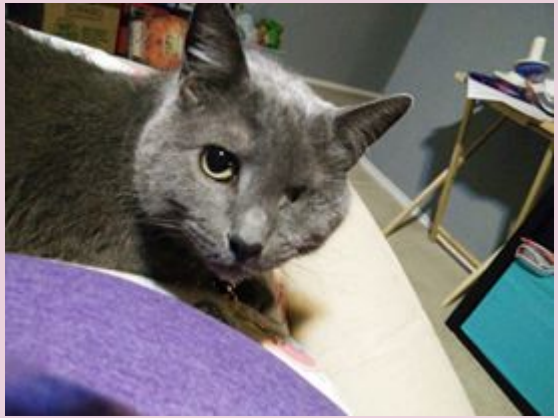
Iron Sights for Your Data

Predictive Analytics for your Blue Team

Leah Figueroa

Who am I?

- Master’s in Education, ABD in research psychology
- 13 years in data analysis
- Data analyst in higher education
- Data aficionado
- Fiber artist (aka knitter)
- Cat-lover
- @Sweet_Grrl
- leahfigueroa22@gmail.com



Where We Are

- Security incidents and data breaches are common
- 2016 DBIR
  - More than 64,000 security incidents
  - More than 3,100 confirmed data breaches
- 2015 DBIR
  - More than 79,000 security incidents
  - More than 2,100 confirmed data breaches
- 2014 DBIR
  - More than 63,000 security incidents
  - More than 1,300 confirmed data breaches

From the Verizon Data Breach Investigation Reports*

It Can’t Possibly Be That Bad, Can It?

2016 Year in Review

- January – Xoom, Chick-fil-a and OneStopParking (payment card breach) Adobe Flash zero-day
- February – BCBS - Anthem, Deep Panda, Dyre, Vawtrak, Carbanak, Ramnit botnet returns
- March – BCBC - Premera, Mandarin Hotel Group, POS NEXTEP
- April – Great Cannon - GitHub, GreatFire, Pawn Storm, CozyDuke
- May – InterContinental Hotel Group, Partners HealthCare, CareFirst BCBS, MetroHealth, Bellvue Hospital
- June – OPM, Cisco
- July – Harvard, Penn State, Trump Hotels, UCLA, Ashley Madison, Hacking Team
- August – American Airlines, DOD, DHHS, IRS, Carphone Warehouse (combined attack), Ubiquity, Malvertising
- September – Excellus - BCBS, Blue Termite
- October – Experion - T-Mobile, TalkTalk, Pawn Storm, Dridex
- November – Dridex, The Armada, Farmer's Direct
- December – Australian Bureau of Meteorology, Screen OS, BlackEnergy

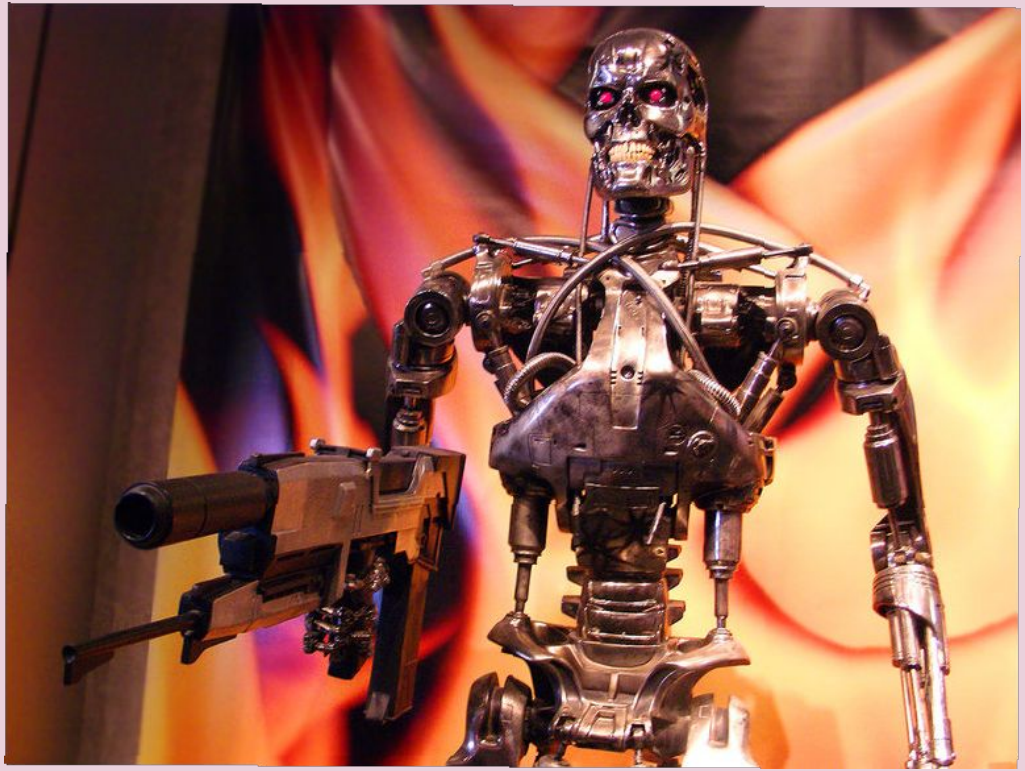
It’s Bad

So what the hell do we do?

Run around and scream

OR

Weaponize Your Data



Underlying Philosophy

Past behavior often predicts future behavior.

In other words, people often repeat the same behaviors.

Underlying Philosophy

What the heck does that psycho-mumbo-jumbo have to do with security?

Past behaviors (data breaches, attacks, threats, vulnerabilities, etc.) often predict future behaviors.

Attacks (and attackers and attack vectors) typically follow patterns you can recognize.

So How do I Weaponize my Data?

- Before you can weaponize your data, you have to understand how to approach it
- Two pronged approach
  - Data Mining
  - Predictive Analytics

Why?

- Why would we bother mining our data?
- Why would we use predictive analytics?
- Big Brother does it…you should, too

More Seriously, Why Bother?

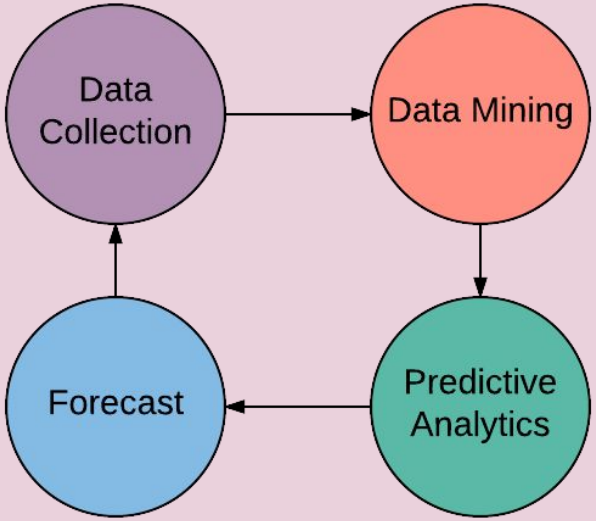
- Forecasts can predict attack surfaces, vectors, actors
- Too many breach reports, not enough time
- Too many ice giants, not enough Nordic Gods
- While not perfect, using data can put you in a much more defensible position
- This is an ongoing process that requires a team effort
- In the end, though, this process means you can fight back effectively

But I Have One of Those Awesome Tools?!?!

- I have _____ tool for working with data.
- You can (and should!) use them, BUT only after you examine and understand your data.
- Otherwise, you are using a firehose to fill a shot glass.

Properly Weaponized Data

- Properly weaponized data provides a feedback loop



- Feedback loop ensures that the forecasts become more accurate

Frameworks

- A good framework for data collection should include:
  - Incident Tracking
  - Victim Demographics
  - Incident Description
  - Discovery & Response
  - Impact Assessment

Frameworks

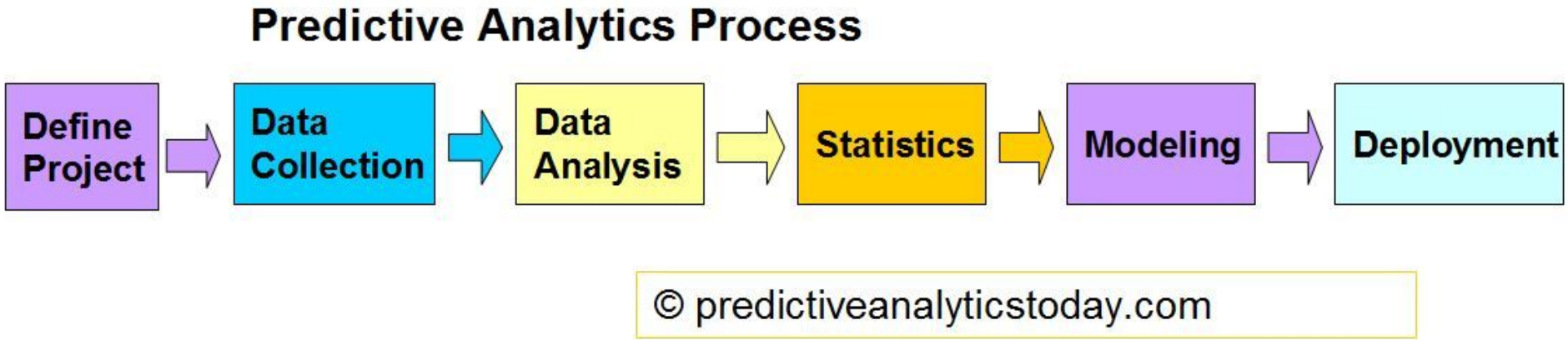
- Verizon VERIS (Vocabulary for Event Recording and Incident Sharing)
  - VERIS has a nice schema (!) for making sure your framework can accommodate the data
  - Ask your DBA for help (If they can’t figure it out, get a new DBA. No, seriously, get a new one)
- Less Verizony ones
  - CERT Insider Threat Database
  - Factor Analysis of Information Risk (FAIR)
  - National Institute of Standards and Technology’s Risk Management Framework (NIST RMF)
  - Operationally Critical Threat, Asset and Vulnerability Evaluation (OCTAVE)
  - SANS 20 Critical Security Control
  - Threat Agent Risk Assessment (TARA)
  - Others

Data Collection

- Incident reports provide data for trend analysis
- System logs provide data for for broad analysis
- Application logs provide data for focused analysis
- All data collected provides data for analysis
  - The more you get, the better it gets
- Many tools currently exist to help you extract this data (and if you're lucky, help with pre-processing)
  - Logstash (free, flexible)
  - Splunk (costly, powerful)
  - Security Onion (kitchen sink)
- When in doubt, ask Operations!

Data Mining and Predictive Analytics

- Data Mining and Predictive Analytics are sometimes used interchangeably
- Data Mining produces decisions based on normal reports
- Predictive Analytics uses Data Mining and builds upon it



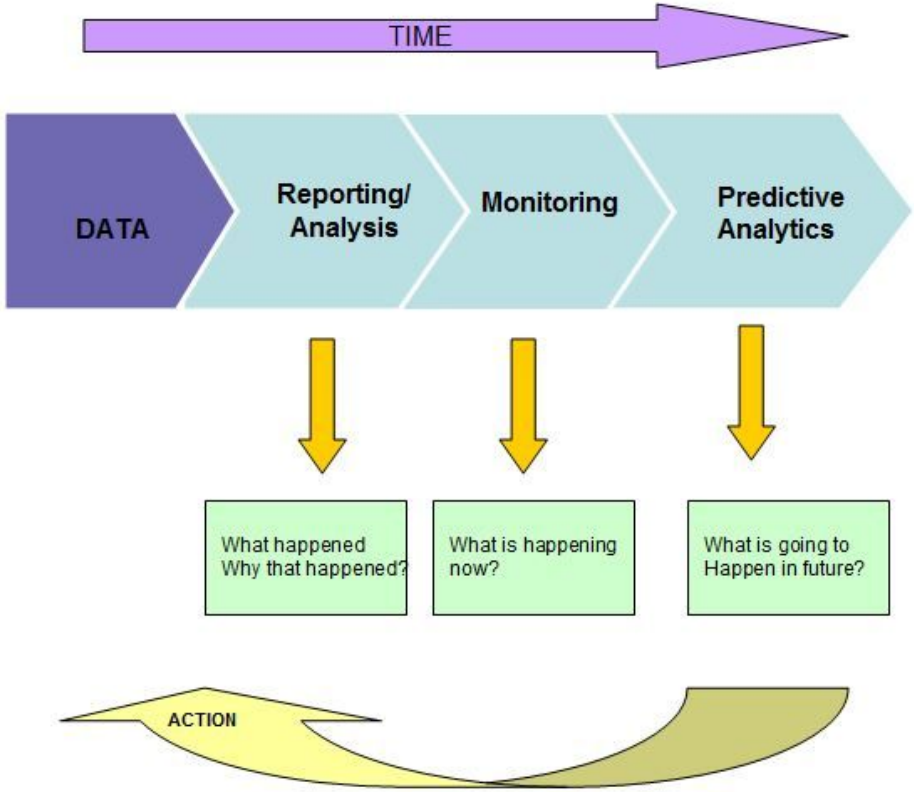
Data Mining

- Data Mining finds valuable information that is hidden in large volumes of data
- Consists of five major elements
  - Extract, transform, and load (ETL) transaction data
  - Store and manage data
  - Provide data access
  - Analyze data
  - Present data
- Different levels of analysis are available for use in data mining

Predictive Analytics

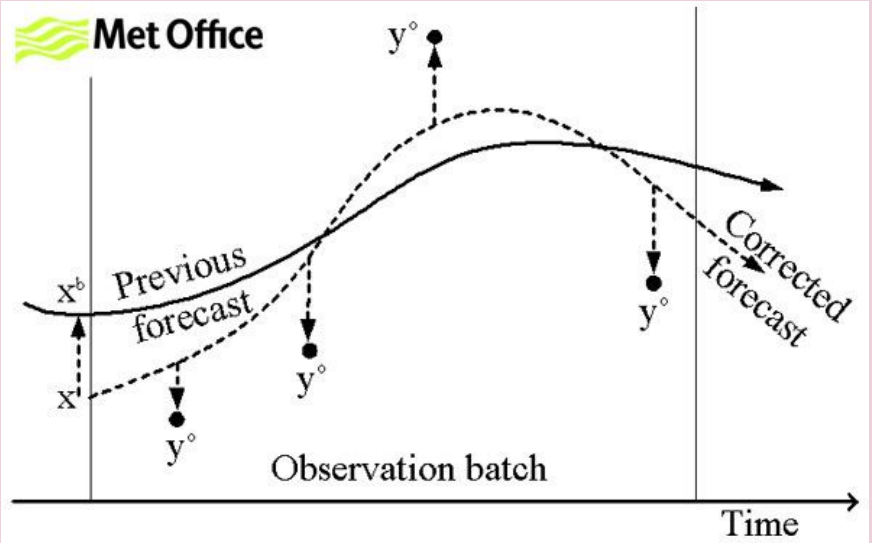
- Predictive analytics extracts data from existing data sets to identify trends and patterns, which are used to predict future outcomes and trends.
- Predictive analytics allows the user
  - Identify trends and patterns.
  - Improve performance.
  - Drive strategic decision making.
  - Predict future outcomes/trends/behaviors
- Predictive analytics is NOT an absolute science

Predictive Analytics



Forecast

- Predictive analytics allows models to be created.
- Models allow forecasting:
  - What will happen next?
  - What trends are expected to continue?
  - How we can continue to improve?
- This provides mineable data for future forecasts



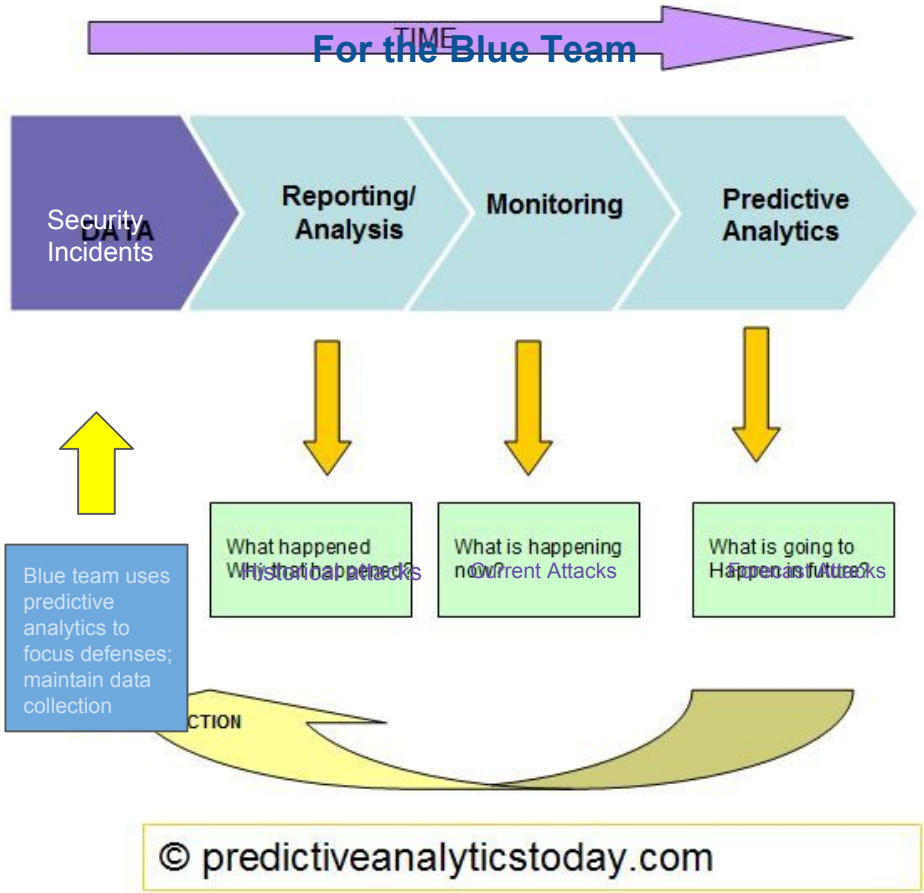
Before you begin

- Select a tool
- Learn to use the tool effectively
- Use it in the feedback loop

Data Mining / Predictive Analytics

- The best approach combines both Data Mining and Predictive Analytics tools with a big friendly GUI (otherwise, it just gets messy)
- Top tools in use
  - Enterprise tools
    - IBM Predictive Analytics
    - SAS Predictive Analytics
  - Open source projects
    - R (https://www.r-project.org/)(Free)
    - WEKA (http://www.cs.waikato.ac.nz/ml/weka/) (Free)
    - KNIME (https://www.knime.org/)
    - RapidMiner (https://rapidminer.com/)

Predictive Analytics



Front End Tools

- Remember your front end tools earlier?
  - Kibana
  - Splunk
  - Maltego
  - Palantir
  - Graphite
  - Greylog2
- With your data focused they can now work efficiently

- Data is everywhere
- Data is useful
- Data is POWER!

Questions?

- @Sweet_Grrl
- sweetgrrl1222@protonmail.com
- leahfigueroa22@gmail.com