Security Through Data Analysis

Security with Numbers

1. Unleashing the Power of Data

Opening chapter will describe some of the challenges in information security and why data analysis is an important component for us moving forward.

1. Unleashing your inner data scientist

Explain how we can do a lot of really cool analysis and tasks without needing a degree in statistics. This chapter will also outline the types of skills we will cover in this book (statistics, programming, scripting, database management and visualization techniques) and explain why each one is important and how much skill the reader should expect to develop.

1. Why 35 is no different than 37

This chapter will be the first (of two) chapters on inferential statistics and will begin with a (brief, very brief) section on descriptive statistics, but jump into sample size, confidence intervals and hypothesis testing (is an observation of 35 different than another of 37?)

1. Why <correlation statement>

Correlation statement: how we see employee productivity in egress firewalls (correlate bandwidth usage with employee productivity), or perhaps, insider misuse with login attempts (there are often indications of “testing the water” for insiders before they go hog wild).

This chapter will cover correlation versus causation and discuss correlation techniques (pearson and scatter plots). This will lead into a high level introduction to regression techniques, but only from “what is it” discussion not necessarily how to perform and interpret regression analysis.

1. Data Munging

This chapter will cover data sources, data collection and cleaning and/or normalizing of data. (maybe we want to talk about normalizing across disparate comparisons too here - through proportions or z-scores, etc)

1. Storing and accessing Data

This will cover traditional databases and the new breed of NoSQL solutions. Should go into strengths and weaknesses of each and help in selecting and using a data storage mechanism and hit the buzz words: Hadoop, mongo, etc.

1. (Avoiding) Spatial data

Projecting the virtual world onto the physical may not be useful. and once the pitfalls of mapping are covered, we could cover map projections and basic mapping techniques, maybe get into lat/long calculations.

1. Data Viz

Intro to data viz concepts, mapping data types to visualization types end up with making pretty excel charts.

1. Making the machine learn for you

Intro to machine learning concepts, give 2 examples: supervised learning and unsupervised learning both from infosec. - detecting failed logins

1. Making the machine read for you

natural language processing - need example (not spam) from infosec here.

1. Looking to (and predicting) the Future (like a boss)

End with a chapter on predictive analytics.