Chapter 1: Unleashing The Securing Power Of Data

A Brief History Of Learning From Data

19th Century Data Analysis

20th Century Data Analysis

21st Century Data Analysis

Gathering Data Analysis Skills

Domain Expertise

Programming Skills

Data Management

Statistics

Visualization (a.k.a. Communication)

Combining The Skills

Centering On A Question

Steps To Creating A Good Research Question

Exploratory Data Analysis

Chapter 2: Building Your Analytics Toolbox: A Primer on Using R & Python for Security Analysis

About This Chapter

Why Python; Why R; ***And***, Why Both?

Jumpstarting Your Python Analytics With Canopy

Understanding The Python Data Analysis And Visualization Ecosystem

Setting Up Your R Environment

Introducing Data Frames

Organizing Analyses

Chapter 3: Learning The "Hello World" Of Security Data Analysis

Solving A Problem

Getting Data

Reading In Data

Exploring Data

Honing In On A Question

In Summary

Recommended Reading

Chapter 4: Analyzing “Badness”

Dissecting the “IP Address”

Representing IP Addresses

Segmenting And Grouping IP Addresses

Locating IP Addresses

Augmenting IP Address Data

Association/Correlation, Causation And Security Operations Center Analysts Gone Rogue

Mapping Outside The Continents

Visualizing The ZeUS Botnet

Visualizing Firewall “Badness”

In Summary

Recommended Reading

Chapter 5: From Maps To Regression (Jay)

Simplifying Maps

Discovering How Many ZeroAccess Infections There Are Per Country

Understanding “Potwin Effect”

Is This Weird?

Counting In Countries

Does Variation Stack Or Cancel?

Moving Down To Counties

Introducing (Briefly) Linear Regression

Understanding Common Pitfalls In Regression Analysis

We Cannot Extrapolate Beyond Our Data

Outliers Have A Lot Of Influence

Hidden Relationships Hide Well

Too Many Variables

Visualize And Apply The Sniff Test

Apply Linear Regression To Linear Relationships

Regression On ZeroAccess Infections

What Is Correlated To ZeroAccess Infections?

Chapter 6: Improving Your Security-oriented Visualizations (Jay)

*[This introduces the core concepts of data visualization using firewall session data for the examples; this will set up future discussions on both visualizations and the data set]*

Chapter Use Case: Exploring Your Firewall Data (Jay: Severski’s Data)

Understanding The Foundations Of Good Visual Communication

Position, Length, Angle, Slope, Area, Volume, Color Saturation and Hue

Addressing the silent minority (colorblind)

Leverage use case to run through basics and then comparisons of core (bar/dot/line/pie/scatterplots) charting techniques

Remembering the communication

Leveraging more than visuals, tables and words

Visualizing tabular data

WEB CONTENT: [visual defaults use case] R code, python code

WEB CONTENT: Firewall data set, R code, visuals

Chapter 7: Getting A Handle On Your Security Data With Descriptive Statistics And Descriptive Visualization (Jay)

*[Advanced visuals with time-series data and distributions; visuals as a descriptive tool for univariate (or even multivariate) data. Will also create a video or two from the firewall data]*

Chapter Use Case: Expanding from Chapter 5 with Severski’s Data

Describing Data Over Time

Moving averages

Time-series plots

Probability density function

Describing Collections of Data with Numbers

How to use descriptive statistics in a security context

distributions, probability density, quantiles, measures of center, standard deviation and skew

Describing Collections of Data Visually

How to use descriptive visualizations in a security context

Box plots, histograms, etc

Moving with Animations

Understanding The Challenges Of Visualizing Lots Of Data

Radial Graphs Example and other hair-ball diagrams

WEB CONTENT: R code, visuals

Chapter 8: Learning From Security Breaches (Bob)

*[This chapter will talk about collecting data from processes rather than logs and how we can learn from failures by collecting and analyzing breach data]*

Turning Chaos Into Structure

The Power Of Structured Recording During An Incident

Understanding & Using VERIS

Comparisons To Other Methods (Strengths and limitations)

Being Cautious About Inferential Estimations

USE CASE: Visualizing VERIS Community Database (Breach Data)

Callout: The Cost-per-datum Challenge

Looking At And Learning From Other Community Breach Data Sets

USE CASE: PRC Aggregated Breaches

WEB CONTENT: breach data, R code, visuals

Chapter 9: Breaking Up With Your Relational Database (Bob)

*[We haven’t touched on data management too much, but this chapter will dive into different approaches to storing and accessing data and the options we have and the recent advancements that open up our options]*

Realizing The Container Has Constraints

Understanding The Limitations Of A Monolithic Data Store

Tables can introduce unnecessary complexity

Exploring Alternative Data Stores

Survey of core SQL alternatives, identifying strengths and uses each in context

USE CASE: "Have we seen this IP address?"

practical example of how a traditional monolithic approach can hinder use of critical threat intelligence and how re-thinking how you intake, crunch and store data can open up new possibilities

WEB CONTENT: sample code for the use case

Chapter 10: Having The Machine Learn For You (Jay)

*[This will give an introduction to Machine Learning in the context of either netflow data or the same firewall data from chapters 6 and 7. Purpose here is to show methods of ML, not to do fancy analysis here]*

De-mystifying Machine Learning

Will discuss the surprisingly straightforward underpinnings of ML and setup the rest of the chapter

Understanding The Security Potential of ML

Unsupervised Learning: Clustering Host Activity

Having the computer figure out the patterns we can’t see

Applying Multi-dimensional scaling and visualizing Euclidian distances

K-nearest neighbors method

K-means method

Learning point: machine can profile “good” and “bad” behaviors

Supervised Learning: Classifying Host Activity *(note: may not do these, depending on data sources)*

Training the computer on patterns we see

Logistic regression

Random forests

Chapter 11: Designing Effective Security Dashboards (Bob)

Designing Dashboards For Effective Security Response

Dashboards are a call to action

Making differences stand out

The never-ending quest for "so what?"

Applying Appropriate Visualizations To Your Security Data Streams

Knowing when and how to use line graphs, bar charts, maps, etc

Recognizing The Importance of Baselines And Thresholds

Don’t Be “Mean”: There’s A Reason For The Five Number Summary

Communicating With Dashboards

Understanding Your Audience

Designing For The Single Screen/“Pane Of Glass”

Building Static Dashboards

USE CASE: The Security Operations Center Daily Operational Dashboard

USE CASE: The Threat Management Dashboard

Developing Dynamic Dashboards

USE CASE: The GRC Drill Down

USE CASE: The Vulnerability Management View

Chapter 12: Building Interactive Security Visualizations

*[This chapter expands the chapter 10 “dashboards” chapter to show readers how to make richer and appropriately informative interactive security visualizations.]*

Moving From Static To Interactive

Knowing when static makes sense

Adding Interaction To Enhance Understanding

Avoiding Interactivity “Land Mines”

Developing Interactive Visualizations

Reviewing Your Choices

Choosing A Storage Medium

Transporting The Data From Storage To Browser

Developing User-centric Visualizations

USE CASE: Building An Interactive Nessus Vulnerability Explorer

WEB CONTENT: self-contained VM image with the stack & visualizations

Chapter 13: Keeping It Simple (Bob/Jay)

Putting Security Data Analysis Into Perspective

Comparing A "Drilling For Oil" Approach To a "Pan For Gold" Approach

Understanding The Reality Of Our Environments

Re-iterating That Data Analysis Assists Our Thinking, Not Replaces It

What Lies Ahead In Security Analytics?

Appendix A: List of Resources and Tools

Data Cleansing

Analytics

Visualization

Misc

Timeline:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Today’s Date:** | 2013-07-08 |  |  |  |  |
| **Author(s):** | Jay Jacobs | Bob Rudis |  |  |  |
| **Book Title:** | Security Data Science: Security Using Data Analysis, Visualization and Dashboards |  |  |  |  |
| **Ms Page/Word Count:** |  |  |  |  |  |
| **Ms Due Date:** |  |  |  |  |  |
| **Chapter #** (only sample numbering here-you may have more or less chapters and you may have Parts; label Parts as Part I, Part II, etc.) | **Chapter Title** (assignment of writing to individual author if there is more than one author) | **Approximate number of pages per chapter** | **Estimated figure count for each chapter** | **Date** you will submit to Wiley the complete chapter ready for editing & development by Wiley | **Author for who is writing the chapter** |
| Preface/Introduction |  |  |  | 10/14/2013 |  |
| 1 | Unleashing The Securing Power Of Data | 20 | 4 | 7/21/13 | Jay |
| 2 | Building Your Analytics Toolbox | 12 | 4 | 8/11/13 | Bob and Jay |
| 3 | Learning The "Hello World" | 26 | 6 | 7/21/13 | Bob |
| 4 | Analyzing “Badness” | 26 | 6 | 7/21/13 | Bob |
| 5 | Mapping “Badness” | 26 | 8 | 7/21/13 | Jay |
| 6 | Improving Your Security-oriented Visualizations | 26 | 12 | 8/11/13 | Jay |
| 7 | Getting A Handle On Your Security Data | 26 | 6 | 9/1/13 | Jay |
| 8 | Learning From Security Breaches | 26 | 4 | 9/1/13 | Bob |
| 9 | Breaking Up With Your Relational Database | 26 | 2 | 8/11/13 | Bob |
| 10 | Having The Machine Learn For You | 30 | 4 | 9/29/13 | Jay |
| 11 | Designing Effective Security Dashboards | 30 | 10 | 9/29/13 | Bob |
| 12 | Building Interactive Security Visualizations | 26 | 8 | 10/20/13 | Bob |
| 13 | Keeping it Simple | 20 | 0 | 10/20/13 | Jay |
| Appendix | Resources and Tools | 6 | 0 | 10/27/13 | Bob and Jay |
|  | References |  | 0 | 10/27/13 | Both |
| Index (Wiley provides) |  |  |  |  |  |
|  | Totals | 326 | 74 |  |  |