# Introduction

1. The power of a story
   1. We already have stories
      1. We could do an experiment here, (shostack, “how common is it for attackers to pile on to a victim?”)
   2. Harnessing the power of stories with the strength of science and the clarity of visualizations.
      1. There is no tool that can spin a customized story
      2. While stories are more or less a natural part of being human, converting data into a clear message is not a natural skill
         1. Hence we need a book like this!

# Finding the Story

1. Data Analysis:

“The good news is that we have the field of statistics to rely on, the bad news is that we have the field of statistics to rely on”

* 1. Strengths and limitations of statistics
     1. An entire field devoted to improving the understanding our environment, but it generally lacks context (the human element)
  2. Fooled by Randomness (pick another title, but same concept)
     1. Seeing patterns where none exists
     2. Not seeing patterns where they do exist
        1. Challenge to conventional wisdom (a la moneyball)
        2. Quick reference to type I and type II errors
  3. Being truthful, due diligence in data analysis

1. Beginning with a question (with or without data)
   1. Has the question already been answered?
      1. i.e. decision already made and analysis is there for support of that decision. (my exhaustive DLP analysis)
   2. Is the question worth answering?
      1. e.g. types of spam being filtered
2. Cleaning data
   1. Preparing data for use
      1. This is not taught in school
      2. Most time of “data analysis” is spent here
   2. Importance of normalization
      1. Normalizing something like “industry”
      2. Techniques (and tools)
   3. Storing data for retrieval (excel, csv, json, database, etc)
      1. Highly dependent on use-once or use-many
3. Harnessing the Power of Little Data
   1. Descriptive statistics
      1. Summarizing to a single number
         1. Since we’re talking about visualization though, we want to focus on sharing as much data and relationships without the reduction of information in single numbers. (video of packets shows all the data)
   2. Correlation, probability and margins of error
   3. Communicating uncertainty versus variability
      1. Estimating system uptime as an example
   4. Communicating complexity
4. Visualizing for Analysis
   1. Going to the wood shop, not the Prom
      1. Example: Minitab output for regression analysis and the stories there
      2. Probably many more examples here of insightful and meaningful visualizations that we wouldn’t share with the world.

# Telling the Story

1. All stories need a listener
   1. The components of communication (sender, channel and recipient)
      1. Success is when sender and receiver share the same meaning
   2. Pitfalls
      1. Focusing on the channel
         1. Carefully (or don’t) adjust stories for the medium
      2. Communicating from the sender to the sender
         1. Curse of knowledge
         2. Always create for someone else
      3. One Story will never fit all
         1. Adjusting stories for the recipient
            1. Manager, decision maker versus ops person
2. Decoding the decoding process
   1. Intro topic: This is not a natural skill, excel defaults do not a dataviz make
   2. Cognitive science (or how we know what we see)
      1. Probably quote Few and Cairo
3. Decoding the data for encoding
   1. Categorical versus Quantitative
      1. Port numbers and IP addresses are categorical
   2. Over Time and Space
4. Color me happy
   1. Introduction to Color Theory
5. Creating a message

*Note: might be good to pick one or two interesting data sets (IDS traffic? firewalls?) and repeatedly show examples of the data, also I think we should exclude code from the book, but offer a website where all the code and examples can be downloaded and toyed with?*

* 1. Scatter Plots
     1. For each section, call out the cognitive strengths and weaknesses
     2. Offer code to do this in R and python?
  2. Bar graphs
  3. Line graphs
  4. Pie charts
  5. Distributions
     1. Histogram
     2. Density plots
     3. Box, violin plots
  6. Relationships
     1. Network diagrams (hosts, ports, subnets, DNS)
     2. Other types of complex diagrams

*Note: We may want to split this chapter into two chapters, “simple charts” and “complex charts” or something like that.*

1. Emphasizing the Story
   1. Chart junk, data-ink ratio, least sig difference
   2. Annotations
   3. Subtle highlights, emphasis
      1. Final touches by hand, not programmatically
   4. A final check: does the story in the visualization:
      1. Match the story in the data?
      2. Answer the question we asked?
   5. Now we can talk about the channel
2. Creating an Interactive Story
   1. Multiple relationships means multiple stories
      1. Often one perspective is not enough
   2. Tableau, D3, and about 100 other tools out there
   3. We could write another whole book here, but interactive visualizations are just an extension of the fundamentals we cover in this book.

# Conclusion

1. Back to Storytelling
   1. Good story tellers train and incorporate feedback
      1. Solutions are more from thinking than buying
      2. Data visualization is not a natural skill, it must be learned
      3. Be truthful: message should match the data
   2. Data helps our understanding of our environment
      1. Visualizations help communicate complexity quickly
      2. Simple tools can be, data scientist you need not be