

splunk> .conf2017

Building Blocks For Analytics Common Sense

Yanpei Chen | Sr. Product Manager, Product Analytics

Archana Ganapathi | Director, Data Strategy

September 2017 | Washington, DC

PRESENTATIONS.

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2017 Splunk Inc. All rights reserved.

Common Pitfall



QUICK
WINS!!!

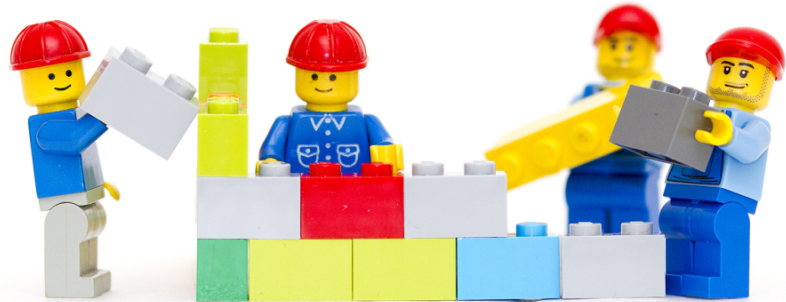


```
130.60.4 - [07/Jan 18:10:57:153] "GET /category.screen?category_id=Q1FT5uJ  
128.241.230.82 - [07/Jan 18:10:57:123] "GET /product.screen?product_id=1  
317.27.160.0 - [07/Jan 18:10:57:153] "GET /category.screen?category_id=Q1FT5uJ  
itemid=EST-10&product_id=RP-LI-02" 468 125.17 /butte  
action=shopping.com/RL-02" 468 125.17 /butte  
action=shopping.com/RL-02" 468 125.17 /butte
```

```
130.60.4 - [07/Jan 18:10:57:153] "GET /category.screen?category_id=Q1FT5uJ  
128.241.230.82 - [07/Jan 18:10:57:123] "GET /product.screen?product_id=1  
317.27.160.0 - [07/Jan 18:10:57:153] "GET /category.screen?category_id=Q1FT5uJ  
itemid=EST-10&product_id=RP-LI-02" 468 125.17 /butte  
action=shopping.com/RL-02" 468 125.17 /butte  
action=shopping.com/RL-02" 468 125.17 /butte
```


Building Blocks For Data-driven Decision Making

- ▶ Clarity of purpose and context
- ▶ Data worth a damn
- ▶ Numerical discipline
- ▶ An appropriate calculator
- ▶ Sanity checked results
- ▶ Strong infrastructure - technical and non-technical



Clarity Of Purpose And Context

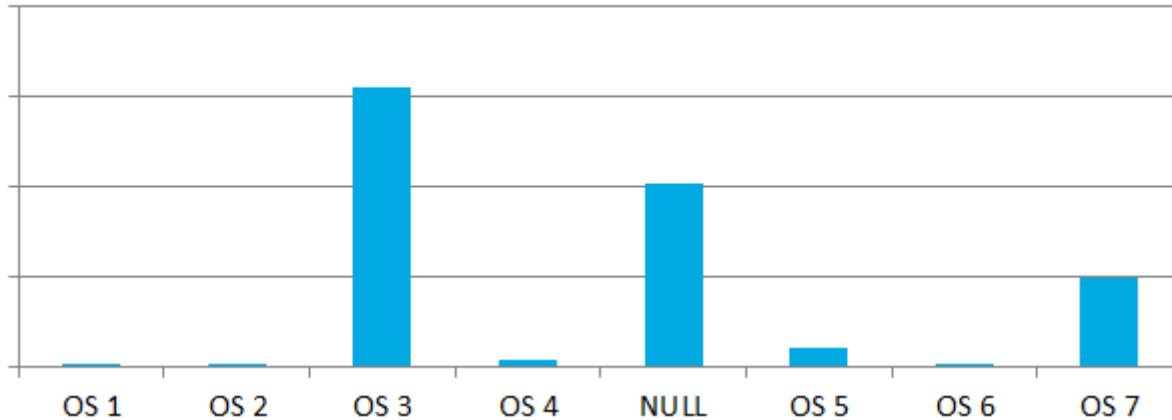
Data Worth A Damn

Bad data means garbage in, garbage out

Example 1: Nulls

- ▶ Real NULLs or missing values?
- ▶ Presence of NULLs relevant or invalidates decision at hand?

Sample customer tickets



```

130.60.4 - - [07/Jan 18:10:57:153] "GET /category/screen?category_id=GLF75&SESSIONID=SD55LAF10ADF10 HTTP/1.1" 404 720 "http://buttercup-shopping.com/cart.do?action=vtw&itemId=EST-6&product_id=3-3w&os=
128.241.230.82 - - [07/Jan 18:10:57:123] "GET /product/screen?product_id=FL-DSH-01&SESSIONID=SD55L7F6ADF9 HTTP/1.1" 404 332 "http://buttercup-shopping.com/category/screen?category_id=GLF75&SESSIONID=SD55LAF10ADF10 HTTP/1.1" 404 720 "http://buttercup-shopping.com/cart.do?action=vtw&itemId=EST-6&product_id=3-3w&os=
317.27.160.0 - - [07/Jan 18:10:57:123] "GET /product/screen?product_id=FL-DSH-01&SESSIONID=SD55L7F6ADF9 HTTP/1.1" 404 332 "http://buttercup-shopping.com/category/screen?category_id=GLF75&SESSIONID=SD55LAF10ADF10 HTTP/1.1" 404 720 "http://buttercup-shopping.com/cart.do?action=vtw&itemId=EST-6&product_id=3-3w&os=
item_id=EST-10&product_id=RP-LI-02" 468 125.17 "http://buttercup-shopping.com/cart.do?action=vtw&itemId=EST-6&product_id=3-3w&os=
/buttercup-shopping.com/rp-li-02" 468 125.17 "http://buttercup-shopping.com/cart.do?action=vtw&itemId=EST-6&product_id=3-3w&os=
shopping.com/purchase&it
/butte

```


Check List For Data Quality

▶ Coverage

- How much data? What time frame?
- Who/what is covered? How much of “total universe” is visible?

▶ Granularity

- Frequency (hourly, weekly)?
- Level of detail (entire system, per component, per session/user)?

▶ Semantics

- Direct vs proxy instrumentation?
- Reference vs divergent interpretations?

▶ Dedicated talk on data quality - see schedule!

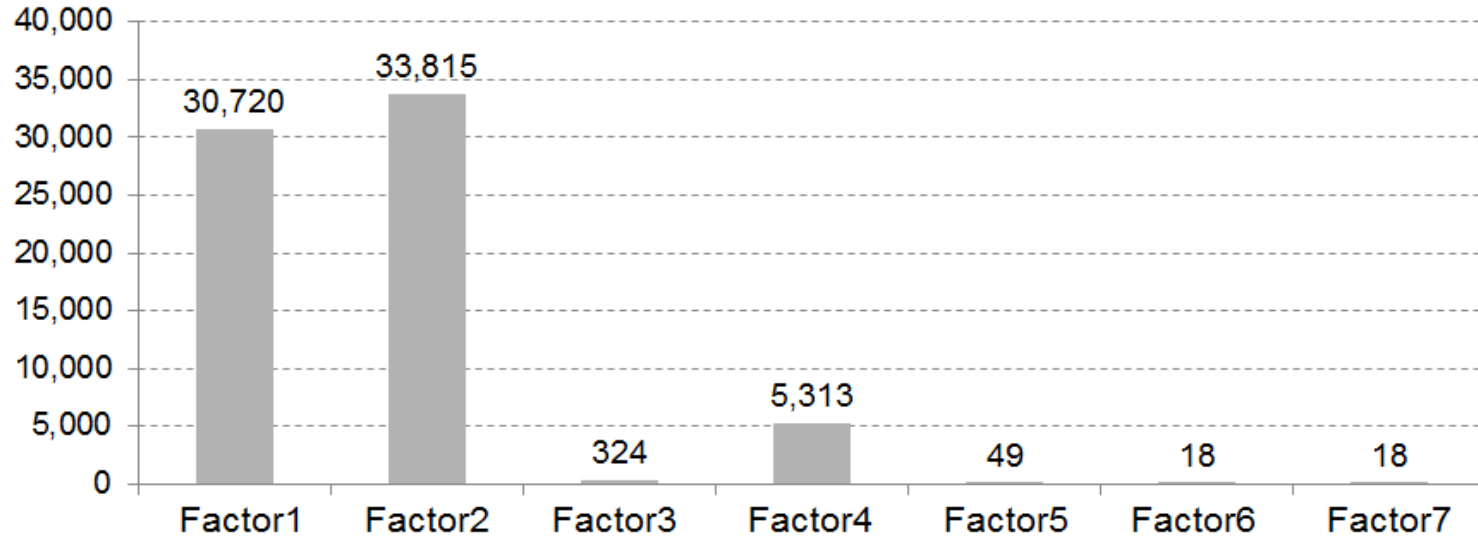
Numerical Discipline

Non-scary techniques that help you be rigorous with data

Count Interesting Things By Interesting Factors

This often reveals valuable info, and everyone understands it!

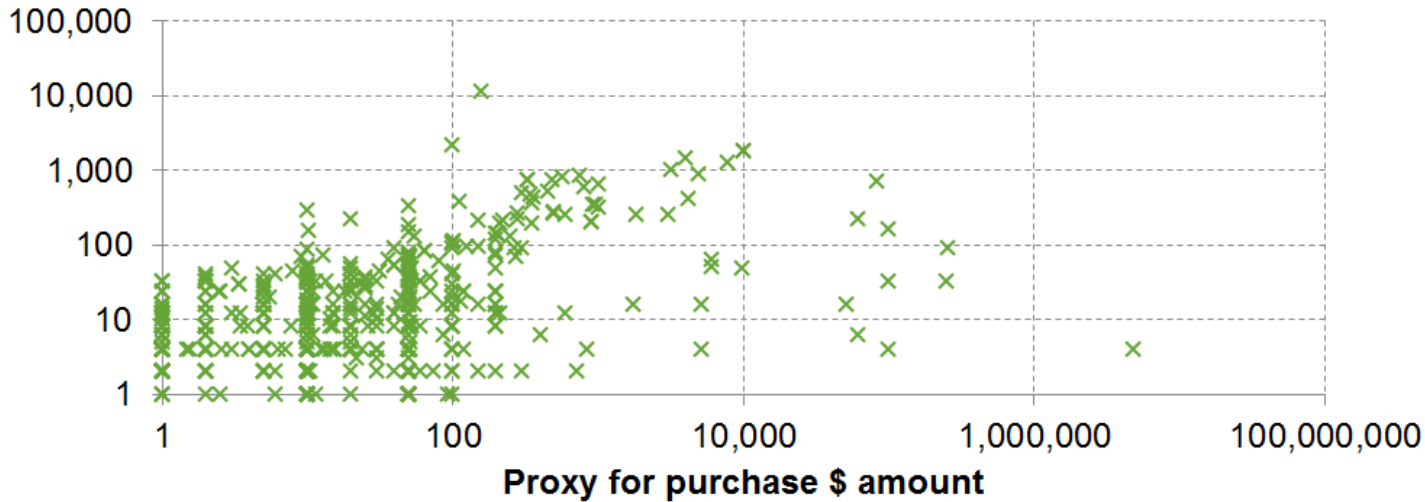
Customer activity



Identify Outliers And Make Sense Of Them

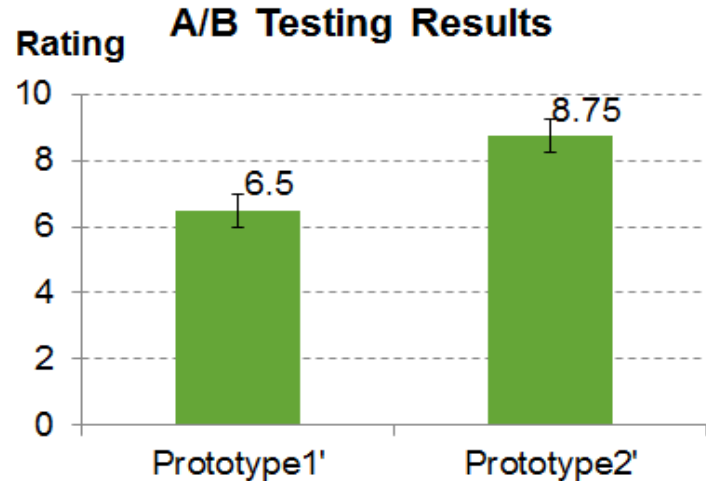
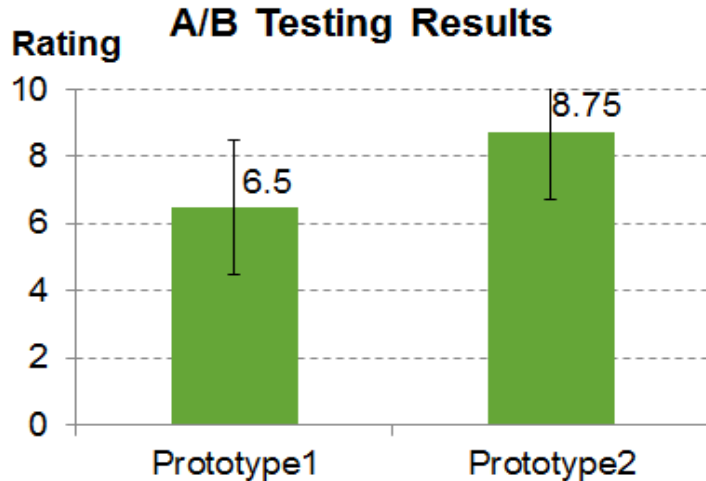
Outliers are as important to understand as “the average”

Related factor



Calculate The Margin Of Error In Results

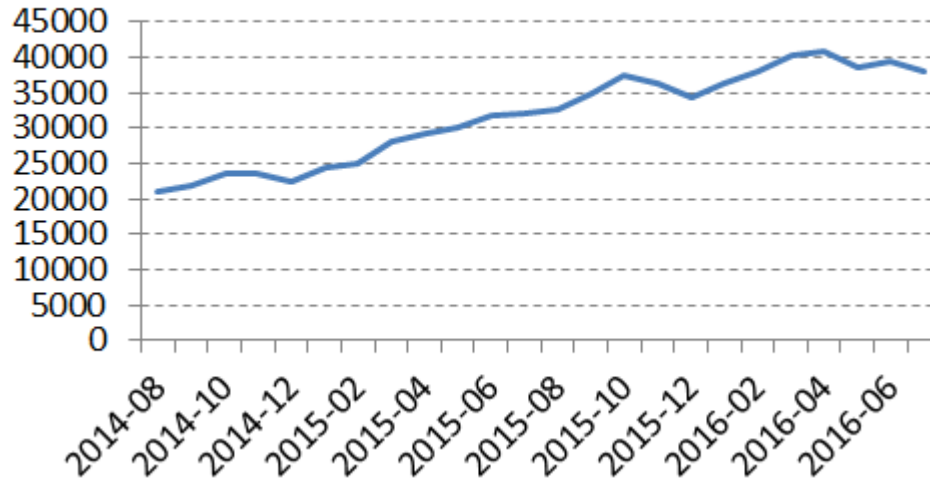
- ▶ Graphs have same numerical results. Left graph has overlapping margin of error
- ▶ Ignoring margin of error could have led to wrong investment decisions



Check For Trend vs Noise vs Seasonality

Same data, 24 months to check for seasonality

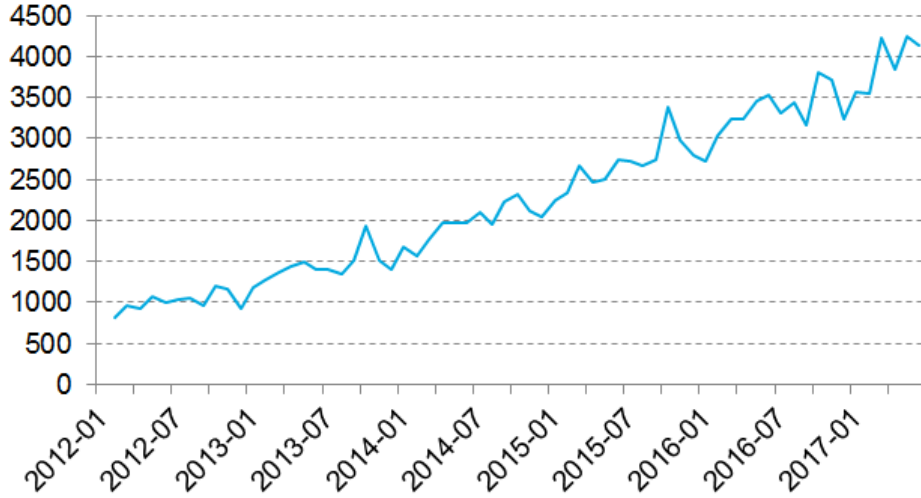
Proxy for size of customer base



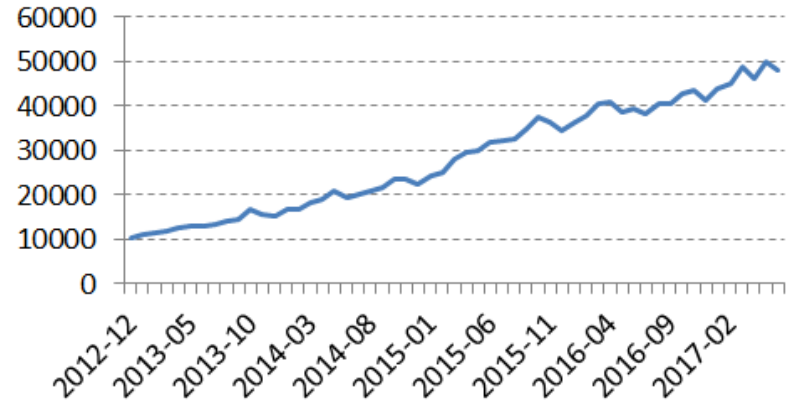
Check If Any Trends Need To Be Normalized

But customer base is also growing ...

Proxy for customer issues



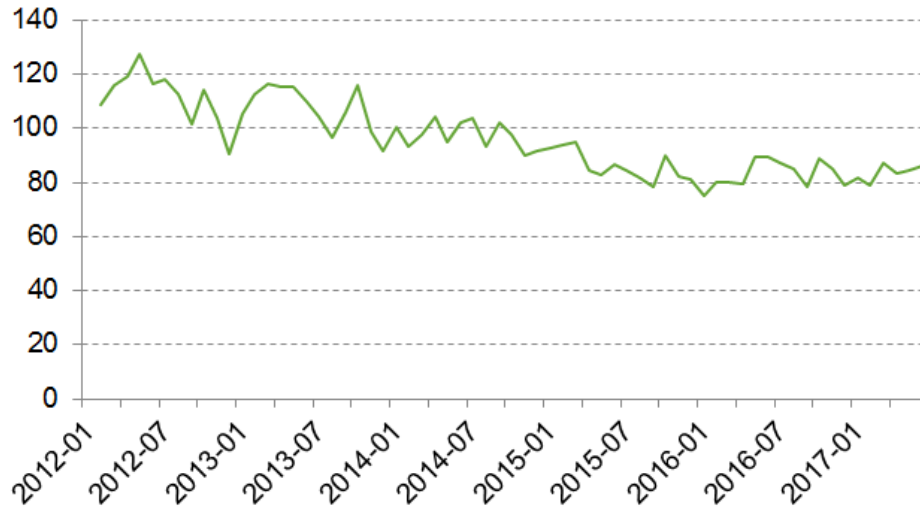
Proxy for size of customer base



Check If Any Trends Need To Be Normalized

Let's normalize customer issues by size of customer base
Normalized data hugely informative

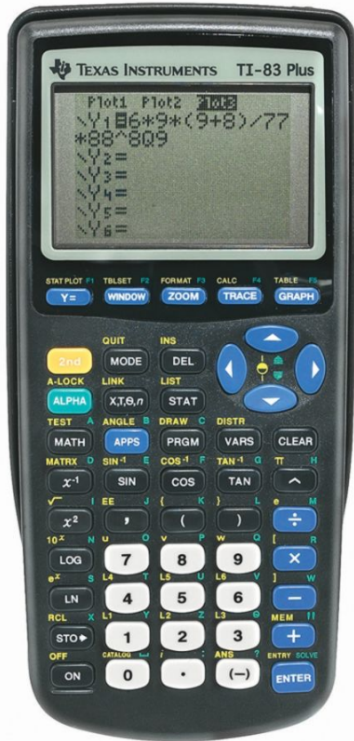
Customer issues per size of customer base



An Appropriate Calculator

Knowing what technique to use when
and how far to go

“Everything should be made as simple as possible, but not simpler.”



- ▶ How many buttons do you need to
 - Compute the dinner bill?
 - Do rocket science?

- ▶ What % of the world's problems are similar to
 - Computing the dinner bill?
 - Doing rocket science?

Splunk Is Actually A Pretty Powerful Calculator!

Save & share results



Visualize

Explore

Analyze
(SPL & MLTK)

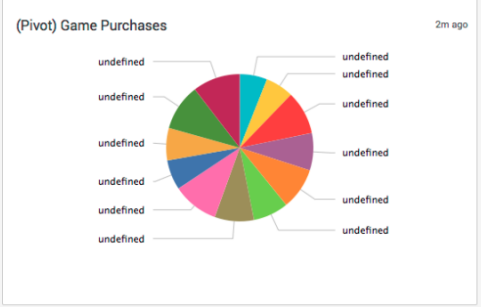
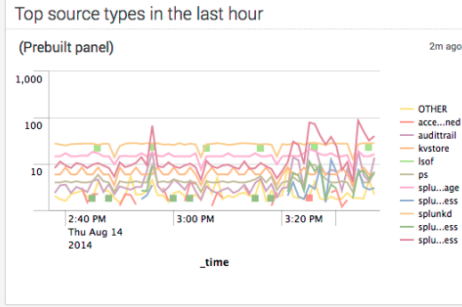
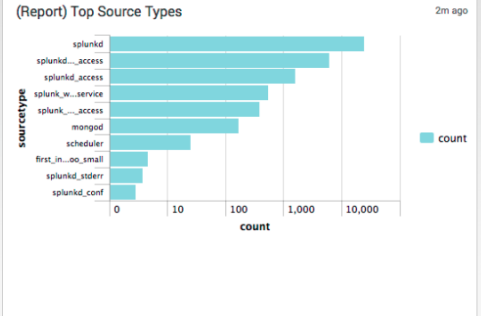
Ingest

Searches power dashboards
Show the various searches to power a panel.

(Inline Search) Top Source Types 2m ago

sourceType	count	percent
1 splunkd	24775	73.43
2 splunkd_ui_access	6207	18.40
3 splunkd_access	1612	4.78
4 splunk_web_service	553	1.64
5 splunk_web_access	388	1.15
6 mongod	169	0.50
7 scheduler	25	0.07
8 first_install-too_small	4	0.01
9 splunkd_stderr	3	0.01
10 splunkd_conf	2	0.01

< prev 1 2 next >



Splunk Is Actually A Pretty Powerful Calculator!

Save & share results



Visualize

Analyze
(SPL & MLTK)

Explore

Ingest

Searches power dashboards
Show the various searches to power a panel.

Security Posture

ACCESS NOTABLES Total Count 1k ↑ +375	ENDPOINT NOTABLES Total Count 753 ↑ +532	NETWORK NOTABLES Total Count 109 ↑ +57	IDENTITY NOTABLES Total Count 7 ↓ -13	AUDIT NOTABLES Total Count 37 ↑ +17	THREAT NOTABLES Total Count 2k ↑ +1k	UEBA NOTABLES Total Count 0
--	---	---	--	--	---	---

Notable Events By Urgency

Notable Events Over Time

Top Notable Events

rule_name	sparkline	count	src	sparkline	correlation_search_count	security_domain_count	count
Watchlisted Event Observed		1459	10.11.36.20		6	4	32
Excessive Failed Logins		559	10.11.36.5		3	2	18
Host With A Recurring Malware Infection		305	10.11.36.13		3	2	17
Threat Activity Detected		286	10.11.36.26		3	2	17
Brute Force Access Behavior Detected		282	10.11.36.10		3	2	16
Host With Multiple Infections		244	10.11.36.11		3	2	16
High Or Critical Priority Host With Malware Detected		140	10.11.36.16		3	2	16
Default Account Activity Detected		128	10.11.36.18		3	2	16
Insecure Or Cleared Authentication Detected		69	10.11.36.19		3	2	16
Host Sending Excessive Email		28	10.11.36.21		3	2	16

Top Notable Event Sources

No investigation is currently loaded. Please create (+) or load an existing one (#).

Alternative: Assembling Way-Too-Many Parts

Save & share results

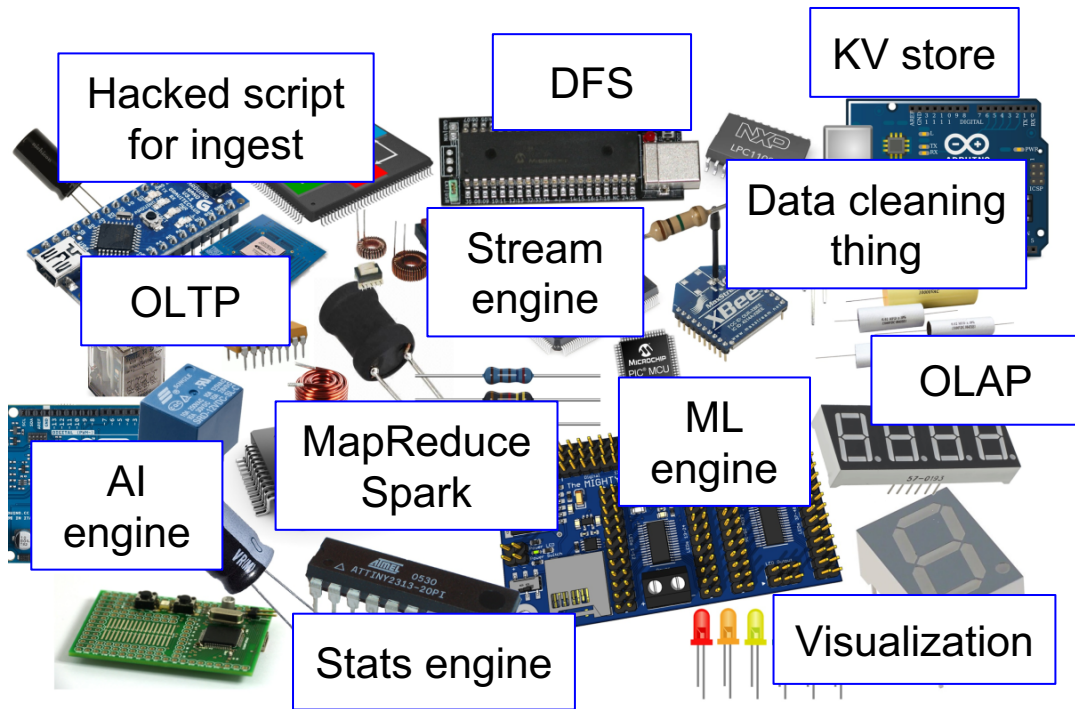
Visualize



Analyze
(SPL & MLTK)

Ingest

Explore



Hacked script for ingest

DFS

KV store

Data cleaning thing

Stream engine

OLTP

OLAP

MapReduce Spark

ML engine

AI engine

Stats engine

Visualization

2x Parts if you also use Cloud

Easy To Get Answers, Easy To Go Wrong



For problems that are rocket science:

- ▶ Did you punch in the right numbers?
- ▶ Do you actually understand the numbers?
- ▶ Did you press the right buttons?
- ▶ Do you actually understand the buttons?
- ▶ Can you explain what you did to stakeholders?
- ▶ Will an astronaut place their life in your hands?
- ▶ What checks did you make on the data?

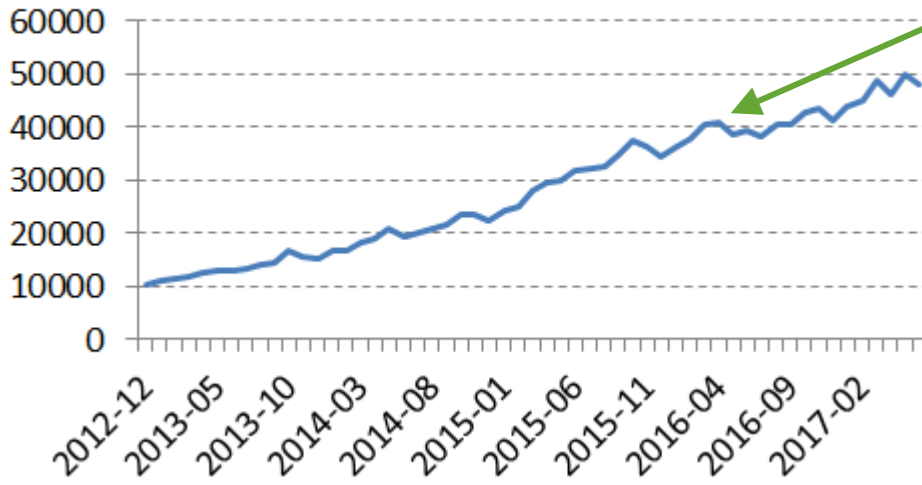
Sanity Checking Results



Confirmation Bias

- ▶ Data often gives a signal to confirm your gut-feel
- ▶ Need to check whether other signals are stronger

Proxy for size of customer base



New product released

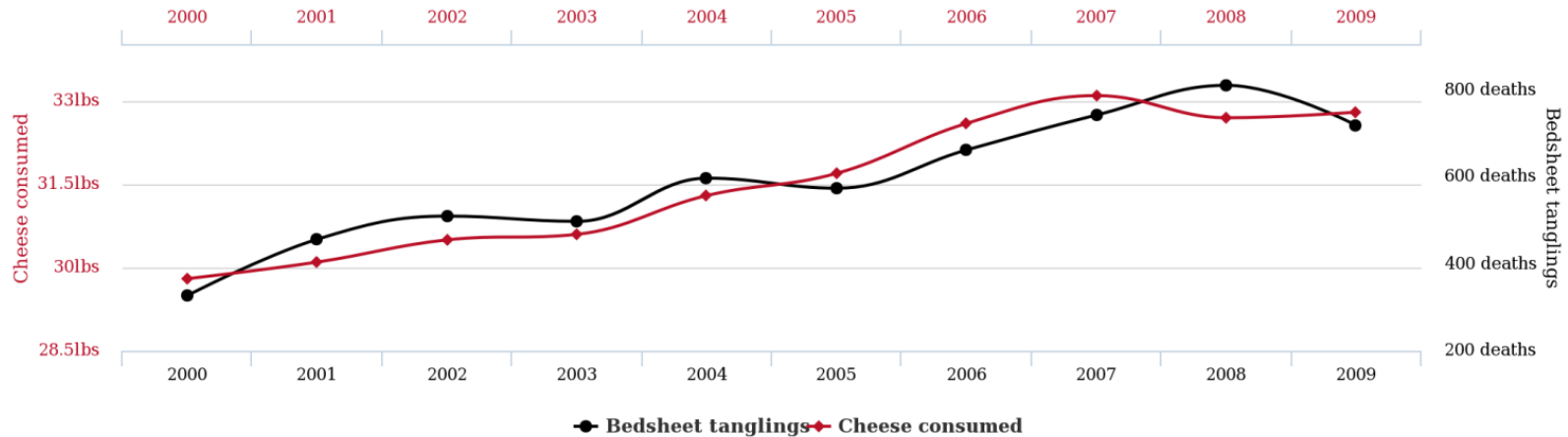
Customer fluctuation
within historical norms

Confusing Correlation With Causation

Per capita cheese consumption

correlates with

Number of people who died by becoming tangled in their bedsheets



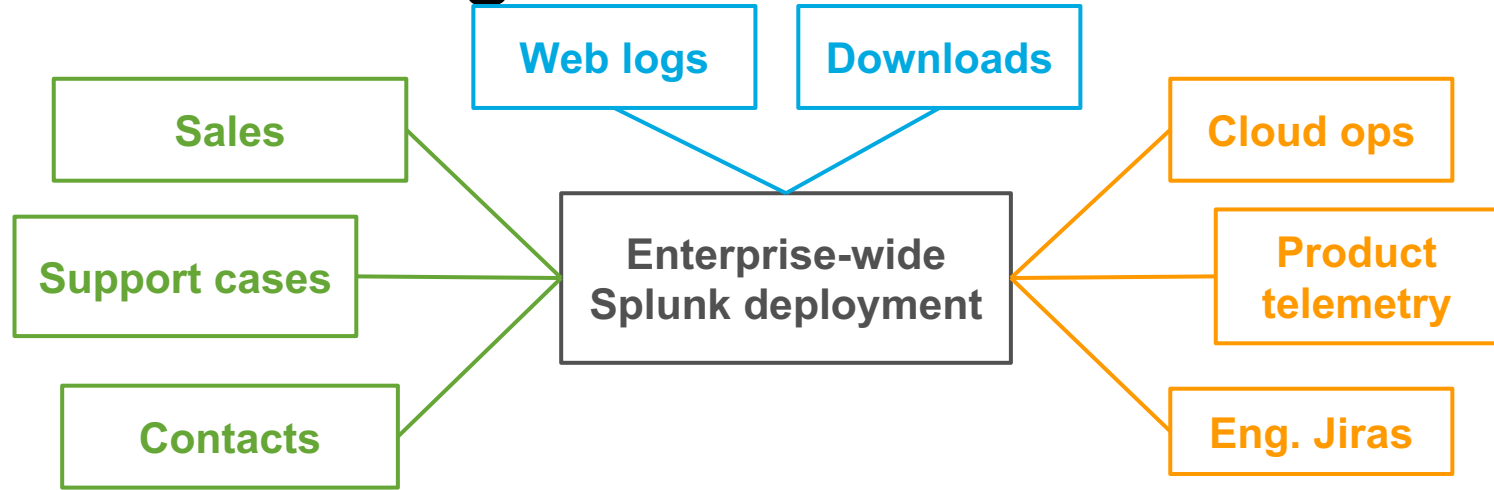
tylervigen.com

Also, umbrellas cause rain ☺

Infrastructure

Technical, cultural, organizational

Translation: Technical, Cultural, Organizational Infra



- ▶ Technical infra: Splunk must scale, especially for cloud ops and product telemetry
- ▶ Cultural infra: To understand ever changing market, learn and adapt continuously
- ▶ Organizational infra: Collaborate across org silos to create much higher value

Closing Thoughts

You Will Be “Data Champions”

- ▶ Your greatest assets: perception of **scientific objectivity and neutrality**
- ▶ You will play many roles as leaders, individual contributors, and/or influencers
 - Translate between data semantics and business semantics
 - Clarify limits and decision boundaries with existing data
 - Nurture data literacy, advocate for long-term investment
 - Create stop gap “data plumbing” while “things improve”
 - Nurture and exemplify openness and transparency
- ▶ You now have more tools to be better data champions
- ▶ PS: See you again for talk on Data Quality!



Thank You

Don't forget to **rate this session** in the
.conf2017 mobile app

splunk> .conf2017