Chapter 1: Unleashing The Securing Power Of Data (Jay)

Standing On The Shoulders Of Giants

Use cases showing brief examples how other fields moved towards being more data drivn, but with different emphases: battling 'expertise', transforming science and applying data to a gut-feel industry)

- Snow

- Fisher

- Demming

Developing the Skills Of A Security Data Scientist

- Curosity

- Domain (security) knowledge

- Programming

- Data munging & management

- Statistics

- Visualization

Recognizing The Importance Exploratory Data Analysis

- Choosing A Good Research Question

\* "What is the…"

\* "Are there similarities…"

\* "Are there differences…"

- Exploratory Security Data Analysis : The Zero Access Botnet

\* Source data: "So what?"

\* Posing a good Question

\* The quest for the answer

WEB CONTENT: ZeroAccess data, visualizations

Chapter 2: Learning The "Hello World" Of Security Data Analysis (Bob)

Reading in data

Getting a feel for the data

Basic Stats

Basic Graphs

Web scraping

Augmenting your analysis with different, relevant data

Asking a basic question of the data

Which country has the most malicious hosts of type “x”

WEB CONTENT: AlienVault data (w/link to source), python code, R code, visuals

Chapter 3: Analyzing “Badness” (Bob)

Chapter Use Case: Crunching AlienVault IP Reputation Data

What is an “IP Address”?

32-bit integer (“how does your computer see an IP address?”) + machine info

Part of a subnet / logical layout / MAC addresses

Perhaps has a hostname (DNS)

Larger context: part of a global network organized by ASNs (BGP)

Lager context: Has a physical location

Mapping Outside The Continents

USE CASE: Visualizing AlienVault ASN data (force-directed network graphs of malhost ASN groupings)

WEB CONTENT: AlienVault data (w/link to source), python code, R code, visuals

Chapter 4: Mapping “Badness” (Jay)

Chapter Use Case: ZeroAccess Botnet Analysis

What can you learn from just a set of lat/lon pairs?

Getting basic lat/long metadata (lat/lon -> country/city)

Visualizing lat/long data

Choropleth

Dot plot

Getting more advanced metadata (internet user population & income)

The quest for correlation

WEB CONTENT: ZeroAccess code & visuals

Chapter 5: Improving Your Security-oriented Visualizations (Jay)

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

Understanding The Foundations Of Good Visual Communication

- Focused overview of the basics & challenges of human visual perception with references to other books for deeper investigation

- USE CASE: Improving visual defaults in [Python|R] to enhance communication

Moving From Tables And Spreadsheets To A More Visual Medium

- Why we use tables (and how to use them better)

- Visualizing tabular data

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

- Scatterplots

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

WEB CONTENT: Firewall data set, R code, visuals

Chapter 6: Getting A Handle On Your Security Data With Descriptive Statistics And Descriptive Visualization (Jay)

Describing Attributes of IP Addresses Over Time

USE CASE: Expanding from Chapter 5 with Severski’s Data

- How to use descriptive statistics in a security context

\* Univariate analysis (distributions, central tendency, dispersion, variance, standard deviation)

Re-Orienting Your Analyses With Visualizations

- Box-plots

- Histograms

- Animation

Understanding The Challenges Of Visualizing Lots Of Data

Radial Graphs Example

WEB CONTENT: R code, visuals

Chapter 7: Learning From Security Breaches (Bob)

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 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 8: Breaking Up With Your Relational Database (Bob)

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 9: Having The Machine Learn For You (Jay)

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

- Applying MDS, KNN & K-Means techniques to security data

- USE CASE: Predicting potential rogue behaviour with security data

Supervised Learning: Classifying Host Activity

- Using logistic regression / random forests to detect network intrusions

Chapter 10: Building Dynamic 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

Understanding The Importance of Baselines And Thresholds

Communicating With Dashboards

- There is no One Security Dashboard to rule them all; designing security dashboards

USE CASE: The Incident Response Manager's Dashboard

- At-a-glance overview

- current incidents in play

- broad view of incidents over time

- insight about the incident handler team members

- etc …

USE CASE: The Threat Management Dashboard

- Tactical overview of internal threat landscape

- Communication attempts (success/fail) to "badness" (i.e. Matched IoCs)

\* to IPs/ASNs

\* from what (servers/workstations)

\* possibly geo-located

- Successful malware infections trends w/emphasis on known 0-days

- etc …

Chapter 11: 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?