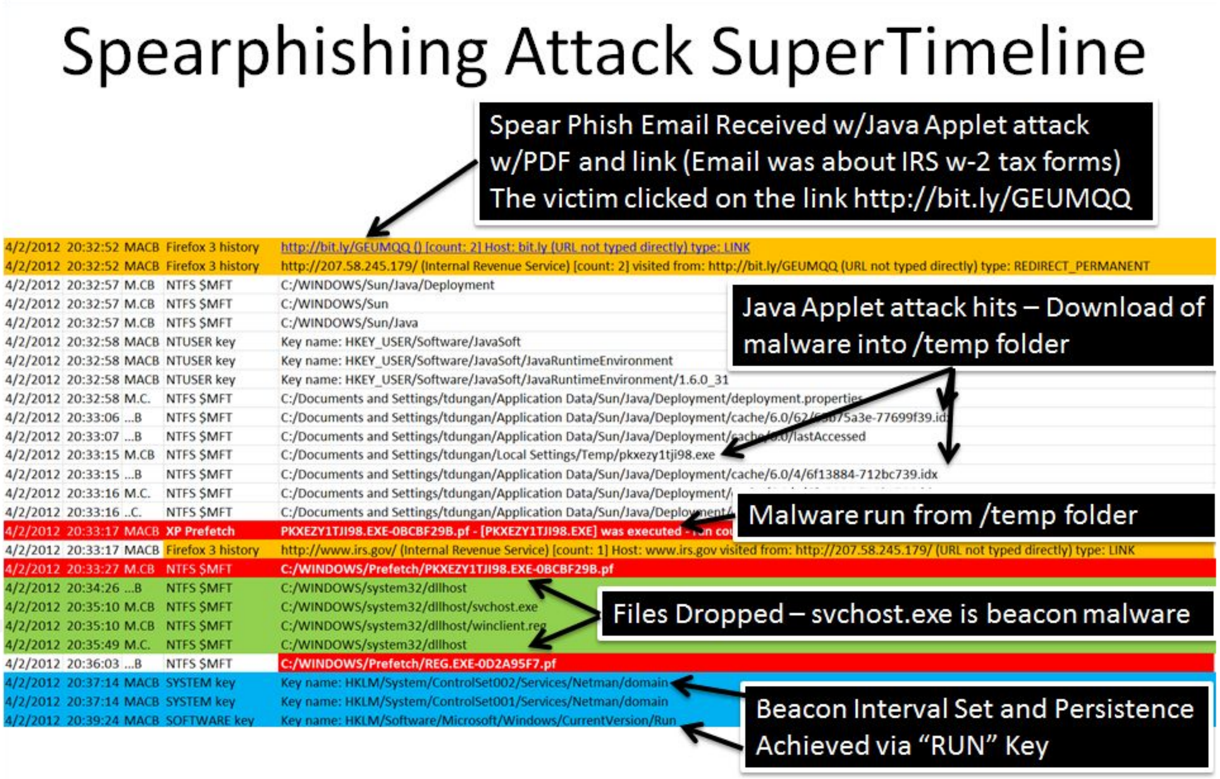
Experimentation of Visualization – 1 week - 2 months

During the visualization phase the team develops the visualizations and refines the visualization. Once the graph is determined it can be easily developed using tools like R which is covered in detail in Appendix C. If the visualization requires additional tools section 6 explains setting up visualization toolbox.

In some cases the visualization can be simple flow chart representing incidents in cyber kill chain to identify which parts of the kill chain was successful so the organization can strengthen the controls. Another example of visualization is shown below with the Pareto plot,

Management Laboratory. (n.d.). Retrieved December 15, 2014, from http://www.sans.edu/research/management-laboratory/article/mgt421-scott-pareto

Another example of visualization can be just showing the different stages in a malware infection with colors indicating which stages were successful as timeline analysis flowchart. Below example shows how visual timeline analysis helps explain the chronology of a spear phishing attack.



Spearphishing Attack SuperTimeline

Spear Phish Email Received w/Java Applet attack w/PDF and link (Email was about IRS w-2 tax forms) The victim clicked on the link http://bit.ly/GEUMQQ

```
4/2/2012 20:32:52 MACB Firefox 3 history  http://bit.ly/GEUMQQ () [count: 2] Host: bit.ly (URL not typed directly) type: LINK
4/2/2012 20:32:52 MACB Firefox 3 history  http://207.58.245.179/ (Internal Revenue Service) [count: 2] visited from: http://bit.ly/GEUMQQ (URL not typed directly) type: REDIRECT_PERMANENT
4/2/2012 20:32:57 M.CB NTFS $MFT          C:/WINDOWS/Sun/Java/Deployment
4/2/2012 20:32:57 M.CB NTFS $MFT          C:/WINDOWS/Sun
4/2/2012 20:32:57 M.CB NTFS $MFT          C:/WINDOWS/Sun/Java
4/2/2012 20:32:58 MACB NTUSER key         Key name: HKEY_USER/Software/JavaSoft
4/2/2012 20:32:58 MACB NTUSER key         Key name: HKEY_USER/Software/JavaSoft/JavaRuntimeEnvironment
4/2/2012 20:32:58 MACB NTUSER key         Key name: HKEY_USER/Software/JavaSoft/JavaRuntimeEnvironment/1.6.0_31
4/2/2012 20:32:58 M.C. NTFS $MFT          C:/Documents and Settings/tdungan/Application Data/Sun/Java/Deployment/deployment.properties
4/2/2012 20:33:06 ...B NTFS $MFT          C:/Documents and Settings/tdungan/Application Data/Sun/Java/Deployment/cache/6.0/62/6b75a3e-77699f39.id
4/2/2012 20:33:07 ...B NTFS $MFT          C:/Documents and Settings/tdungan/Application Data/Sun/Java/Deployment/cache/6.0/lastAccessed
4/2/2012 20:33:15 M.CB NTFS $MFT          C:/Documents and Settings/tdungan/Local Settings/Temp/pkxezy1tji98.exe
4/2/2012 20:33:15 ...B NTFS $MFT          C:/Documents and Settings/tdungan/Application Data/Sun/Java/Deployment/cache/6.0/4/6f13884-712bc739.idx
4/2/2012 20:33:16 M.C. NTFS $MFT          C:/Documents and Settings/tdungan/Application Data/Sun/Java/Deployment/
4/2/2012 20:33:16 ..C. NTFS $MFT          C:/Documents and Settings/tdungan/Application Data/Sun/Java/Deployment/
4/2/2012 20:33:17 MACB XP Prefetch        PKXEZY1TJI98.EXE-0BCBF29B.pf - [PKXEZY1TJI98.EXE] was executed - run cou
4/2/2012 20:33:17 MACB Firefox 3 history  http://www.irs.gov/ (Internal Revenue Service) [count: 1] Host: www.irs.gov visited from: http://207.58.245.179/ (URL not typed directly) type: LINK
4/2/2012 20:33:27 M.CB NTFS $MFT          C:/WINDOWS/Prefetch/PKXEZY1TJI98.EXE-0BCBF29B.pf
4/2/2012 20:34:26 ...B NTFS $MFT          C:/WINDOWS/system32/dllhost
4/2/2012 20:35:10 M.CB NTFS $MFT          C:/WINDOWS/system32/dllhost/svchost.exe
4/2/2012 20:35:10 M.CB NTFS $MFT          C:/WINDOWS/system32/dllhost/winclient.reg
4/2/2012 20:35:49 M.C. NTFS $MFT          C:/WINDOWS/system32/dllhost
4/2/2012 20:36:03 ...B NTFS $MFT          C:/WINDOWS/Prefetch/REG.EXE-0D2A95F7.pf
4/2/2012 20:37:14 MACB SYSTEM key         Key name: HKLM/System/ControlSet002/Services/Netman/domain
4/2/2012 20:37:14 MACB SYSTEM key         Key name: HKLM/System/ControlSet001/Services/Netman/domain
4/2/2012 20:39:24 MACB SOFTWARE key       Key name: HKLM/Software/Microsoft/Windows/CurrentVersion/Run
```

Java Applet attack hits – Download of malware into /temp folder

Malware run from /temp folder

Files Dropped – svchost.exe is beacon malware

Beacon Interval Set and Persistence Achieved via "RUN" Key

(Spear Phishing. (n.d.). Retrieved December 15, 2014, from https://blogs.sans.org/computer-forensics/files/2013/02/Spearphishing.jpg)

Feedback and fine-tune – Ongoing

This step involves continuous improvement with feedback from the stakeholders and availability of new data. As a simple example having darker colors Blue or Purple color in the graph instead of other light colors like Yellow helped in security part of the presentation since it was fitting well with the rest of the presentation.

4.2 Security Data Visualization Project 2 – Finding Anomalies

The objective of this project is finding anomalies using security data visualization. One of the key strengths of security teams is access to enterprise log data, meta-data , network traffic data, netflow data. The challenge is finding and isolating the bad actors from legitimate traffic. Human mind by evolution is trained to identify patterns and anomalies using visualization. Security professionals can benefit by visualizing enterprise data to find anomalies and identify patterns which will be helpful in isolating events which might indicate compromise.

Below are the high level phases in developing security data visualization project.

- Knowledge gathering phase
    - Statistical knowledge – 1 week - 2 weeks
    - Visualization theory – 1 week - 2 weeks
    - Information Security Domain Expertise – Ongoing
- Visualization Goals – 1 week
- Data Preparation phase – 1 week - 2 weeks
- Exploration phase – 1 week - 2 months
- Visualization phase – 1 week - 2 months

  - Experimentation

- Feedback and fine-tune – Ongoing

Depending upon the expertise level and time provided to the team within 1 month to 6 months effective process can be developed to find anomalies.

Let us see the activities involved in these steps in detail.

Knowledge gathering phase

Statistical knowledge – 1 week - 2 weeks

Statistics knowledge is very important for finding anomalies. Security tools provide lot of numeric data, understanding the different terms like mean, median, mode, standard deviation, linear regression will enable the project team to create more insightful decisions. The other key aspect is running large set of data through Gaussian distribution or Monte Carlo simulation models for predictions. Since most of the security team major would be engineering based it will be easy for the team to revisit these topics within one or two weeks.

Visualization theory – 1 week - 2 weeks

There are two aspects to visualization theory, one is the aesthetics. There is literature around how to use color, hue, thickness and other aspects to make visually pleasing images to intended audience. This aspect is not as important in this project since the focus is on finding anomalies and not necessarily communicating to different sets of audience. As long as the team understands the visualization lot of effort is not necessary for the aesthetics.

The other aspect is to understand the different visualization methods which are available. The 3.3.4 Visualize section of this paper covers some of the visualization methods.

Information Security Domain Expertise – Ongoing

In this case of finding anomalies understanding security log data is the foundational skills required for security data visualization. There are good books on network security monitoring which might augment the domain knowledge along with work experience. For anomaly detection use cases the security monitoring domain knowledge is more important. The security monitoring experience will enable to team to baseline the activities and understand the anomalies. The domain knowledge will enable the team to create as many hypotheses as possible. This is ongoing activity where the team will keep updating their knowledge in this area.

Visualization Goals – 1 week

Security data visualization is only valuable if the security team has the right questions and the right data. It is important to understand our goals and what the organization is trying to achieve before jumping into any security visualization. Visualization should be goal driven and use case driven and not data driven. By thinking and documenting the goals the security team starts with the end objective in mind which is very important to capture the right data and use the right tools for visualization. The team may require some brainstorming session to come up with goals and use cases for anomaly detection.

Data Preparation phase – 1 week - 2 weeks

Data preparation phase takes stock all the available network security log data and takes in to consideration any normalization that needs to happen before the data can be used.

Exploration – 1 week - 2 months

Exploration phase is where the right questions are asked to understand and clarify the scope of the visualization effort. This is where all the current process for anomaly detection is looked to understand the gaps. It is important to start small on a particular type of log, show success and the move to different types of logs and more complex visualizations.

For example the initial use case might be to identify anomalies in firewall log data using visualization. Next step would be to develop use case with all possible options of displaying the data. The use cases might also take in to consideration the statistical methods for future trend predictions. Section 4.3.1 explained some analytical activities which might be useful. For anomaly detection use cases the security monitoring domain knowledge is more important. The security monitoring experience will enable to team to baseline the activities and understand the anomalies. The domain knowledge will enable the team to create as many hypotheses as possible. Once you have a set of hypotheses within the scope the team can start exploring the possibilities of creating graphs and visualizing.

For this use case of Firewall log data visualization to identify anomalies the below GIAC paper has examples using Afterglow.

Visualizing Firewall Log Data to Detect Security Incidents. (n.d.). Retrieved November 30, 2014, from http://www.giac.org/paper/gcia/1651/visualizing-firewall-log-data-detect-security/109883

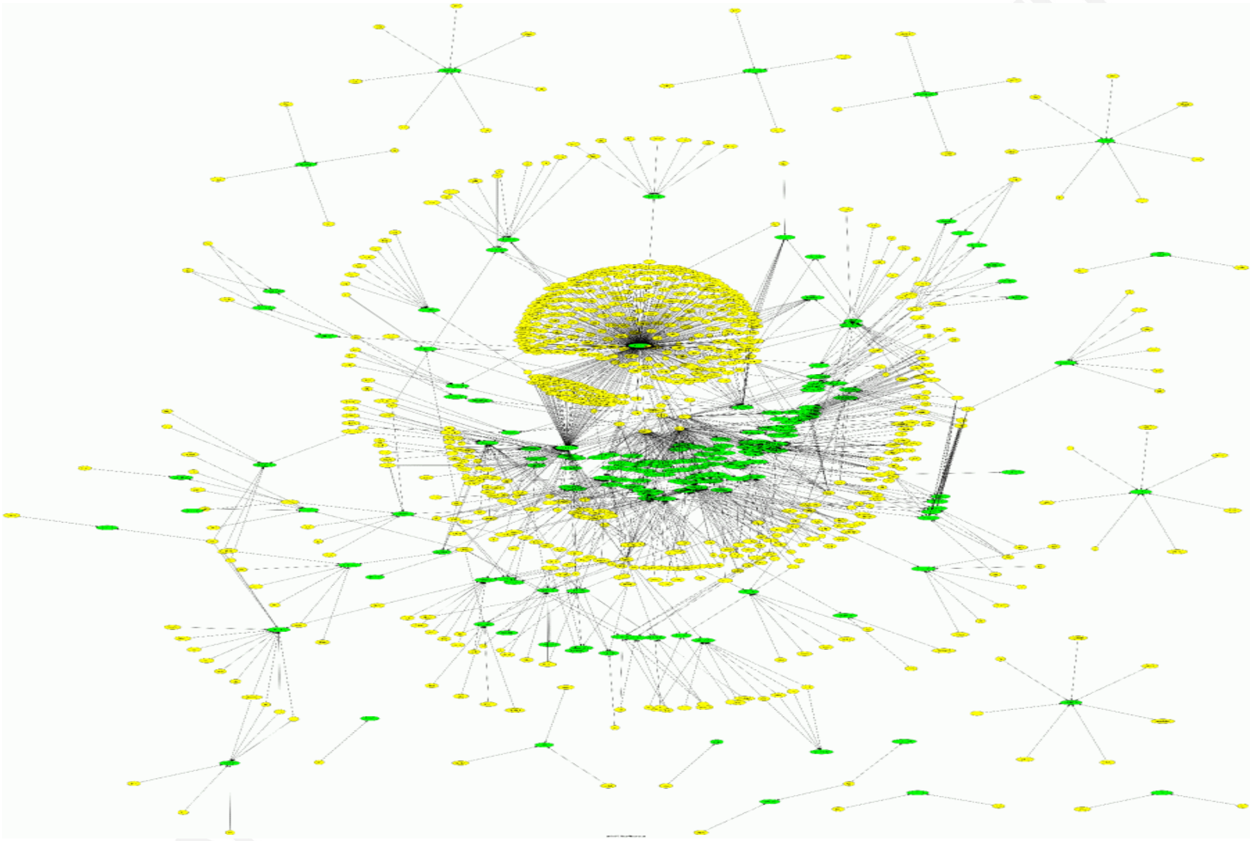
Experimentation of Visualization – 1 week - 2 months
Ben Shneiderman proposed the below methodology:

“Overview first, then zoom and filter, and finally, details on demand.”(Shneiderman, "The Eyes Have it", 1996)

This applies to identifying anomalies, once the major goals are decided, security team can start the visualization process with overview first, then zoom and filer, and finally, details on demand.

Below is a example from SecViz website on SSHD brute force attempts, once we get the overview, we can start zoom and filter on areas of interest and start investigating on the details. This iterative process will assist with finding anomalies.

SSHD brute force attempts - userids and IPs



(SSHD Brute force attempts. (n.d.). Retrieved December 15, 2014, from http://secviz.org/files/images/test-neato-e5-scaled.preview.gif)

Once the team is comfortable with firewall logs, analyzing login failures/brute force attempts they can move to other areas. Below flowchart from Marty, R. (2009) Applied security visualization helps explain which network security monitoring process can use visualization.

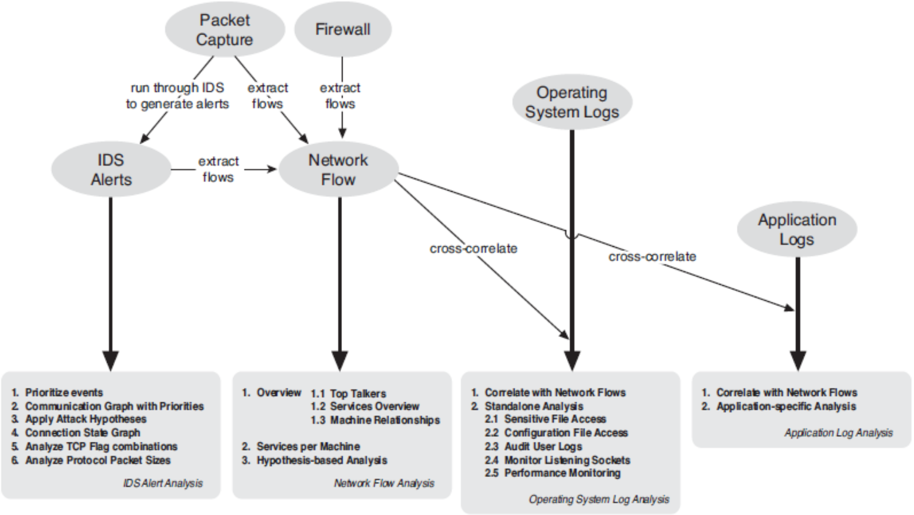

Packet Capture
Firewall
run through IDS to generate alerts
extract flows
extract flows
Operating System Logs
IDS Alerts
extract flows
Network Flow
Application Logs
cross-correlate
cross-correlate
1. Prioritize events
2. Communication Graph with Priorities
3. Apply Attack Hypotheses
4. Connection State Graph
5. Analyze TCP Flag combinations
6. Analyze Protocol Packet Sizes
IDS Alert Analysis
1. Overview 1.1 Top Talkers
1.2 Services Overview
1.3 Machine Relationships
2. Services per Machine
3. Hypothesis-based Analysis
Network Flow Analysis
1. Correlate with Network Flows
2. Standalone Analysis
2.1 Sensitive File Access
2.2 Configuration File Access
2.3 Audit User Logs
2.4 Monitor Listening Sockets
2.5 Performance Monitoring
Operating System Log Analysis
1. Correlate with Network Flows
2. Application-specific Analysis
Application Log Analysis

Figure 5-22 Forensic analysis process summary diagram. Ovals represent data sources, and the boxes contain the individual analysis processes.

(Forensic Analysis Process, "Marty, R. (2009). Applied security visualization. Upper Saddle River, NJ: Addison-Wesley")

Afterglow, Gobi and the other tools can be used to visualize, correlate and find anomalies.

If you use Splunk the paper (Discovering Security Events of Interest Using Splunk, sans.org) provides additional information on using Splunk to explore and visualize anomaly detection using Splunk.

The webpage (Visualizing Interesting Log Events Using Splunk's Google Maps Application, danielmiessler.com) explains how to use Google maps and Splunk to find interesting events

The webpage(Forensic timeline Splunking, kleinco.com.au, 2014) explains Forensic timeline Splunking for timeline analysis which is very useful for forensic analysis.

Feedback and fine-tune – Ongoing

Feedback is very vital in this process to share with the team. As the team starts using Afterglow and other tools like R, the feedback on what steps were successful in finding anomalies and what steps created false-positives will save a lot of time for the team. There is lot of scope to continuously improve based on feedback and progress. The team can slowly move in to correlating of all events and logs to find anomalies and keep iteratively improving the process.

5.0 Conclusion

This paper tries to emphasize the importance of security visualization. Security visualization can be used in many areas in information security. In this paper as one example, security metrics was used and shared examples of security visualization that can be used to communicate security metrics effectively. Security monitoring, anomaly detection, forensics, and malware analysis are examples of other use cases where security visualization can play a vital role and make us better security professionals.

The good news is, that there are a lot of resources available to learn "R" and other visualization tools. Data analytics/ data mining are well-established disciplines and there are many practical implications for data analytics. Coursera and EdX MOOC offer free courses on "R" and data analytics. If security practitioners are passionate and believe there can be new ways to analyze and visualize data, there is a lot of help available online.

Another benefit, especially for security professionals, is that, as part of information security jobs, security teams have access to security logs, incidents, and all the key data. Meta-data, combined with security visualization and data analysis skill sets, security teams will be able to identify anomalies, improve metrics and make security teams very efficient in our jobs.

Security data visualization also plays key role in emerging fields such as data science, machine learning, and exploratory data analytics.

I hope more security practitioners learn these data analysis and visualization techniques and by sharing these techniques. Hopefully together we can stay a step ahead in securing the organization from the information security threats/attacks.

References

Security Tools: Visualization Is Power. (n.d.). Retrieved September 16, 2014, from http://www.csoonline.com/article/2121183/data-protection/security-tools--visualization-is-power.html

Heat Map Love R Style. (n.d.). Retrieved September 16, 2014, from http://risktical.com/2012/01/20/heat-map-love-r-style/

#005: The Sight and Sound of Cybercrime - The Office For Creative Research. (2014, November 15). Retrieved December 15, 2014, from http://o-c-r.org/2014/11/15/specimen_box/

Data visualization with R: How to get and show meaningful metrics for a scrum team. (n.d.). Retrieved September 16, 2014, from http://www.ibm.com/developerworks/library/bd-operationalmetrics/index.html?ca=drs

Marty, R. (2009). Applied security visualization. Upper Saddle River, NJ: Addison-Wesley

Graphic Sociology. (n.d.). Retrieved September 16, 2014, from http://thesocietypages.org/graphicsociology/2012/11/20/visualizing-email-traffic/

Visualizing Firewall Log Data to Detect Security Incidents. (n.d.). Retrieved November 30, 2014, from http://www.giac.org/paper/gcia/1651/visualizing-firewall-log-data-detect-security/109883

Free Data Visualization and Analysis Tools. (n.d.). Retrieved September 16, 2014, from http://semanticommunity.info/Data_Science/Free_Data_Visualization_and_Analysis_Tools

How to Make a Heatmap – a Quick and Easy Solution. (n.d.). Retrieved September 16, 2014, from http://flowingdata.com/2010/01/21/how-to-make-a-heatmap-a-quick-and-easy-solution/

Data Exploration of a publicly available dataset. (n.d.). Retrieved September 16, 2014, from http://nbviewer.ipython.org/github/ClickSecurity/data_hacking/blob/master/mdl_exploration/MDL_Data_Exploration.ipynb

Data Exploration of a publicly available dataset (inÂ R). (n.d.). Retrieved September 16, 2014, from http://datadrivensecurity.info/blog/posts/2014/Jan/data-exploration-in-r/

One Look Is Worth A Thousand Words. (1913, August 1). Piqua Leader-Dispatch

Conway, D. (n.d.). The Data Science Venn Diagram. Retrieved November 29, 2014, from http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Iliinsky, N. (n.d.). Properties and best uses of visual encoding. Retrieved December 15, 2014, from http://complexdiagrams.com/wp-content/2012/01/VisualPropertiesTable.pdf

Data visualization. (2014, November 29). Retrieved November 29, 2014, from http://en.wikipedia.org/wiki/Data_visualization

Mondrian - Interactive Statistical Graphics in JAVA. (n.d.). Retrieved November 29, 2014, from http://www.theusrus.de/Mondrian/Mondrian.html

Mondrian. (n.d.). Retrieved November 29, 2014, from http://www.interactivegraphics.org/Slides_files/Mondrian.pdf

Data Visualization, Discovery and Visual Analytics - Use Cases, Tools, CoE, Vendors. (n.d.). Retrieved November 29, 2014, from https://practicalanalytics.wordpress.com/2014/09/24/data-visualization-discovery-and-visual-analytics-tools-coe/

IBM - Many Eyes. (n.d.). Retrieved November 29, 2014, from http://www-958.ibm.com/software/data/cognos/manyeyes/page/Visualization_Options.html

Improving visualisation -Links and reference. (n.d.). Retrieved November 29, 2014, from http://www.improving-visualisation.org/links

A Periodic Table of Visualization Methods. (n.d.). Retrieved November 29, 2014, from http://www.visual-literacy.org/periodic_table/periodic_table.html

Techniques. (n.d.). Retrieved November 29, 2014, from http://www.wikiviz.org/wiki/Techniques

Tnv : Screenshots. (n.d.). Retrieved November 29, 2014, from http://tnv.sourceforge.net/grabs.php

Visualization for Monitoring Network Security Events. (n.d.). Retrieved November 29, 2014, from http://ercim-news.ercim.eu/en90/special/visualization-for-monitoring-network-security-events
Fabian Fischer. (n.d.). Retrieved November 29, 2014, from http://ff.cx/eventvis/
VIS-SENSE -Links. (n.d.). Retrieved November 29, 2014, from http://www.vis-sense.eu/Links/
Visualanalyticseu About the book and download. (n.d.). Retrieved November 29, 2014, from http://www.visual-analytics.eu/book/aboutbook/

Honeynet Challenge. (n.d.). Retrieved November 29, 2014, from http://www.honeynet.org/sites/default/files/files/Fabian_Fischer_-_Forensic_Challenge_2011_-_Challenge_10.pdf

Presentation. (n.d.). Retrieved November 29, 2014, from http://www.cs.arizona.edu/~collberg/Teaching/466-566/2012/Resources/presentations/2012/topic13-final/report.pdf

Visualization. (n.d.). Retrieved November 29, 2014, from http://slagell.name/Adam_J._Slagell/Publications_files/lakkaraju05b.pdf

Black Hat - Open DNS presentation. (n.d.). Retrieved November 29, 2014, from https://www.blackhat.com/docs/us-14/materials/us-14-Hay-Unveiling-The-Open-Source-Visualization-Engine-For-Busy-Hackers-WP.pdf

FIRST Presentation. (n.d.). Retrieved November 29, 2014, from http://www.first.org/conference/2006/papers/banerjee-uday-papers.pdf

Visualization. (n.d.). Retrieved November 29, 2014, from http://www.diva-portal.org/smash/get/diva2:347722/FULLTEXT01.pdf

Microsoft Research. (n.d.). Retrieved November 29, 2014, from http://research.microsoft.com/en-us/um/people/shliu/Infovis-TVCJ.pdf

RUMINT. (n.d.). Retrieved November 29, 2014, from http://www.rumint.org/gregconti/publications/insecure_conti.pdf

Resources - Visualising Data. (n.d.). Retrieved November 29, 2014, from http://www.visualisingdata.com/index.php/resources/

The Data Visualization Beginner's Toolkit #1: Books and Other Resources. (n.d.). Retrieved November 29, 2014, from http://fellinlovewithdata.com/guides/data-vis-beginners-toolkit-1

Lectures for Information Visualization. (n.d.). Retrieved November 29, 2014, from http://comminfo.rutgers.edu/~aspoerri/Teaching/InfoVisOnline/Lectures/Lectures.htm

Presentation. (n.d.). Retrieved November 29, 2014, from http://vis.stanford.edu/files/2010-Narrative-InfoVis.pdf

Hype-free. (n.d.). Retrieved November 29, 2014, from http://www.agent31.eu/2008/05/visualization-techniques-for-networking.html

Complex Data Visualized. (n.d.). Retrieved November 29, 2014, from http://complexdatavisualized.com/

The Role of Visualization in Cyber Intelligence. (n.d.). Retrieved November 29, 2014, from https://www.recordedfuture.com/cyber-intelligence-visualization/

Tufte, Edward R (2001) [1983], The Visual Display of Quantitative Information (2nd ed.), Cheshire, CT: Graphics Press, ISBN 0-9613921-4-2.

Tufte, Edward R (2006), Beautiful Evidence, Cheshire, CT: Graphics Press, ISBN 0-9613921-7-7.

Tufte, Edward R (1997), Visual Explanations: Images and Quantities, Evidence and Narrative, Cheshire, CT: Graphics Press, ISBN 0-9613921-2-6.

Few, S. (n.d.). Perceptual Edge-Selecting the Right Graph for Your Message-2004. Retrieved December 5, 2014, from http://www.perceptualedge.com/articles/ie/the_right_graph.pdf

Amar, R. (n.d.). Low-Level Components of Analytic Activity in Information Visualization. Retrieved December 6, 2014, from http://www.cc.gatech.edu/~stasko/papers/infovis05.pdf

Few, S. (n.d.). Graph Selection Matrix. Retrieved December 6, 2014, from http://www.perceptualedge.com/articles/misc/Graph_Selection_Matrix.pdf

Few, S. (n.d.). Tapping the Power of Visual Perception. Retrieved December 6, 2014, from http://www.perceptualedge.com/articles/ie/visual_perception.pdf

Few, S. (n.d.). Selecting the Right Graph for Your Message. Retrieved December 6, 2014, from http://www.perceptualedge.com/articles/ie/the_right_graph.pdf

A Periodic Table of Visualization Methods. (n.d.). Retrieved December 8, 2014, from http://www.visual-literacy.org/periodic_table/periodic_table.html

Forensic timeline Splunking. (n.d.). Retrieved December 15, 2014, from http://kleinco.com.au/thoughts-events/item/forensic-timeline-splunking

Visualizing Interesting Log Events Using Splunk's Google Maps Application. (2011, November 14). Retrieved December 15, 2014, from http://danielmiessler.com/blog/visualizing-interesting-log-events-using-splunks-google-maps-application/

Discovering Security Events of Interest Using Splunk. (n.d.). Retrieved December 15, 2014, from http://www.sans.org/reading-room/whitepapers/logging/discovering-security-events-interest-splunk-34272

Domas, C. (n.d.). The 1s and 0s behind cyber warfare. Retrieved December 15, 2014, from http://www.ted.com/talks/chris_domas_the_1s_and_0s_behind_cyber_warfare

Data to Insight: An Introduction to Data Analysis - The University of Auckland - FutureLearn. (n.d.). Retrieved December 16, 2014, from https://www.futurelearn.com/courses/data-to-insight

Appendix A

Security Data Visualization – Background and Inspiration

Inspiration for this paper was the training session from Raffael Marty on Security Visualization. The training was focused on how to use security visualization to help security analysts visualize security logs.

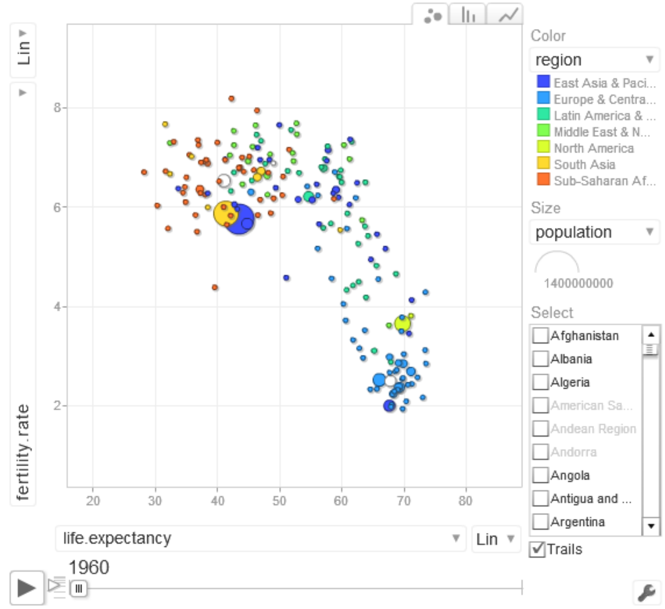
The other inspirations are from many TED talks were many of the TED presenters use visualization to tell powerful stories.

Another good example is gapminder.org. Imagine if you can implement the same visualization to show how security incidents have risen over time, similar to the video showed in gapminder.org.

The following example of the data visualization using Google Chart API is the motion chart, popularized by Hans Rosling in his 2006 TED talk.

Motion Chart data visualization link: https://gist.github.com/mages/5180860#file-worldbank_demo_with_googlevis-r :

It was a learning moment when the R code was executed and the browser opened with the motion chart.

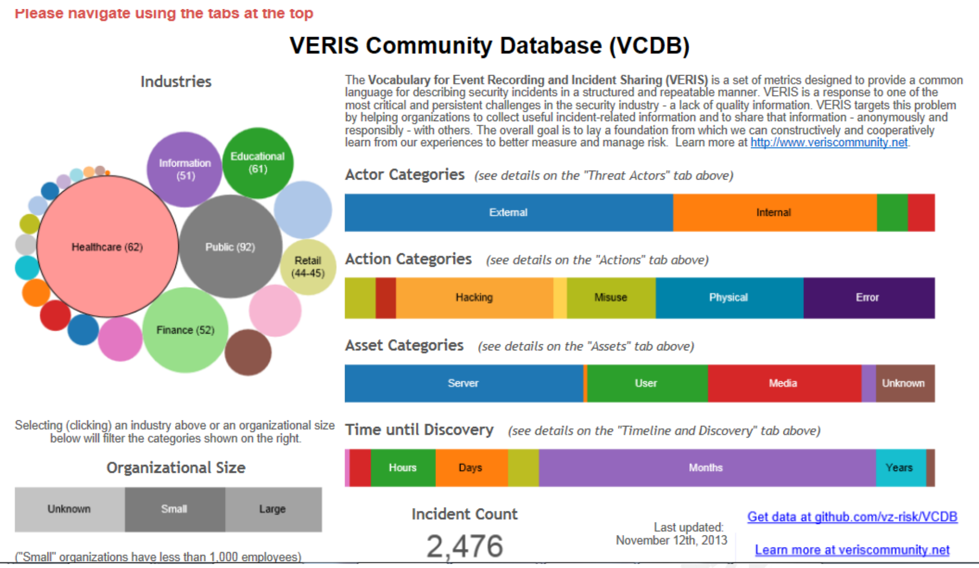


Data: WorldBank • Chart ID: MotionChartID37e46e005d25 • googleVis-0.5.5
R version 3.1.1 (2014-07-10) • Google Terms of Use • Documentation and Data Policy

(Gap Minder)

The final inspiration is from the data-Driven Security book which has well-structured information on security data analysis and visualization. There is a lot of guidance in the resources, blog, and podcast section of the Data-Driven Security book website.

The best known example for security visualization is the VERIS Community database. It can be accessed at http://public.tableausoftware.com/profile/jay.jacobs#!/vizhome/vcdb/Overview



Please navigate using the tabs at the top

VERIS Community Database (VCDB)

The **Vocabulary for Event Recording and Incident Sharing (VERIS)** is a set of metrics designed to provide a common language for describing security incidents in a structured and repeatable manner. VERIS is a response to one of the most critical and persistent challenges in the security industry - a lack of quality information. VERIS targets this problem by helping organizations to collect useful incident-related information and to share that information - anonymously and responsibly - with others. The overall goal is to lay a foundation from which we can constructively and cooperatively learn from our experiences to better measure and manage risk. Learn more at http://www.veriscommunity.net.

Selecting (clicking) an industry above or an organizational size below will filter the categories shown on the right.

("Small" organizations have less than 1,000 employees)

Last updated: November 12th, 2013

Get data at github.com/vz-risk/VCDB

Learn more at veriscommunity.net

Industries
Information (51)
Educational (61)
Healthcare (62)
Public (92)
Retail (44-45)
Finance (52)
Actor Categories (see details on the "Threat Actors" tab above)
External
Internal
Action Categories (see details on the "Actions" tab above)
Hacking
Misuse
Physical
Error
Asset Categories (see details on the "Assets" tab above)
Server
User
Media
Unknown
Time until Discovery (see details on the "Timeline and Discovery" tab above)
Hours
Days
Months
Years
Organizational Size
Unknown
Small
Large
Incident Count
2,476

(Jacobs, "VERIS Community Database", 2014)

Another security visualization example in the public domain on the world's biggest data breaches is the web page shown below.

Data analysis features on this web site serve as a good example on how security metrics can be extended to a dynamic format creatively.

Imagine a presentation for senior management with similar dynamic security metrics for your organization; it will be very handy to explain any questions immediately since the actual incident events or risk events can be drilled down right away. Usually, we provide static presentations to senior management. By providing this dynamic content, the senior management can be empowered with immediate answers to "what-if" scenarios' based questions.

These are just few inspirations which enlighten us on the value of security data visualization.

There are a lot of books and leaders in this space who can be followed to keep up to date in security data visualization area.

If you need additional information visit the data visualization reference network for wealth of information in this field which is visually catalogued: http://moebio.com/datavisnetwork/

Appendix B

B.0 Information Security Metrics

Before we dive into security visualization, we will review the information security metrics literature which summarizes the current available information on information security metrics. Information security metrics has to be customized to each and every organization, depending on various factors, including senior management preference. Depending on the organization, security metrics should be constantly updated.

B.1 Importance of Information Security Metrics

The following are some of the reasons information security metrics are used for senior management:

- Measurement and communication of whether security investments are done effectively.
- Regular communication of the Information security strategy and the yearly information security work program.
- Regular communication of the value of information (data classification).
- Regular communication of the risk prioritization and how the work program is focused on what is important to the organization.
- Regular explanation of critical incidents along with their impact to the organization.
- Explanation of the different threats the organization would face, prioritized by risk factors.

There are tons of operational security metrics for optimizing operations and to highlight any operational issues related to vulnerability management, incident management, change management, application security reviews, patch management, etc. Some of the operational security metrics are good for technical audience and CISO for enhancing the services.

B.2 Establishing Security Metrics Program

It is beneficial to have a security metrics program within the security team with a process owner instead of generating different ad-hoc metrics from different sub-teams. The security metrics program leader can be empowered with all the data and metrics which are already available. Once all the available information security metrics are reviewed, they have to be compared with the metrics currently generated in the organization. It is useful for the security metrics process owner to conduct a brainstorming session to update the information security metrics and use creative and innovative security visualization to display the data.

NIST 800-55 Rev 1 has candidate measures (metrics) which is a useful short list.

2.3 Information Security Metrics Sources

There is a good amount of material on information security metrics. Below are some of the books and publications that provide a methodology on creating a security metrics work program and candidate metrics which can be chosen to improve the current metrics or to create new metrics.

The below materials and books are valuable resources for selecting and developing good information security metrics.

Andrew Jaquith: Security Metrics (Jaquith, Security metrics: replacing fear, uncertainty, and doubt, 2007)

This books covers broad areas in security metrics and best practices, and offers many effective examples. It contains dedicated chapters on security visualization, creating balanced scorecards, and many foundational ideas.

M. Hoehl: Security Scorecards(Hoehl, "Security Scorecards", 2010)

Creating a monthly information security scorecard for CIO and CFO – describes a good process on creating scorecards. It also covers the relevant legislation and contracts that organizations must comply with.

PRAGMATIC Security Metrics (Brotby, 2013)

This is a comprehensive book with information about various information security metrics. It provides clear examples, case studies, and frameworks for selecting metrics. This is certainly a valuable source of information for reference.

Data Driven Security (Jacobs, J 2014)

This book focuses on security visualization which is the topic of the second part of this paper. It provides a lot of guidance, examples, and ideas to build a data driven culture.

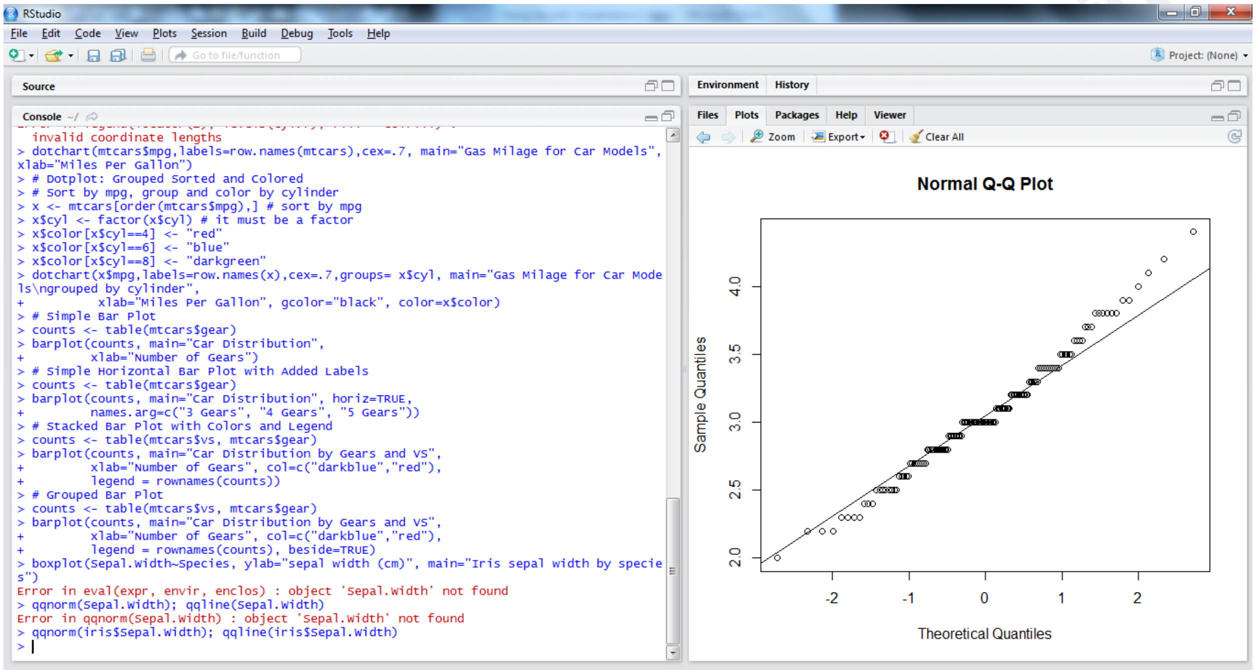
There are many other books and resources in the reference section like CIS metrics and Metricon metrics. The idea was to highlight some important resources available on information security metrics, hoping this will be an inspiration to research more - as required - for additional information, depending on the needs of the organization.

Appendix C

Security Visualization Examples using R

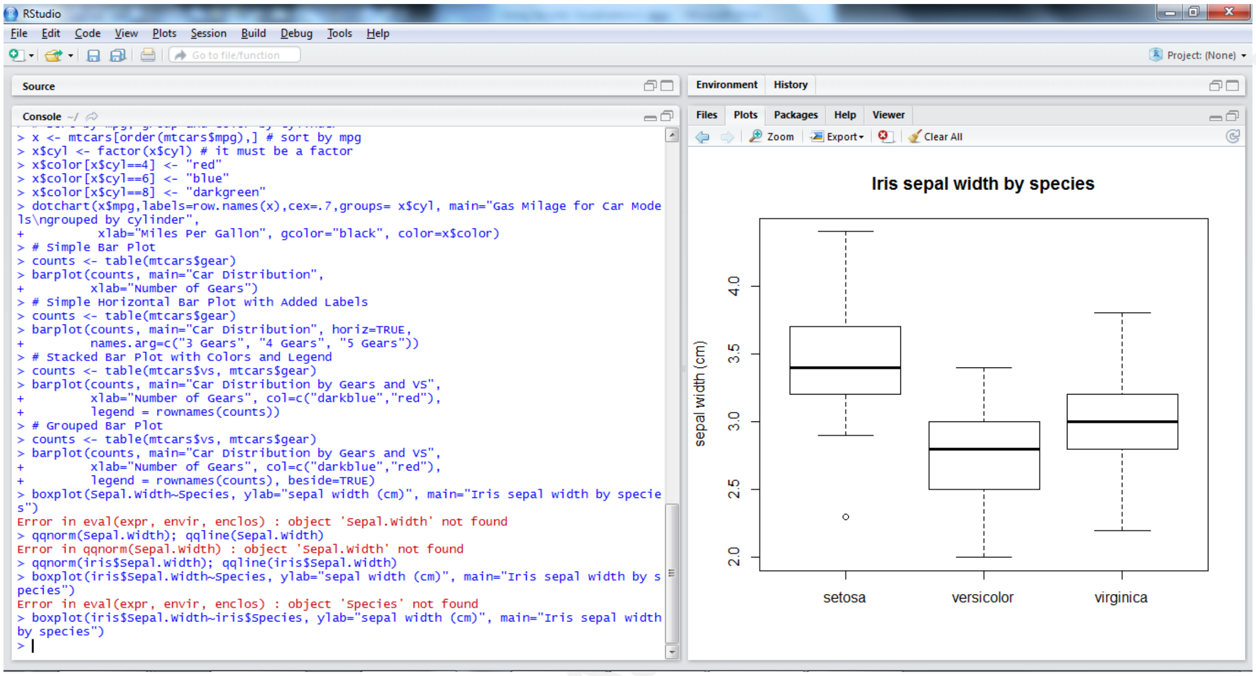
We will use specific examples using R for visualizing some example data already available in R.

Example R code for Normal Q-Q plot



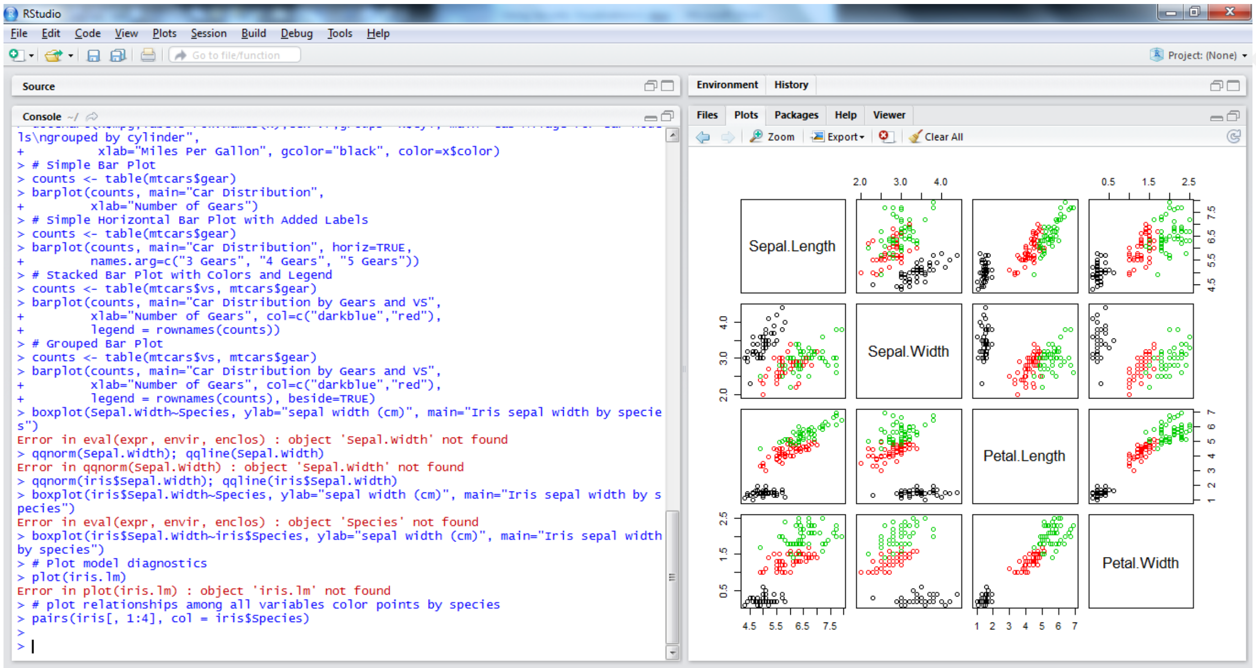
```
  invalid coordinate lengths
> dotchart(mtcars$mpg,labels=row.names(mtcars),cex=.7, main="Gas Milage for Car Models",
xlab="Miles Per Gallon")
> # Dotplot: Grouped Sorted and Colored
> # Sort by mpg, group and color by cylinder
> x <- mtcars[order(mtcars$mpg),] # sort by mpg
> x$cyl <- factor(x$cyl) # it must be a factor
> x$color[x$cyl==4] <- "red"
> x$color[x$cyl==6] <- "blue"
> x$color[x$cyl==8] <- "darkgreen"
> dotchart(x$mpg,labels=row.names(x),cex=.7,groups= x$cyl, main="Gas Milage for Car Mode
ls\ngrouped by cylinder",
+          xlab="Miles Per Gallon", gcolor="black", color=x$color)
> # Simple Bar Plot
> counts <- table(mtcars$gear)
> barplot(counts, main="Car Distribution",
+         xlab="Number of Gears")
> # Simple Horizontal Bar Plot with Added Labels
> counts <- table(mtcars$gear)
> barplot(counts, main="Car Distribution", horiz=TRUE,
+         names.arg=c("3 Gears", "4 Gears", "5 Gears"))
> # Stacked Bar Plot with Colors and Legend
> counts <- table(mtcars$vs, mtcars$gear)
> barplot(counts, main="Car Distribution by Gears and VS",
+         xlab="Number of Gears", col=c("darkblue","red"),
+         legend = rownames(counts))
> # Grouped Bar Plot
> counts <- table(mtcars$vs, mtcars$gear)
> barplot(counts, main="Car Distribution by Gears and VS",
+         xlab="Number of Gears", col=c("darkblue","red"),
+         legend = rownames(counts), beside=TRUE)
> boxplot(Sepal.Width~Species, ylab="sepal width (cm)", main="Iris sepal width by specie
s")
Error in eval(expr, envir, enclos) : object 'Sepal.Width' not found
> qqnorm(Sepal.Width); qqline(Sepal.Width)
Error in qqnorm(Sepal.Width) : object 'Sepal.Width' not found
> qqnorm(iris$Sepal.Width); qqline(iris$Sepal.Width)
> 
```

Example R code for Box Plot



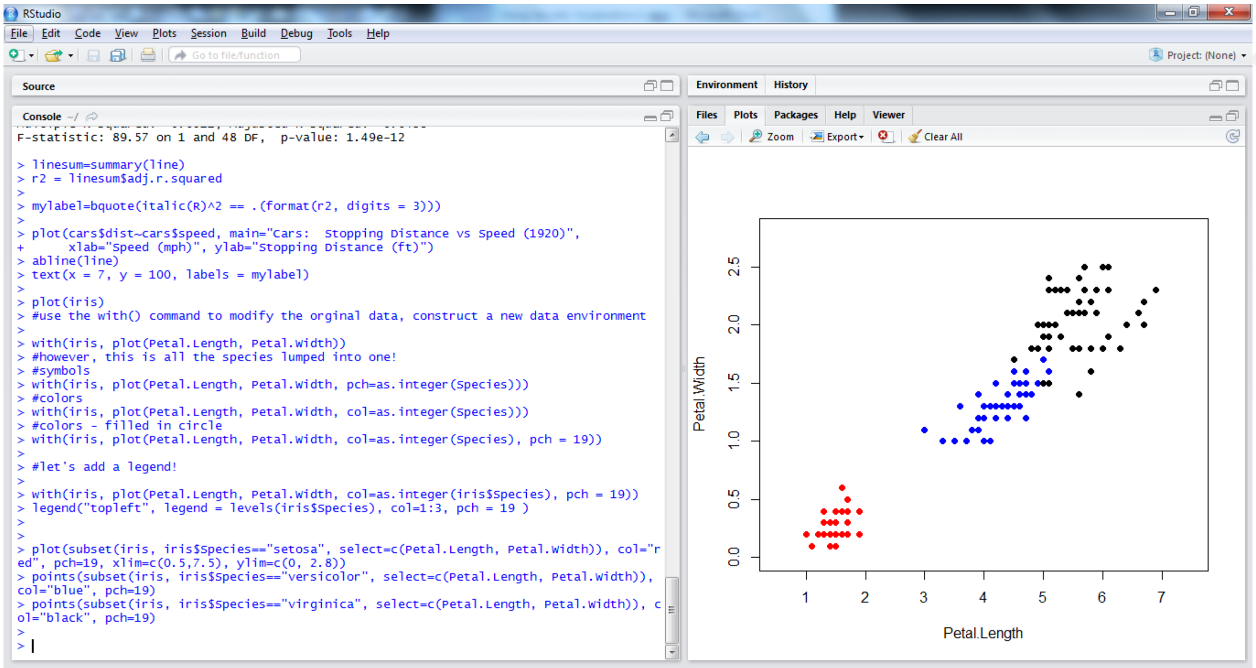
```
> x <- mtcars[order(mtcars$mpg),] # sort by mpg
> x$cyl <- factor(x$cyl) # it must be a factor
> x$color[x$cyl==4] <- "red"
> x$color[x$cyl==6] <- "blue"
> x$color[x$cyl==8] <- "darkgreen"
> dotchart(x$mpg,labels=row.names(x),cex=.7,groups= x$cyl, main="Gas Milage for Car Mode
ls\ngrouped by cylinder",
+          xlab="Miles Per Gallon", gcolor="black", color=x$color)
> # Simple Bar Plot
> counts <- table(mtcars$gear)
> barplot(counts, main="Car Distribution",
+         xlab="Number of Gears")
> # Simple Horizontal Bar Plot with Added Labels
> counts <- table(mtcars$gear)
> barplot(counts, main="Car Distribution", horiz=TRUE,
+         names.arg=c("3 Gears", "4 Gears", "5 Gears"))
> # Stacked Bar Plot with Colors and Legend
> counts <- table(mtcars$vs, mtcars$gear)
> barplot(counts, main="Car Distribution by Gears and VS",
+         xlab="Number of Gears", col=c("darkblue","red"),
+         legend = rownames(counts))
> # Grouped Bar Plot
> counts <- table(mtcars$vs, mtcars$gear)
> barplot(counts, main="Car Distribution by Gears and VS",
+         xlab="Number of Gears", col=c("darkblue","red"),
+         legend = rownames(counts), beside=TRUE)
> boxplot(Sepal.Width~Species, ylab="sepal width (cm)", main="Iris sepal width by specie
s")
Error in eval(expr, envir, enclos) : object 'Sepal.Width' not found
> qqnorm(Sepal.Width); qqline(Sepal.Width)
Error in qqnorm(Sepal.Width) : object 'Sepal.Width' not found
> qqnorm(iris$Sepal.Width); qqline(iris$Sepal.Width)
> boxplot(iris$Sepal.Width~Species, ylab="sepal width (cm)", main="Iris sepal width by s
pecies")
Error in eval(expr, envir, enclos) : object 'Species' not found
> boxplot(iris$Sepal.Width~iris$Species, ylab="sepal width (cm)", main="Iris sepal width
by species")
> 
```

Example R code for Grouped Bar Plot



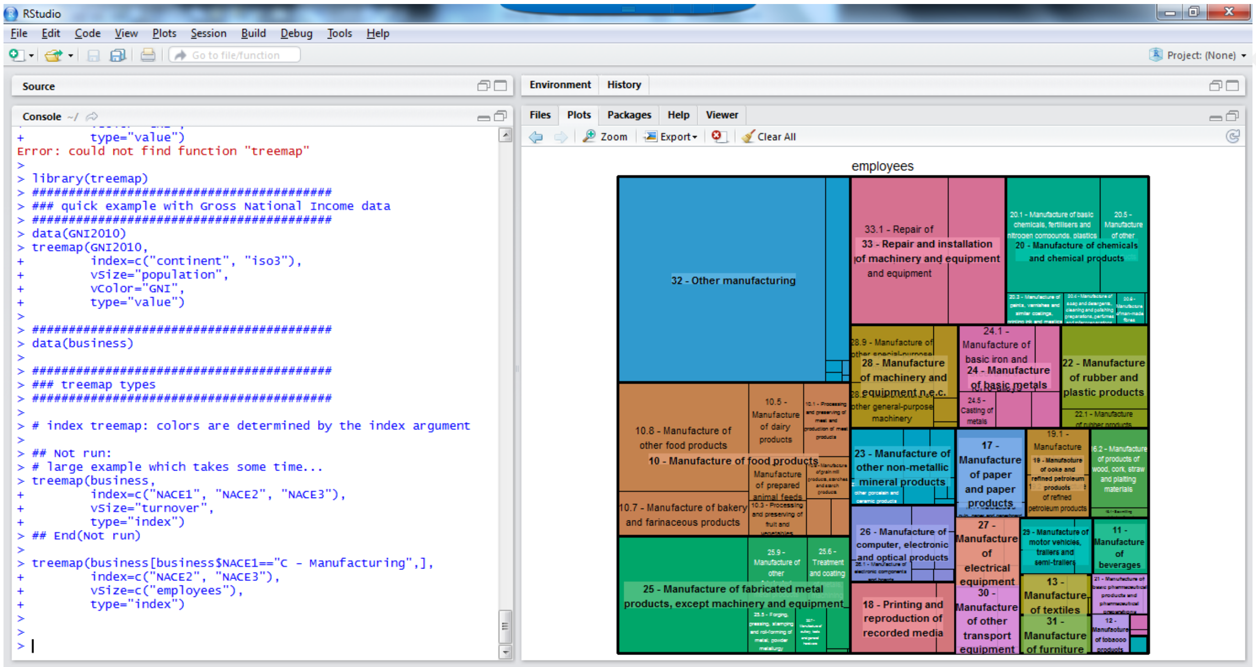
```
ls\ngrouped by cylinder",
+          xlab="Miles Per Gallon", gcolor="black", color=x$color)
> # Simple Bar Plot
> counts <- table(mtcars$gear)
> barplot(counts, main="Car Distribution",
+         xlab="Number of Gears")
> # Simple Horizontal Bar Plot with Added Labels
> counts <- table(mtcars$gear)
> barplot(counts, main="Car Distribution", horiz=TRUE,
+         names.arg=c("3 Gears", "4 Gears", "5 Gears"))
> # Stacked Bar Plot with Colors and Legend
> counts <- table(mtcars$vs, mtcars$gear)
> barplot(counts, main="Car Distribution by Gears and VS",
+         xlab="Number of Gears", col=c("darkblue","red"),
+         legend = rownames(counts))
> # Grouped Bar Plot
> counts <- table(mtcars$vs, mtcars$gear)
> barplot(counts, main="Car Distribution by Gears and VS",
+         xlab="Number of Gears", col=c("darkblue","red"),
+         legend = rownames(counts), beside=TRUE)
> boxplot(Sepal.Width~Species, ylab="sepal width (cm)", main="Iris sepal width by specie
s")
Error in eval(expr, envir, enclos) : object 'Sepal.Width' not found
> qqnorm(Sepal.Width); qqline(Sepal.Width)
Error in qqnorm(Sepal.Width) : object 'Sepal.Width' not found
> qqnorm(iris$Sepal.Width); qqline(iris$Sepal.Width)
> boxplot(iris$Sepal.Width~Species, ylab="sepal width (cm)", main="Iris sepal width by s
pecies")
Error in eval(expr, envir, enclos) : object 'Species' not found
> boxplot(iris$Sepal.Width~iris$Species, ylab="sepal width (cm)", main="Iris sepal width
by species")
> # Plot model diagnostics
> plot(iris.lm)
Error in plot(iris.lm) : object 'iris.lm' not found
> # plot relationships among all variables color points by species
> pairs(iris[, 1:4], col = iris$Species)
>
> |
```

Example R code for Dot Plot - color



```
F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12

> linesum=summary(line)
> r2 = linesum$adj.r.squared
>
> mylabel=bquote(italic(R)^2 == .(format(r2, digits = 3)))
>
> plot(cars$dist~cars$speed, main="Cars:  Stopping Distance vs Speed (1920)",
+      xlab="Speed (mph)", ylab="Stopping Distance (ft)")
> abline(line)
> text(x = 7, y = 100, labels = mylabel)
>
> plot(iris)
> #use the with() command to modify the orginal data, construct a new data environment
>
> with(iris, plot(Petal.Length, Petal.Width))
> #however, this is all the species lumped into one!
> #symbols
> with(iris, plot(Petal.Length, Petal.Width, pch=as.integer(Species)))
> #colors
> with(iris, plot(Petal.Length, Petal.Width, col=as.integer(Species)))
> #colors - filled in circle
> with(iris, plot(Petal.Length, Petal.Width, col=as.integer(Species), pch = 19))
>
> #let's add a legend!
>
> with(iris, plot(Petal.Length, Petal.Width, col=as.integer(iris$Species), pch = 19))
> legend("topleft", legend = levels(iris$Species), col=1:3, pch = 19 )
>
>
> plot(subset(iris, iris$Species=="setosa", select=c(Petal.Length, Petal.Width)), col="r
ed", pch=19, xlim=c(0.5,7.5), ylim=c(0, 2.8))
> points(subset(iris, iris$Species=="versicolor", select=c(Petal.Length, Petal.Width)),
col="blue", pch=19)
> points(subset(iris, iris$Species=="virginica", select=c(Petal.Length, Petal.Width)), c
ol="black", pch=19)
>
> |
```

Example of R code for Treemap



RStudio
File Edit Code View Plots Session Build Debug Tools Help
Go to file/function
Project: (None)
Source
Console ~/
Environment History
Files Plots Packages Help Viewer
Zoom Export Clear All
employees
32 - Other manufacturing
33.1 - Repair of
33 - Repair and installation of machinery and equipment
and equipment
20.1 - Manufacture of basic chemicals, fertilisers and nitrogen compounds, plastics
20.5 - Manufacture of other
20 - Manufacture of chemicals and chemical products
28.9 - Manufacture of other special-purpose
28 - Manufacture of machinery and equipment n.e.c.
other general-purpose machinery
24.1 - Manufacture of basic iron and
24 - Manufacture of basic metals
24.5 - Casting of metals
22 - Manufacture of rubber and plastic products
22.1 - Manufacture of rubber products
10.8 - Manufacture of other food products
10.5 - Manufacture of dairy products
10.1 - Processing and preserving of meat and production of meat products
10 - Manufacture of food products
Manufacture of prepared animal feeds
10.7 - Manufacture of bakery and farinaceous products
10.3 - Processing and preserving of fruit and vegetables
23 - Manufacture of other non-metallic mineral products
other porcelain and ceramic products
17 - Manufacture of paper and paper products
19.1 - Manufacture
19 - Manufacture of coke and refined petroleum products
of refined petroleum products
16.2 - Manufacture of products of wood, cork, straw and plaiting materials
26 - Manufacture of computer, electronic and optical products
26.1 - Manufacture of electronic components and boards
27 - Manufacture of electrical equipment
29 - Manufacture of motor vehicles, trailers and semi-trailers
11 - Manufacture of beverages
25.9 - Manufacture of other
25.6 - Treatment and coating
25 - Manufacture of fabricated metal products, except machinery and equipment
25.5 - Forging, pressing, stamping and roll-forming of metal; powder metallurgy
18 - Printing and reproduction of recorded media
30 - Manufacture of other transport equipment
13 - Manufacture of textiles
31 - Manufacture of furniture
21 - Manufacture of basic pharmaceutical products and pharmaceutical preparations
12 - Manufacture of tobacco products

```
+         type="value")
Error: could not find function "treemap"
>
> library(treemap)
> ########################################
> ### quick example with Gross National Income data
> ########################################
> data(GNI2010)
> treemap(GNI2010,
+         index=c("continent", "iso3"),
+         vSize="population",
+         vColor="GNI",
+         type="value")
>
> ########################################
> data(business)
>
> ########################################
> ### treemap types
> ########################################
>
> # index treemap: colors are determined by the index argument
>
> ## Not run:
> # large example which takes some time...
> treemap(business,
+         index=c("NACE1", "NACE2", "NACE3"),
+         vSize="turnover",
+         type="index")
> ## End(Not run)
>
> treemap(business[business$NACE1=="C - Manufacturing",],
+         index=c("NACE2", "NACE3"),
+         vSize=c("employees"),
+         type="index")
>
>
> |
```

Security Incident Trending Metrics Example using R

The challenge for some of the examples is that there are very few public security datasets available. Some data sources used in this paper are VERIS framework data and some sources stem from the "Data-Driven Security" blog.

In this example, dummy security incident data is created in the format shown below with file name incidentpriority1.csv.

Date ID Severity Type:

1 1/1/2014 IM-21 High Hacking

2 1/2/2014 IM-25 Medium Phishing

This script was adapted to meet incident metrics, originally developed for Bugs by Barker, Tom - Pro Data Visualization using R and JavaScript.

In this script, the dummy incident data will be used to generate time charts which are effective metrics to compare changes over time.

```
incidentpriority1 <- read.csv("C:/Temp/incidentpriority1.csv")
```

Here the data is loaded in R using read.csv and viewing the records using “view command”.

```
View(incidentpriority1)
```

This displays the records.

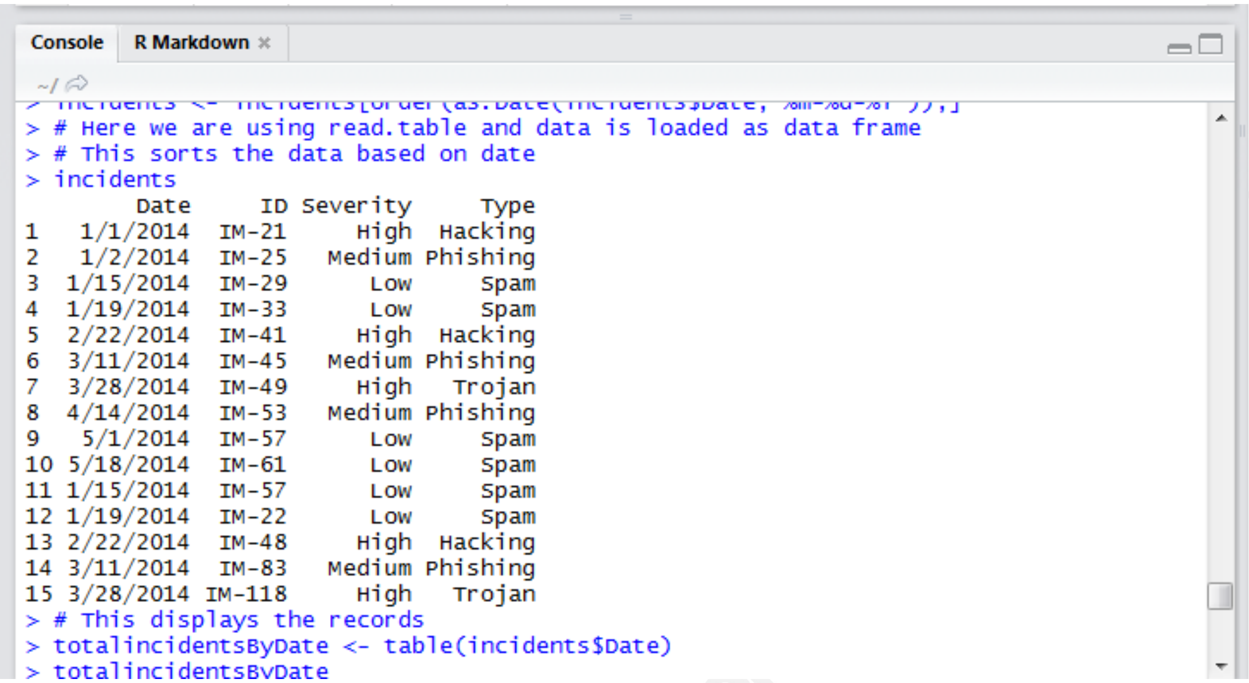
```
incidents <- read.table("C:/Temp/incidentpriority1.csv", header=TRUE, sep=",")
```

```
incidents <- incidents[order(as.Date(incidents$Date,"%m-%d-%Y")),]
```

Here we are using a read.table and the data is loaded as a data frame. This sorts the data based on date.

```
incidents
```

This displays the records.



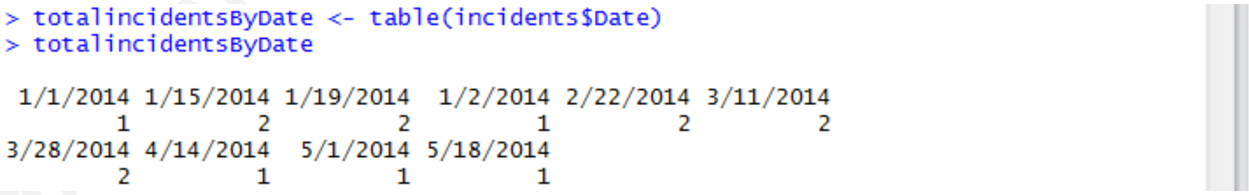
```
> incidents <- incidents[order(as.Date(incidents$Date, "%m-%d-%Y")),]
> # Here we are using read.table and data is loaded as data frame
> # This sorts the data based on date
> incidents
         Date     ID Severity     Type
1    1/1/2014  IM-21     High  Hacking
2    1/2/2014  IM-25   Medium Phishing
3   1/15/2014  IM-29      Low     Spam
4   1/19/2014  IM-33      Low     Spam
5   2/22/2014  IM-41     High  Hacking
6   3/11/2014  IM-45   Medium Phishing
7   3/28/2014  IM-49     High   Trojan
8   4/14/2014  IM-53   Medium Phishing
9    5/1/2014  IM-57      Low     Spam
10  5/18/2014  IM-61      Low     Spam
11  1/15/2014  IM-57      Low     Spam
12  1/19/2014  IM-22      Low     Spam
13  2/22/2014  IM-48     High  Hacking
14  3/11/2014  IM-83   Medium Phishing
15  3/28/2014 IM-118     High   Trojan
> # This displays the records
> totalincidentsByDate <- table(incidents$Date)
> totalincidentsByDate
```

totalincidentsByDate <- table(incidents$Date)

Here we are passing incidents$Date to table() function to build a data structure of counts of incidents each day.

totalincidentsByDate

This displays the number of incidents per day.



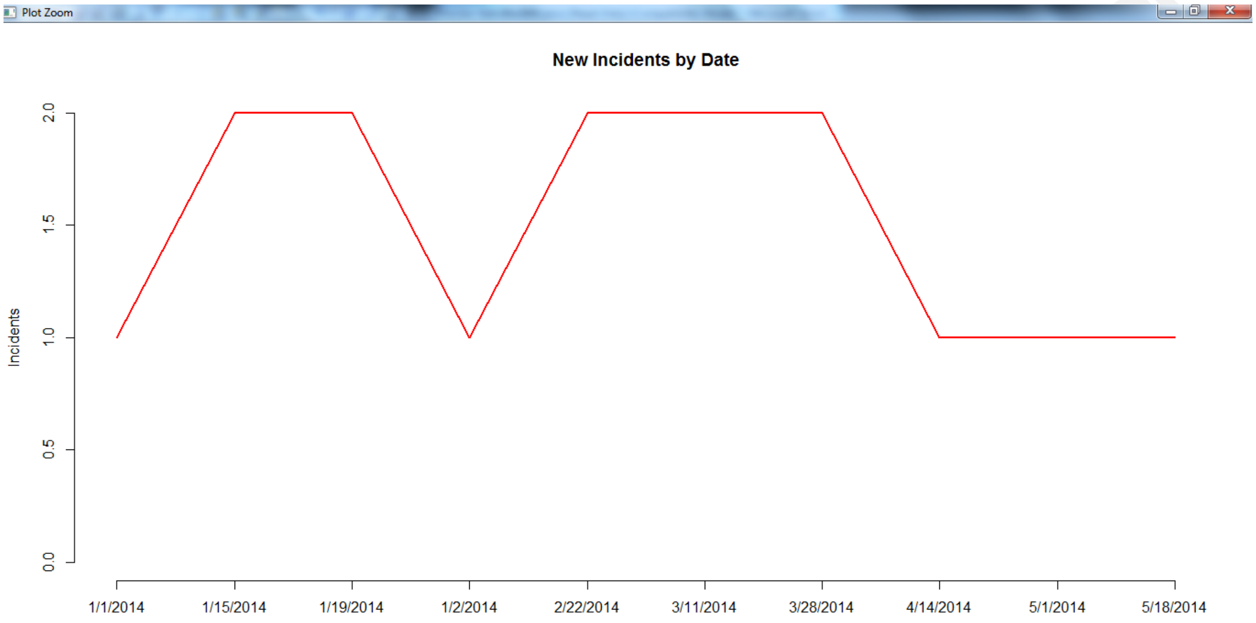
```
> totalincidentsByDate <- table(incidents$Date)
> totalincidentsByDate

 1/1/2014 1/15/2014 1/19/2014  1/2/2014 2/22/2014 3/11/2014 
        1         2         2         1         2         2 
3/28/2014 4/14/2014  5/1/2014 5/18/2014 
        2         1         1         1 
```

Creating a Basic Plot

plot(totalincidentsByDate, type="l", main="New Incidents by Date", col="red", ylab="Incidents")

The plot is used to chart how many incidents are created per day, additional examples are in the appendix for plot.

This creates the chart showing how many incidents are created per day.



runningTotalincidents <- cumsum(totalincidentsByDate)

runningTotalincidents

This created a cumulative sum of total incidents by date.

This is just a short example to show how R can be used for effectively visualizing security incident trend metrics.

Building Security Data Visualization Toolbox

There are many security data visualization tools available.

The website http://secviz.org/ has a list of all available tools that have been packaged in the DAVIX distribution.

The book Marty, R. (2009). Applied security visualization also has lot of examples and guidance on developing security visualization using most of the tools in DAVIX.

Apart from the tools in DAVIX package there is a lot of other data visualization tools.

Some of the sophisticated ones are Tableau, Gephi, RapidMiner, LightSide, Afterglow, FlowBAT, iNZight, R , Opengraphiti.

R is a free, open source language with access to powerful, cutting-edge analytics with more than 2000 packages.

Mondrian

Once you have the data , in few clicks Mondrian can be used to generate Plots, Histograms, Boxplots, Scatterplots, Barcharts and Mosaic Plots.

This does not require any coding and very simple for security practitioner to generate interactive visualizations.

The below web page have numerous additional examples of information security data visualization for getting inspiration on use cases and applications of security data visualization:

http://secviz.org/category/image-galleries/graph-exchange

New Tools

OpenGraphiti - OpenDNS Data Visualization Framework

OpenGraphiti is a free and open source 3D data visualization engine for data scientists to visualize semantic networks and to work with them. It offers an easy-to-use API with several associated libraries to create custom-made datasets. It leverages the power of GPUs to process and explore the data and sits on a homemade 3D engine.

The below white-paper explains in detail the OpenGraphiti framework

https://www.blackhat.com/docs/us-14/materials/us-14-Hay-Unveiling-The-Open-Source-Visualization-Engine-For-Busy-Hackers-WP.pdf

Use Cases with OpenGraphiti

There are a seemingly endless number of applicable use cases for the visualization of loosely related data. Some examples include the analysis of security data, malware infection alerts could be visualized to expose a previously unrecognized patterns in a malicious actor activity, or even a misconfiguration of a technical control that allows too much, or too little, access to data, files, or networks.

Appendix C - References for R

R - Getting Started. (n.d.). Retrieved December 16, 2014, from http://data.princeton.edu/R/gettingStarted.html

Anthony Tanbakuchi. (n.d.). Retrieved December 16, 2014, from http://www.tanbakuchi.com/Resources/R_Statistics/RBasics.html

Kembel @ UQAM. (n.d.). Retrieved December 16, 2014, from http://phylodiversity.net/skembel/r-workshop/introR/SK_Intro_R.html

Intro to Graphing in R. (n.d.). Retrieved December 16, 2014, from http://gradquant.ucr.edu/wp-content/uploads/2013/11/GraphingRworkshop.txt

Images/WorldBank_demo_with_googleVis.R. (n.d.). Retrieved December 16, 2014, from https://gist.github.com/mages/5180860#file-worldbank_demo_with_googlevis-r

Pro Data Visualization using R and JavaScript. (2013). Berkeley, CA: Apress.