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

In Summary

Recommended Reading

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

In Summary

Recommended Reading

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

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?

In Summary

Recommended Reading

Chapter 6: Communicating Visually (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)

Visual Thinking: how visually decoding impacts how we encode

Setting up dashboard chapters: pre-attentive processing, iconic memory, etc

Understanding The Foundations Of Good Visual Communication

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

Addressing the silent minority (colorblind)

Core charting techniques: bar/scatter/line/pie, etc

Describing Data Over Time

Moving averages

Time-series plots (variety here, heatmaps, etc)

Probability density function

Describing Collections of Data Visually

Beyond the core charting techniques: Box plots, histograms, etc

Moving with Animations

Understanding The Challenges Of Visualizing Lots Of Data

Radial Graphs Example and other hair-ball diagrams

In Summary

Recommended Reading

Chapter 7: Learning From Security Breaches (Jay)

*[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

In Summary

Recommended Reading

Chapter 8: 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

RDBMS tables & normalization 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

USE CASE: "Finding hidden ‘badness’.”

practical example of how to transform logs into a relational database to quickly discover potentially malicious traffic.

In Summary

Recommended Reading

Chapter 9: Having The Machine Learn For You (Bob)

*[This will give an introduction to Machine Learning in the context of neflow/firewall data previous chapters. 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

Training the computer on patterns we see

Logistic regression

Random forests

In Summary

Recommended Reading

Chapter 10: 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 Summaries

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

In Summary

Recommended Reading

Chapter 11: Building Interactive Security Visualizations (Bob)

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

In Summary

Recommended Reading

Chapter 12: Keeping It Simple (Jay) - Focus

We’ve covered a lot, but you don’t have to know it all

Pick out a couple of simple things you enjoy and develop those, for example we did basic data munging and visualizaitons.

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

--- Don’t shoot for a blazing amount of success, small victories and targeted approach. start small and grow from there.

There is a tremendous amount of iterations, trial and error is the name of the game and things must be repeatable and transferable.

establish a focus,

Venn diagram -

With a focus on security people, focus on “domain expertise”, “hacking (and programming) skills” and statistics. Call out where to go to develop each area.

How do our chapters fit into the Venn Diagram?

Here’s what it means to do development ,source control, unit testing, source code review, moving from “hacking” to “profession” (keep in mind the single-use code).

There is a significant amount of trial and error, hence the source control and the need for good coding skills.

Putting Security Data Analysis Into Perspective

Understanding The Reality Of Our Environments - don’t be disheartened by “noisy” data, mess is the world. Don’t toss out data that is perceived as dirty but at the same time acknowledge the dirty data and don’t give it more meaning than the data can support. It’s also just not the data, the tools may have problems and need some TLC (recompiling libxml2, for example).

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

What Lies Ahead In Security Analytics?

Personalize this.

Top 3: Strat

1) Ask questions with objective answers, - stress this shift in perspective.

2) need data to answer it (reference ch 8, “have we seen this ip address”)

3) try everything, be okay with failure

4) get statistics “Got Stats?”

Operational

1) find data (data inventory), to enable prioritization

2) have we seen this IP address?

Appendix A: The Development Side Of Data Science

Adapting The Development Mindset To Security Data Analysis

Learning From Decades Of Software Development Evolution

Embracing A Cycle Of Continuous Learning

Appendix B: List of Resources and Tools

Data Cleansing

Analytics

Visualization

Color Resources (mentioned in Chapter 6)

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 |  |  |  | 10/27/2013 |  |
| 1 | Unleashing The Securing Power Of Data | 26 | 4 | 7/21/13 | Jay |
| 2 | Building Your Analytics Toolbox: A Primer on Using R & Python for Security Analysis | 16 | 5 | 8/11/13 | Bob and Jay |
| 3 | Learning The "Hello World" | 29 | 14 | 7/21/13 | Bob |
| 4 | Analyzing “Badness” | 28 | 9 | 7/21/13 | Bob |
| 5 | From Maps To Regression | 32 | 11 | 7/21/13 | Jay |
| 6 | Improving Your Security-oriented Visualizations | 26 | 12 | 8/31/13 | Jay |
| 7 | Learning From Security Breaches | 26 | 5 | 9/21/13 | Jay |
| 8 | Breaking Up With Your Relational Database | 26 | 4 | 10/12/13 | Bob |
| 9 | Having The Machine Learn For You | 30 | 6 | 10/12/13 | Bob |
| 10 | Designing Effective Security Dashboards | 30 | 12 | 8/31/13 | Bob |
| 11 | Building Interactive Security Visualizations | 26 | 10 | 9/21/13 | Bob |
| 12 | Keeping it Simple | 20 | 2 | 10/27/13 | Jay |
| Appendix A | The Development Side of Data Science | 12 | 2 | 10/27/13 | Bob and Jay |
| Appendix B | Resources and Tools | 4 | 0 | 10/27/13 | Bob and Jay |
|  | References |  | 0 | 10/27/13 | Both |
| Index (Wiley provides) |  |  |  |  |  |
|  | Totals | 326 | 96 |  |  |