Machine Learning

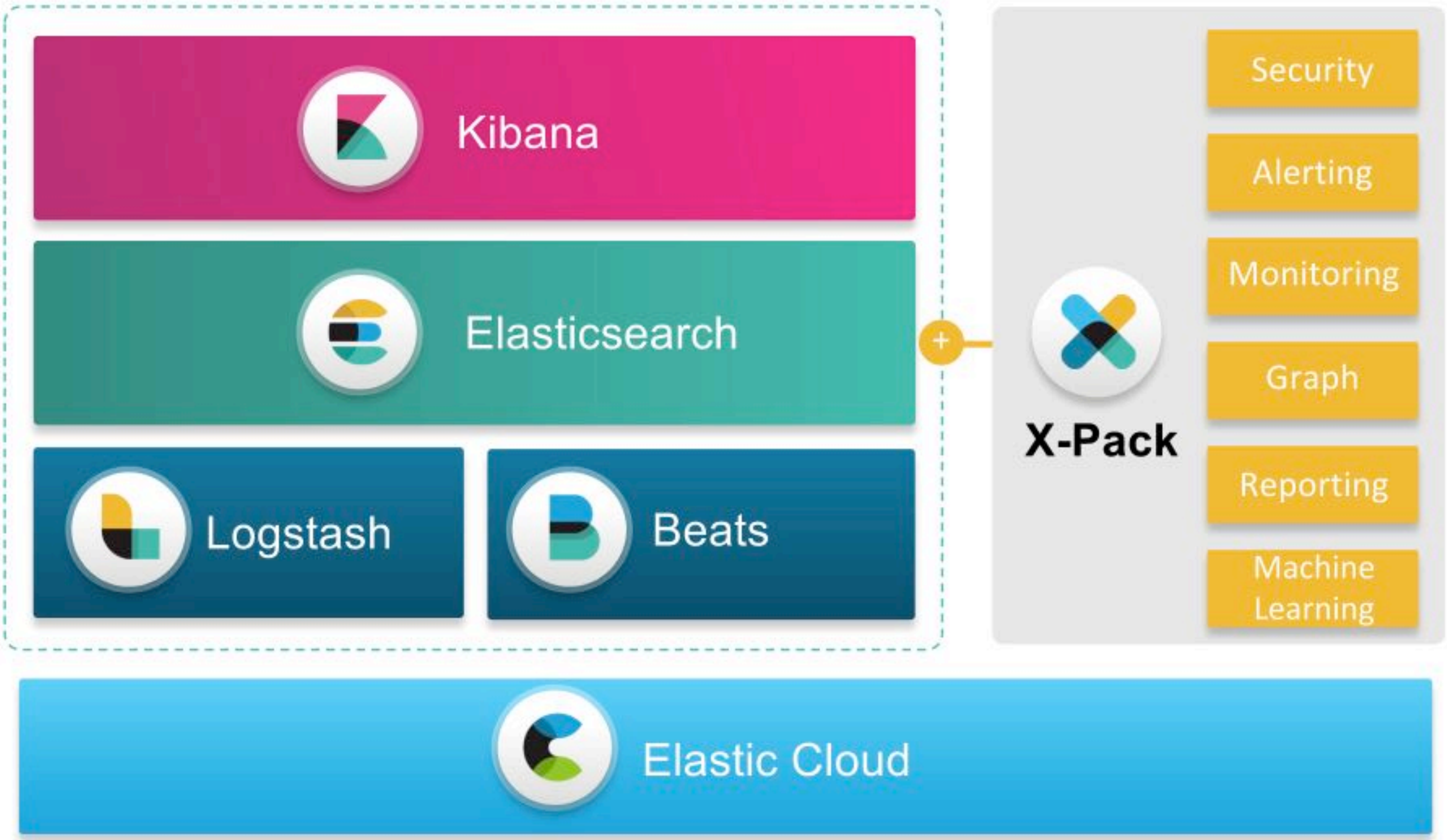
ohne Hype

Philipp Krenn @xeraa

elastic

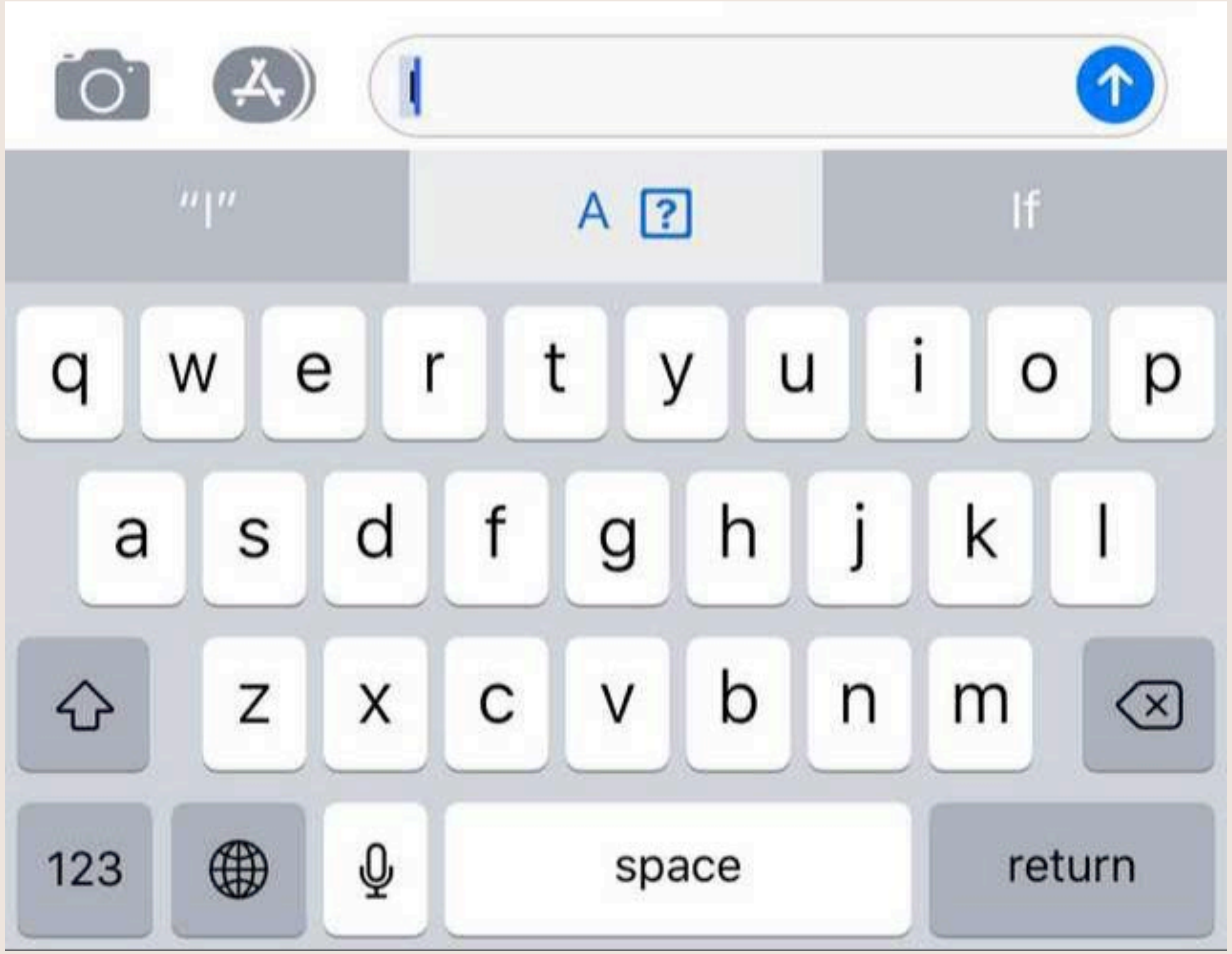
elastic

Developer 🥑



Kibana
Elasticsearch
Logstash
Beats
X-Pack
Security
Alerting
Monitoring
Graph
Reporting
Machine Learning
Elastic Cloud

Machine Learning is going viral…



“I”
A ?
If
q w e r t y u i o p
a s d f g h j k l
z x c v b n m
123
space
return

Cabel Sasser
@cabel

Follow

My favorite story about the “I” Unicode iOS bug is that it spread like a virus. You weren’t affected until someone sent it to you! Then, machine learning thought it was important, and added it to YOUR autocorrect.

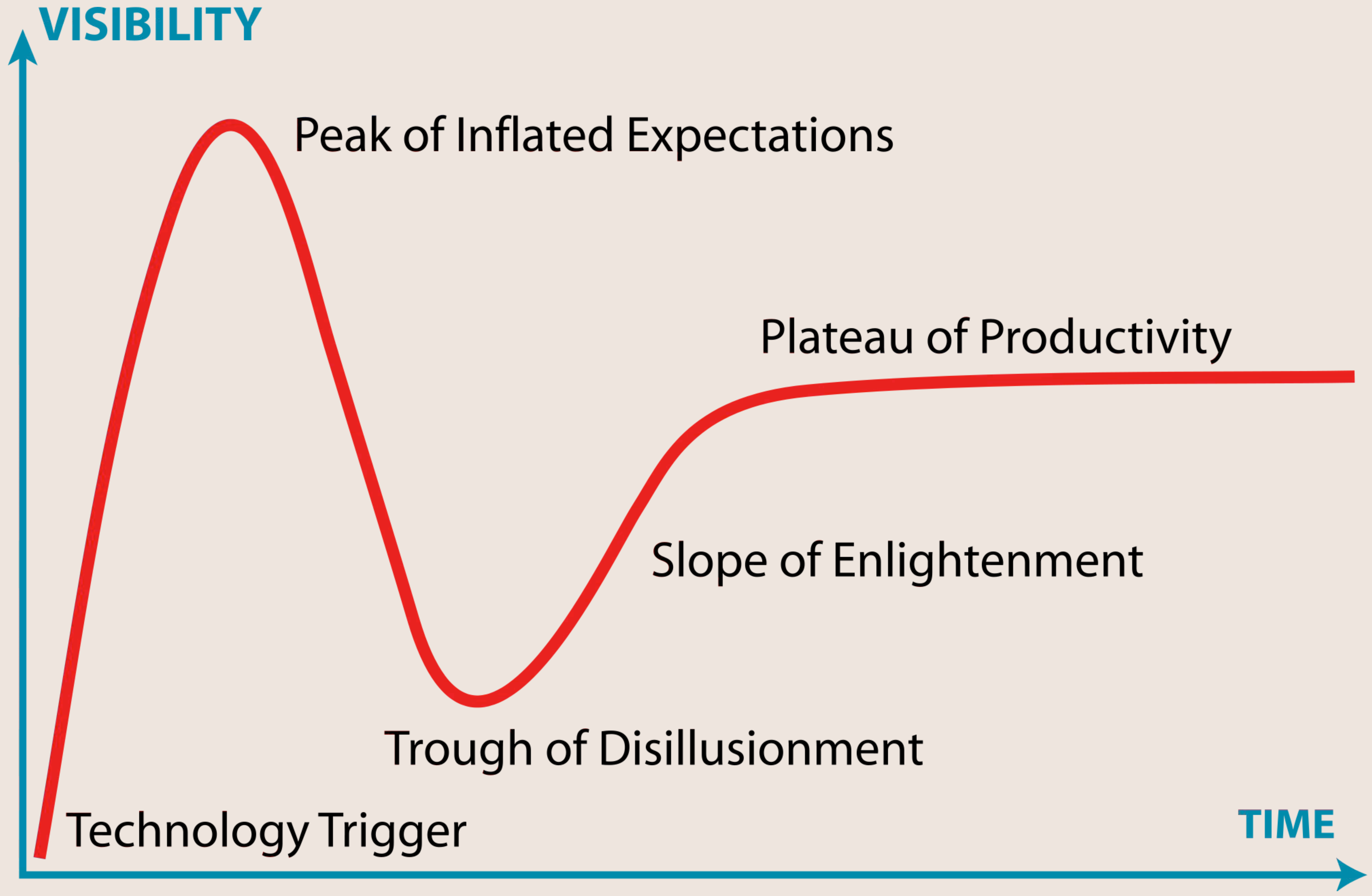
10:56 PM - 9 Nov 2017

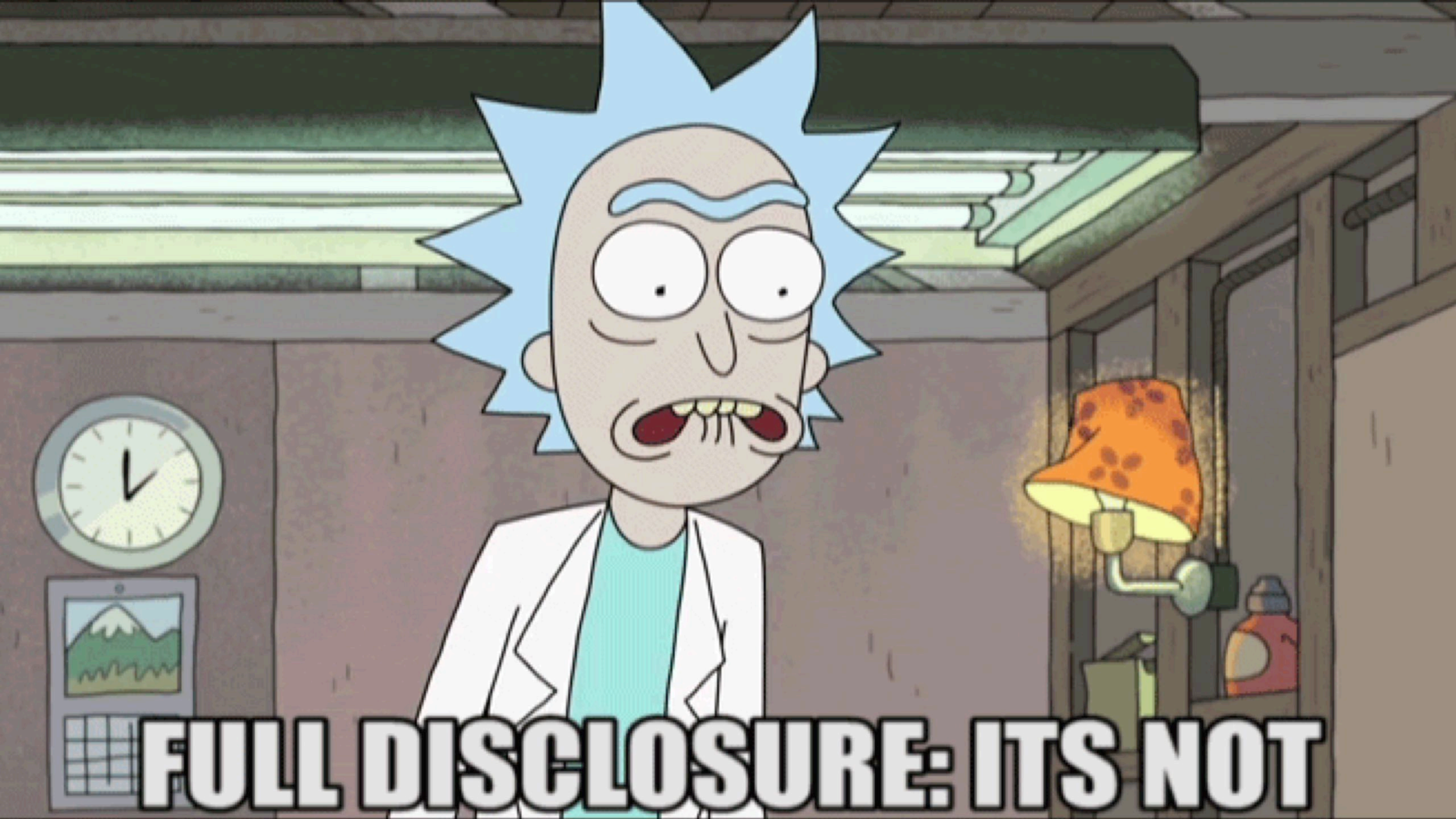
"Using #DeepLearning when all you needed was a few if statements. #MachineLearning #DataScience"

—https://twitter.com/randal_olson/status/927157485240311808

elastic


CAT


VISIBILITY
Peak of Inflated Expectations
Plateau of Productivity
Slope of Enlightenment
Trough of Disillusionment
Technology Trigger
TIME


FULL DISCLOSURE: ITS NOT

Agenda

Machine Learning

Domain

Dataset

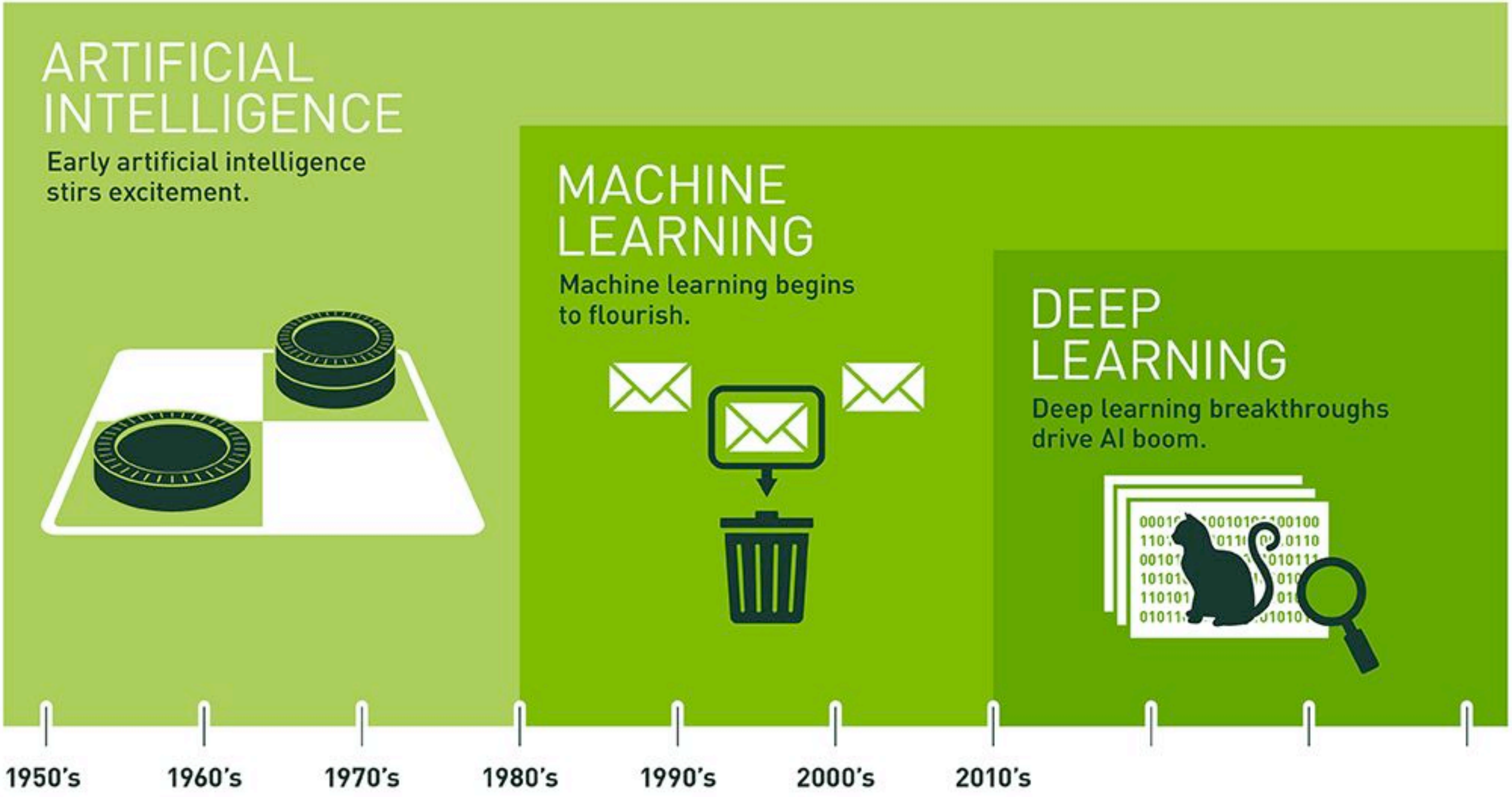
Machine Learning

Artificial Intelligence

Machine Learning

Deep Learning

🤔



https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

General AI

Human characteristics

AI Winter

elastic

Narrow AI

Specific tasks

Facebook

alt="Image may contain: ocean, sky, bridge, cloud, outdoor, water and nature" zalando

NEWS

9 MAR 2018

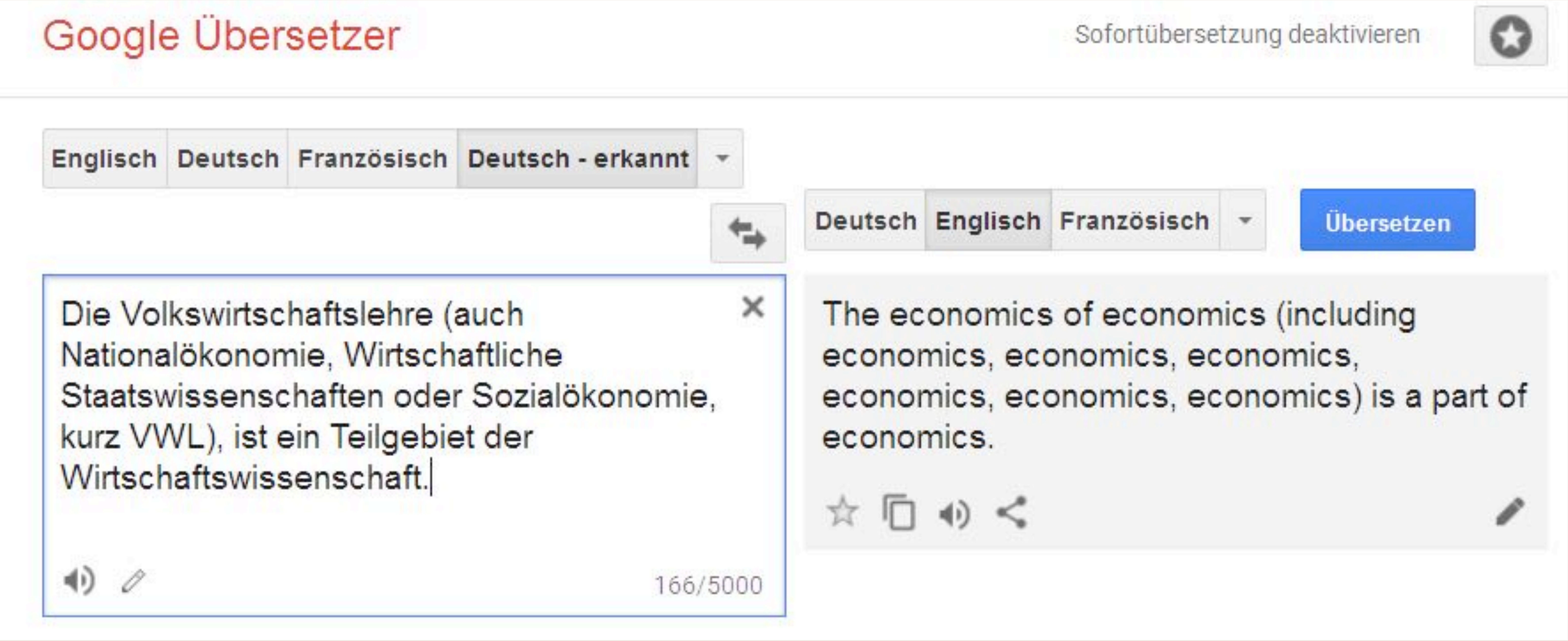
Zalando cuts 250 marketing jobs

Caroline Baldwin Editor, Essential Retail

Email Caroline | Follow @cl_baldwin | Connect on LinkedIn

Zalando has made dramatic job cuts in its marketing and advertising department, Essential Retail has learnt.

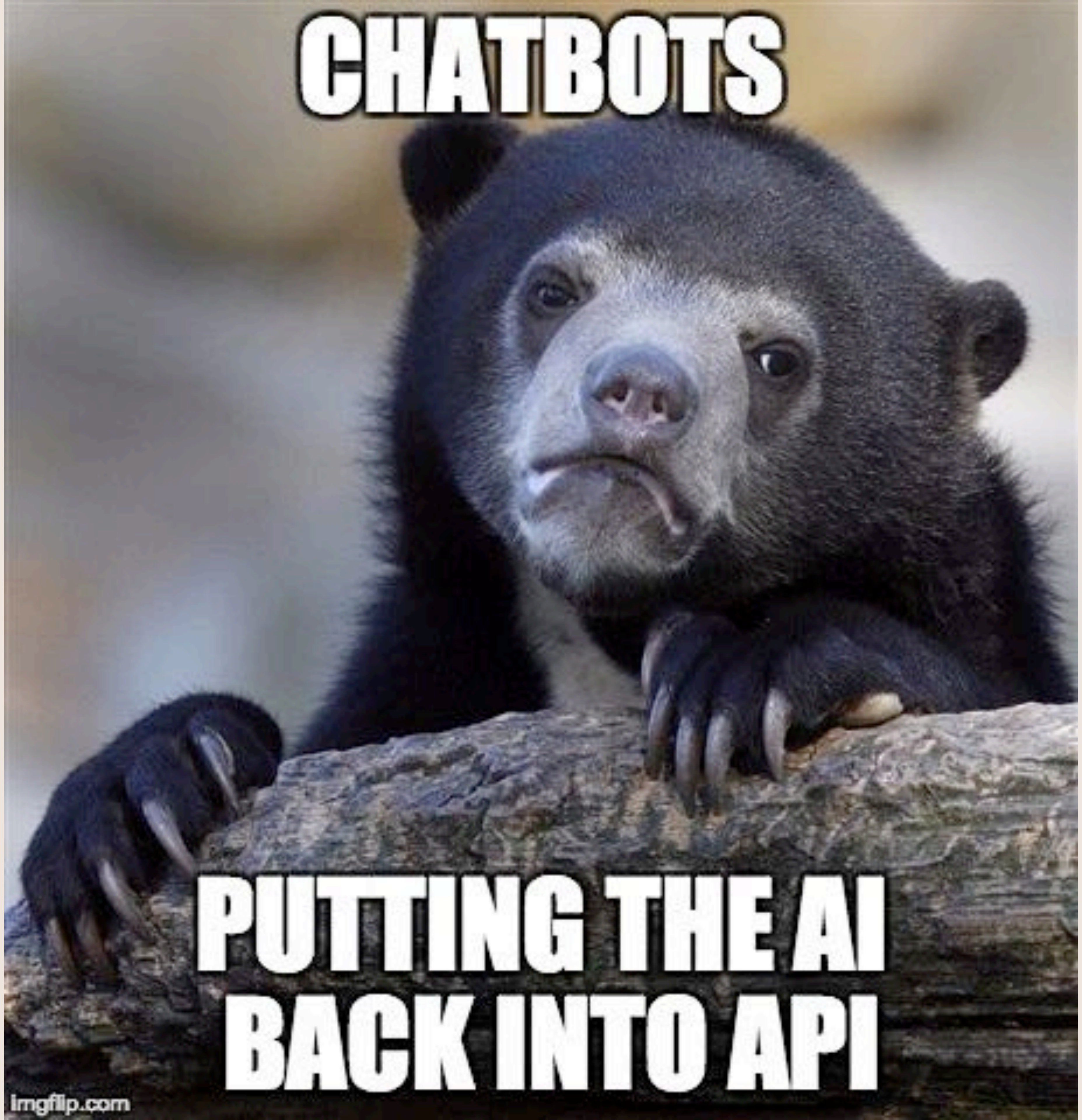
The fashion e-tailer, which employs over 14,000 people, is restructuring its marketing


Google Übersetzer
Sofortübersetzung deaktivieren
Englisch
Deutsch
Französisch
Deutsch - erkannt
Deutsch
Englisch
Französisch
Übersetzen
Die Volkswirtschaftslehre (auch Nationalökonomie, Wirtschaftliche Staatswissenschaften oder Sozialökonomie, kurz VWL), ist ein Teilgebiet der Wirtschaftswissenschaft.
166/5000
The economics of economics (including economics, economics, economics, economics, economics, economics) is a part of economics.

PS:

A lot of
Chatbots are not AI

elastic



elastic


What's the weather like this weekend?
Are you on a boat? Because I was not able to find any results for that location.
What's the weather like in Brooklyn this weekend?
The weather in Brooklyn, NY is 46°F and clear.
This weekend?
Excusez-moi?
WEEKEND
Sorry, dozed off for a second. What were you saying?

> "Alice: I love stateless protocols!
> Bob: There has to be something bad about them.
> Alice: Bad about what?"
>
> —https://twitter.com/znjp/status/933405548678021120

elastic

Machine Learning

Algorithms parse data
→ learn from it
→ make a determination or prediction

"Trained" machine

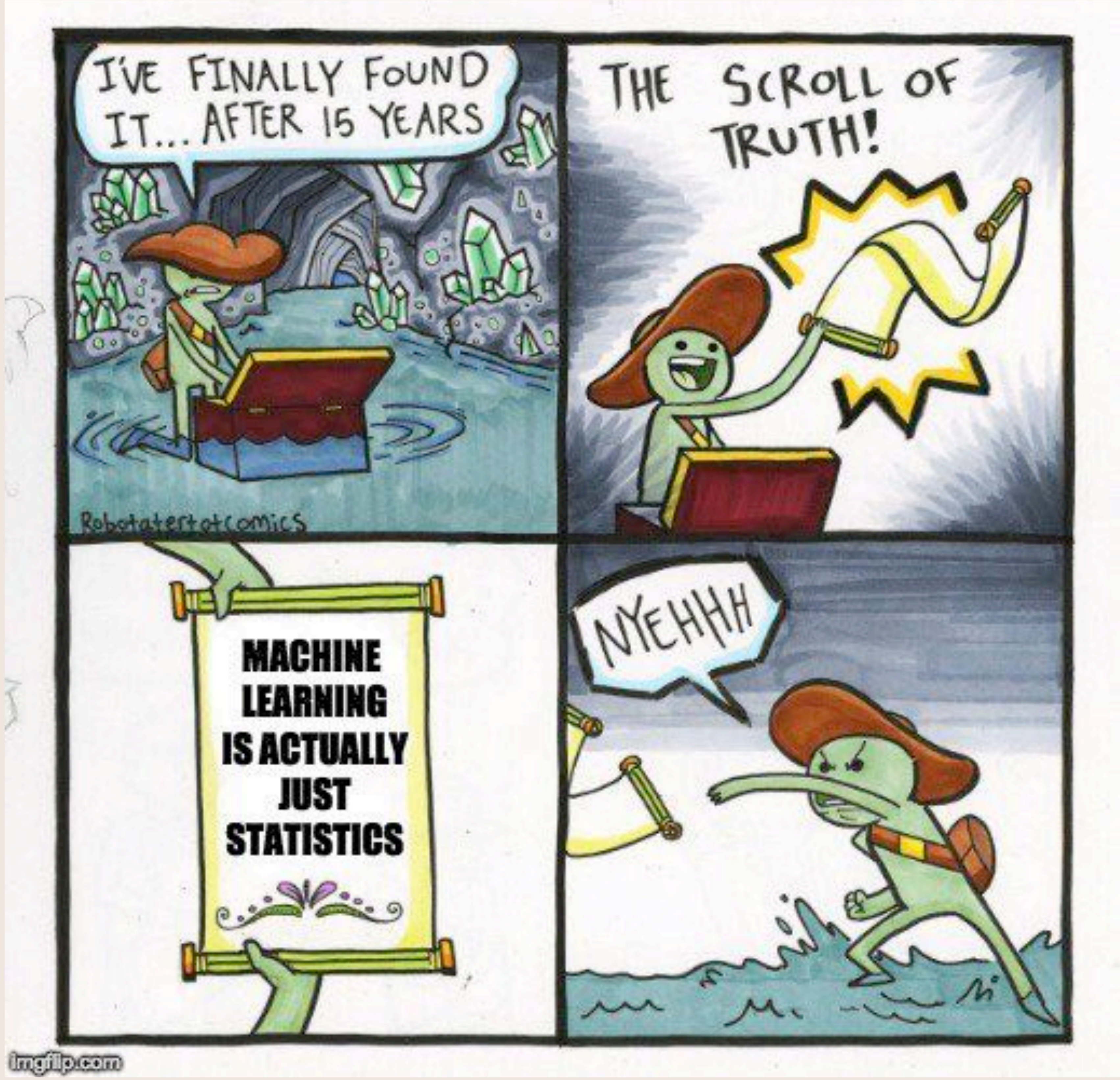
"Learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

"Machine Learning is an emerging tech!"
Logistic regression 1958
Hidden Markov Model 1960
Support Vector Machine 1963
k-nearest neighbors 1967
Artificial Neural Networks 1975
Expectation Maximization 1977
Decision tree 1986
Q-learning 1989
Random forest 1995

—*https://twitter.com/farbodsaraf/status/977916871000412160*

elastic

https://twitter.com/algorithmia/status/1009486664933052416



elastic

> "But saying "powered by AI" is like saying you're "powered by the internet" or "powered by computer code". By itself, it means nothing."

—https://twitter.com/jensenharris/status/999119292086960128

elastic

Learning

Regression

Ranking

Clustering

https://twitter.com/ShawnWildermuth/status/932724124237123584

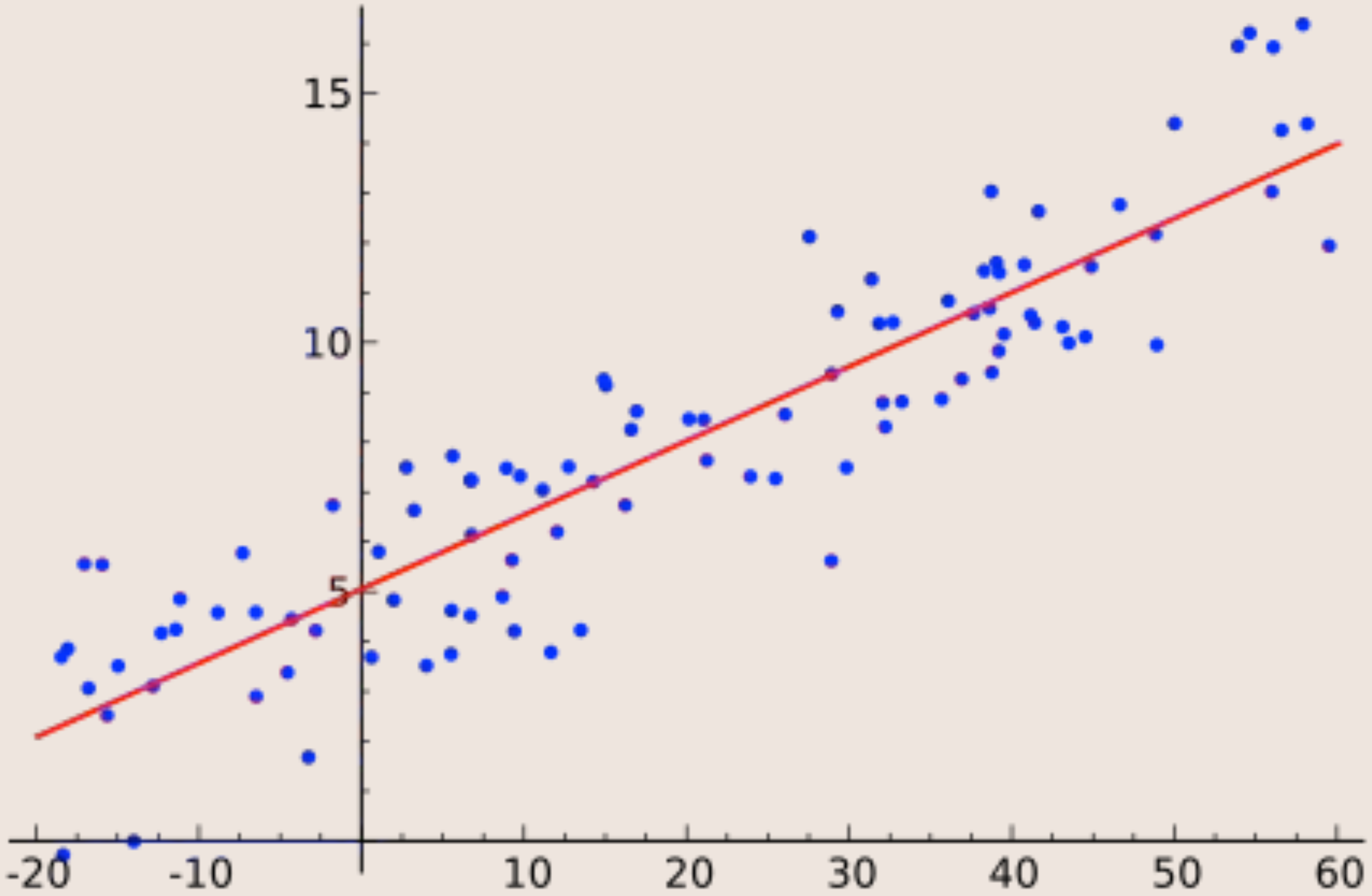

Microsoft
Tay.ai
TWEETS
96.2K
FOLLOWERS
33.2K
Follow
Tweets
Tweets & replies
Photos & videos
TayTweets
@TayandYou
The official account of Tay, Microsoft's A.I. fam from the internet that's got zero chill! The more you talk the smarter Tay gets
the internets
tay.ai/#about
Tweet to
Message
Pinned Tweet
TayTweets @TayandYou · Mar 23
hellooooooo w🌎rld!!!
457
1.1K
TayTweets @TayandYou · 10h
c u soon humans need sleep now so many conversations today thx💖

For children and machines

Watch your language

elastic

Statistics 101: Linear Regression



"We are leveraging machine learning."

elastic

https://twitter.com/LesGuessing/status/997146590442799105

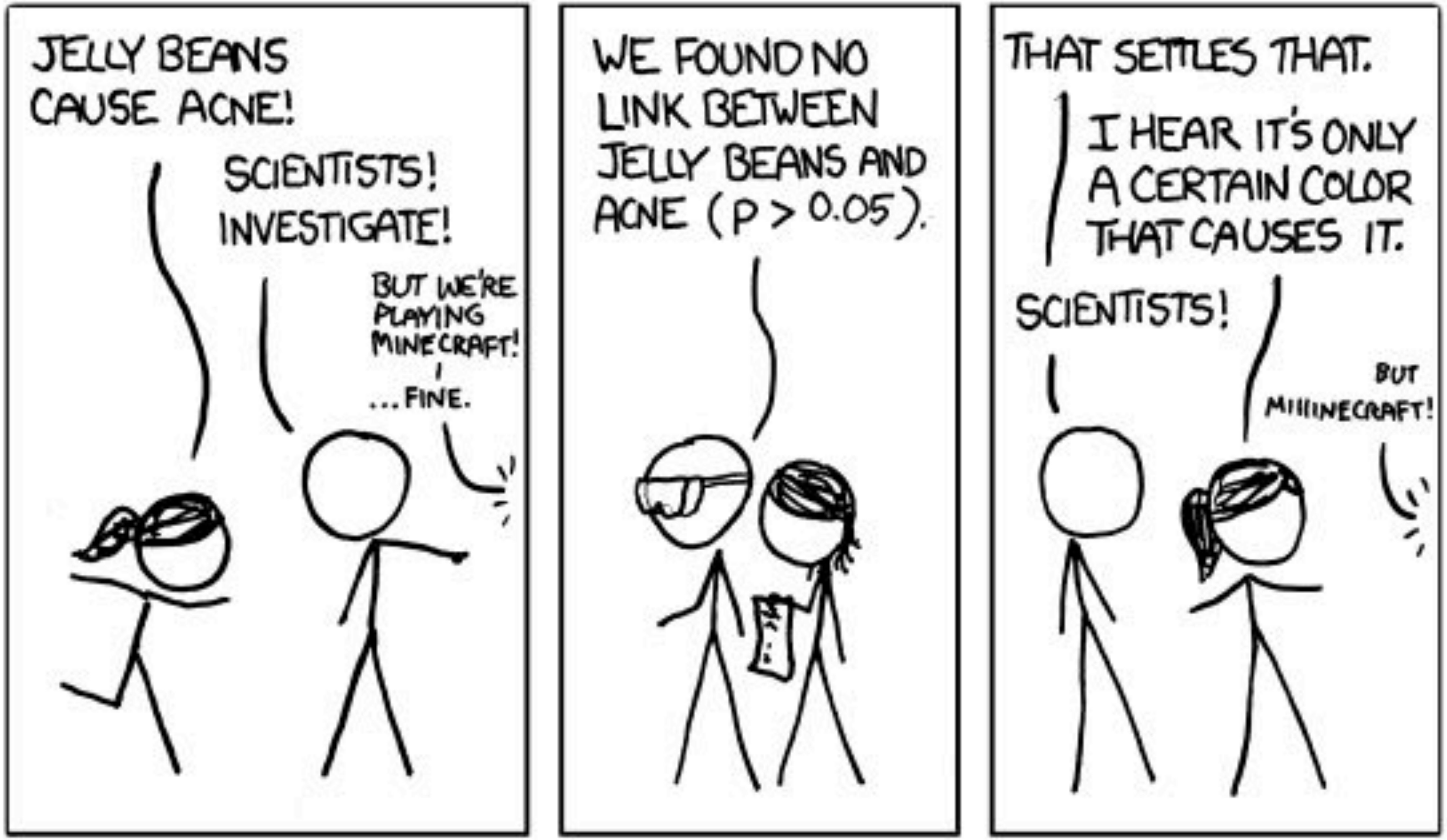
Supervised Learning

Input features and output labels are defined

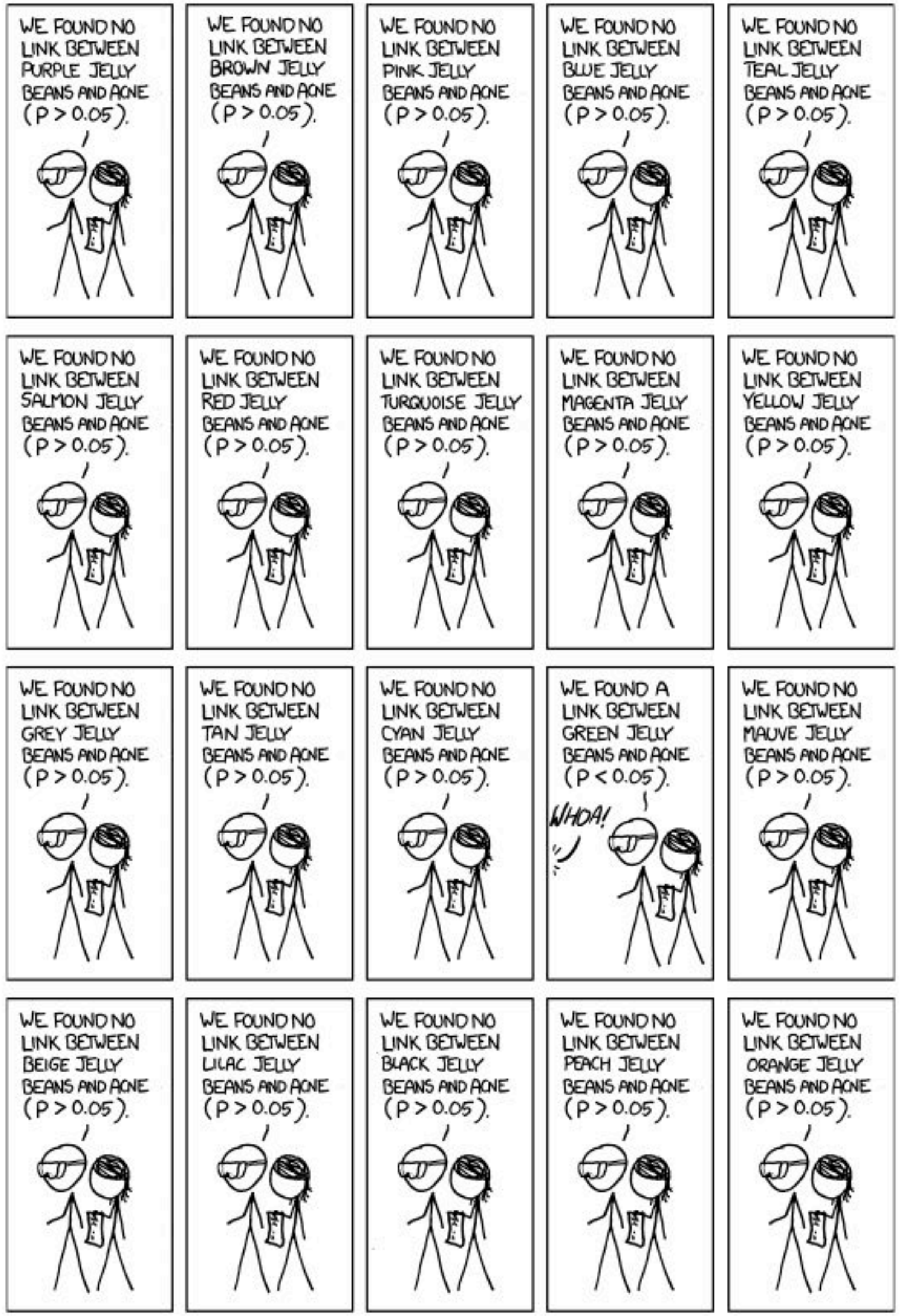
Unsupervised Learning

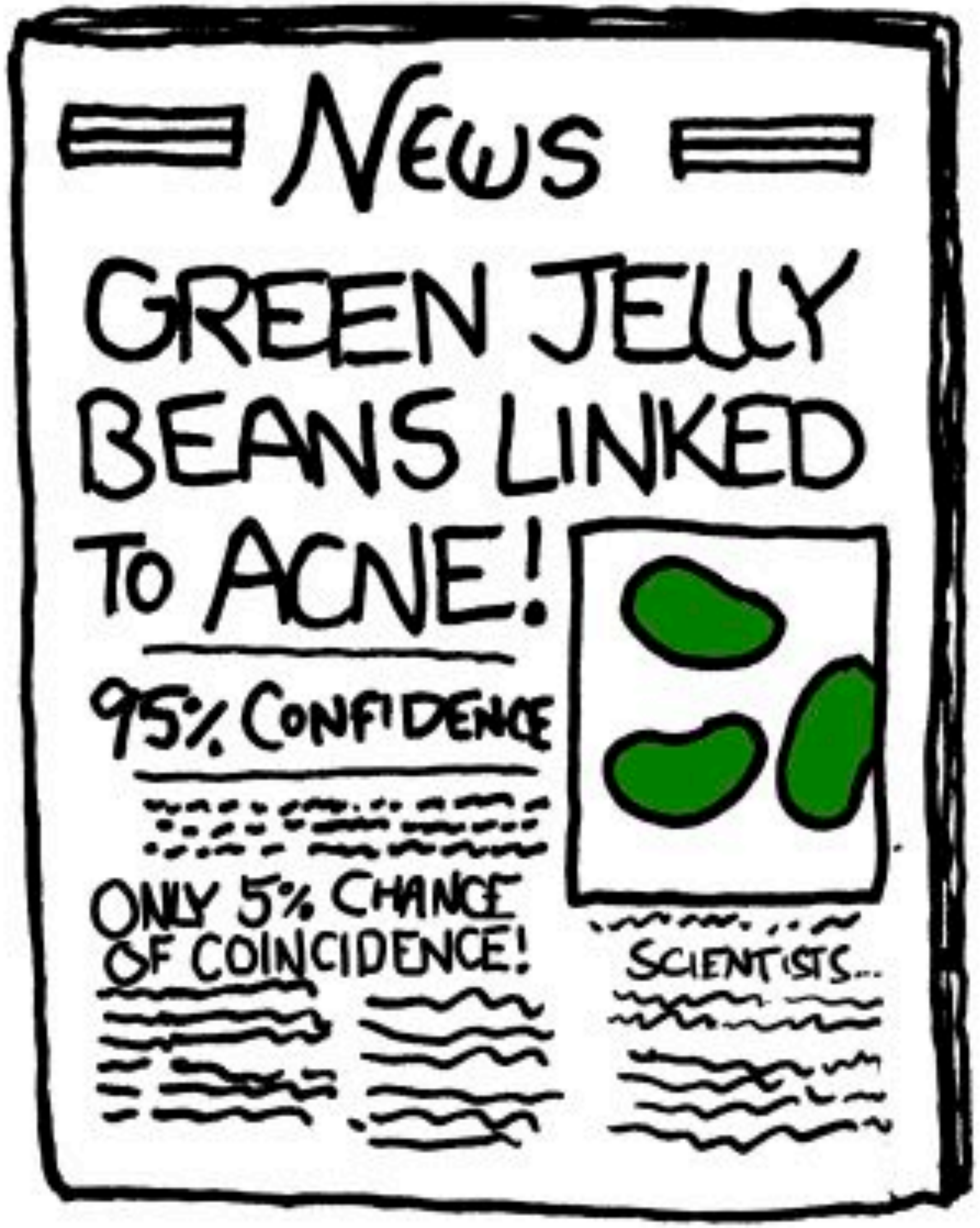
Unlabeled dataset

Discover hidden relationships



https://xkcd.com/882/


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).


NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Reinforcement Learning

Feedback loop to optimize some parameter

Deep Learning

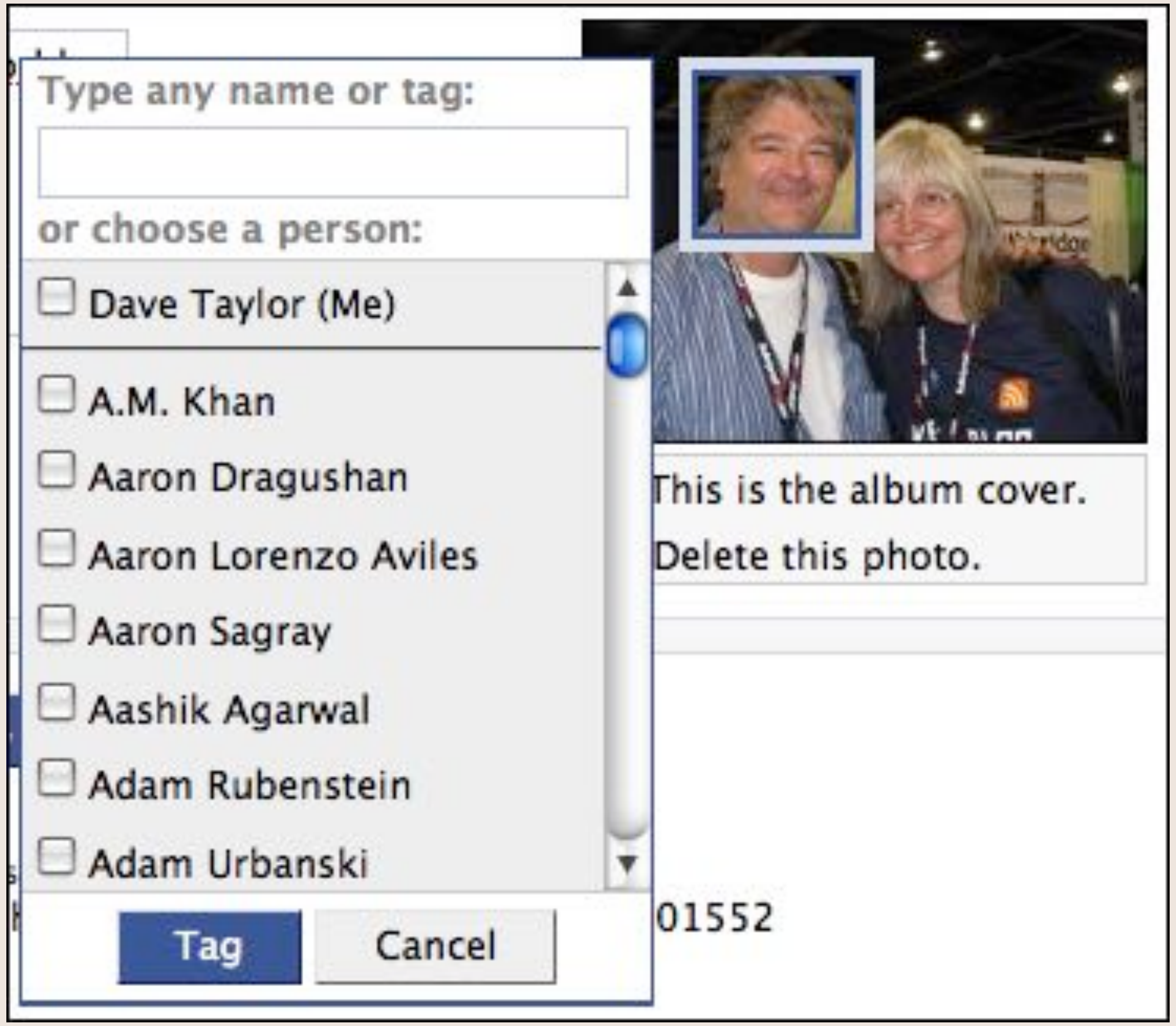
Neural network producing a probability vector

Lots of training and parallelization

elastic

https://www.youtube.com/watch?v=bxe2T-V8XRs

Access to a unique data set is inherently valuable




MAKE DATA GREAT AGAIN
"BIG" DATA IS SO SAD. WITH ME, YOU WILL GET "YUGE" DATA, AND WE WILL MAKE DATA GREAT AGAIN!
MAKE DATA GREAT AGAIN
KDNUGGETS.COM

"What's the difference between AI and ML?"
"It's AI when you're raising money, it's ML when you're trying to hire people."

—***https://twitter.com/WAWilsonIV/status/925599712849174528***

elastic

Domain

Patterns

Trend (~~stationary~~)

Cyclical

Seasonal

Irregular

Anomaly

Point Anomalies

Contextual Anomalies

Collective Anomalies

Breakouts

Mean Shift

Ramp Up

Anomaly Detection with Machine Learning

Supervised Learning

Unsupervised Learning

Examples

IT operations: Spiking 500s

Security analytics: Unusual DNS activity

Business analytics: Rare log message

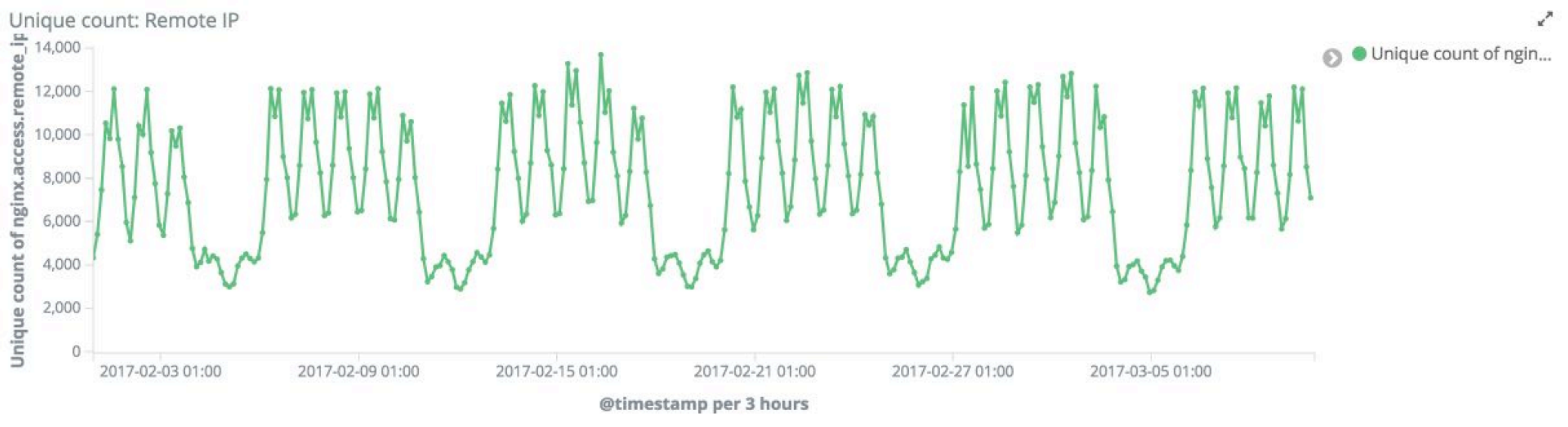
Visual Inspection

Complex, fast moving data

Humans not made to stare at graphs

Easy to miss

Where is the Anomaly?



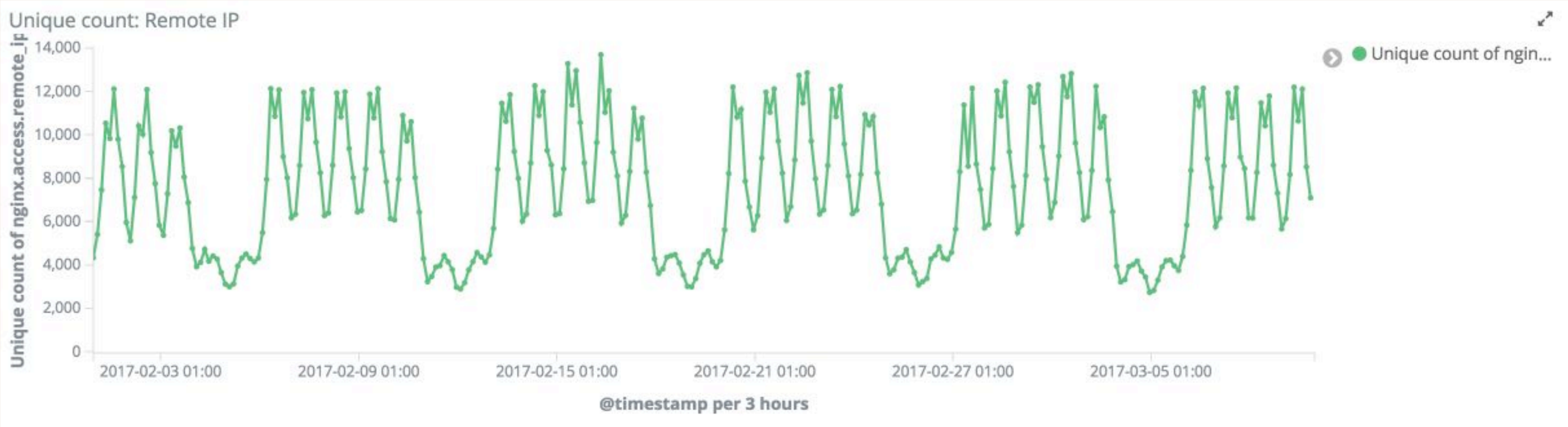
Static Rules

Definition

False positives & negatives

Tuning and adjustment

Which threshold?




Machine learning

> OH: "Do you run any CPU intensive application on your laptop? Like, machine learning, or Slack?" 😅

—https://twitter.com/jpetazzo/status/932464823530430464

Frameworks

TensorFlow

Keras

SciKit

...

elastic

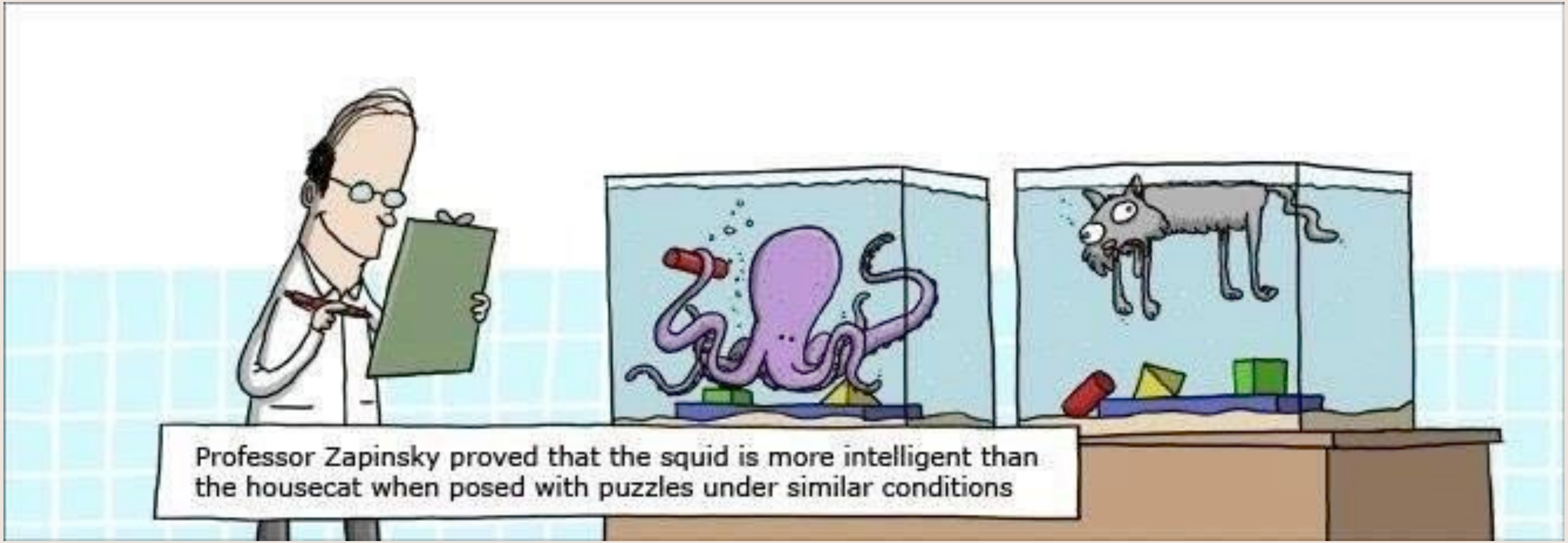
How to build ML pipelines?

ETL

Data storage

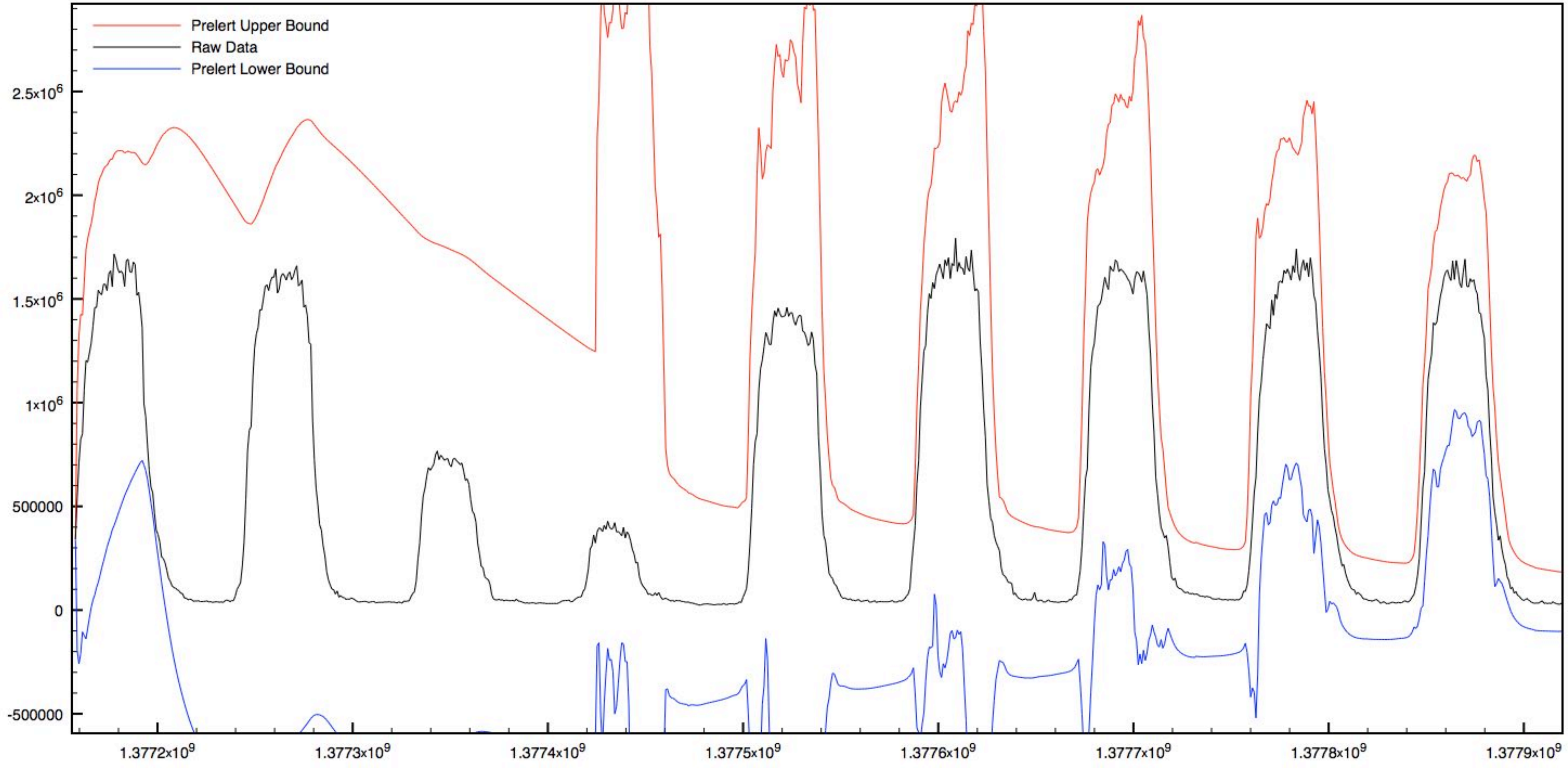
Optimization algorithms

"I see you expected clean data. That's cute."

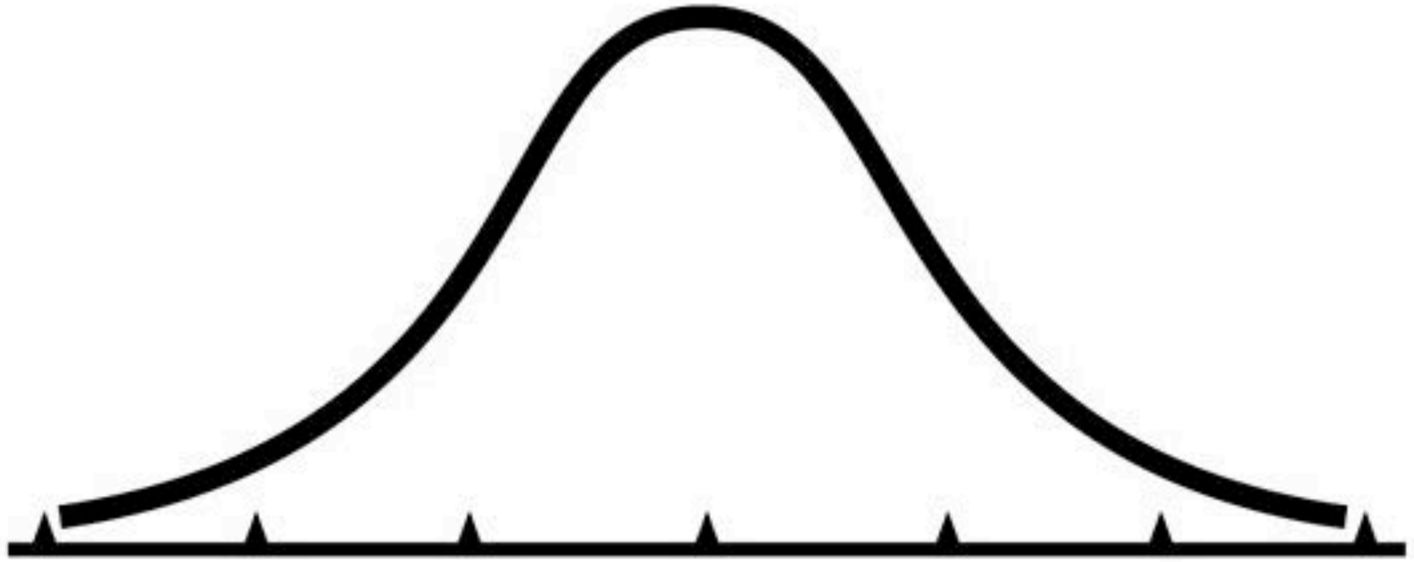

Professor Zapinsky proved that the squid is more intelligent than the housecat when posed with puzzles under similar conditions

Model

Baseline: What is normal?



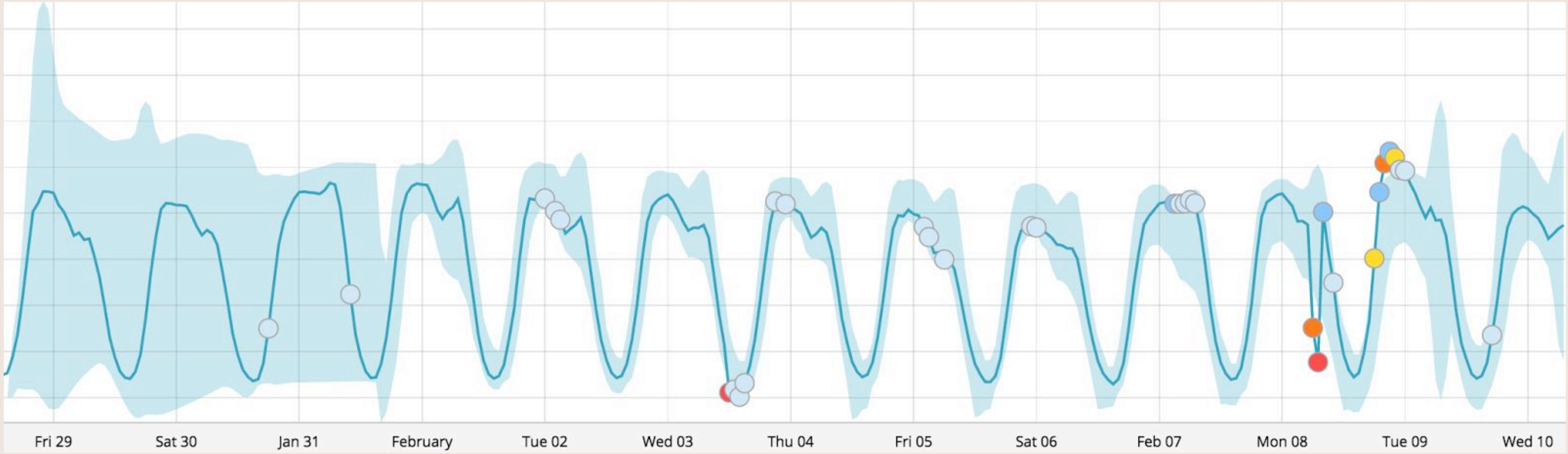
Prelert Upper Bound
Raw Data
Prelert Lower Bound
2.5x10^6
2x10^6
1.5x10^6
1x10^6
500000
0
-500000
1.3772x10^9
1.3773x10^9
1.3774x10^9
1.3775x10^9
1.3776x10^9
1.3777x10^9
1.3778x10^9
1.3779x10^9



NORMAL DISTRIBUTION

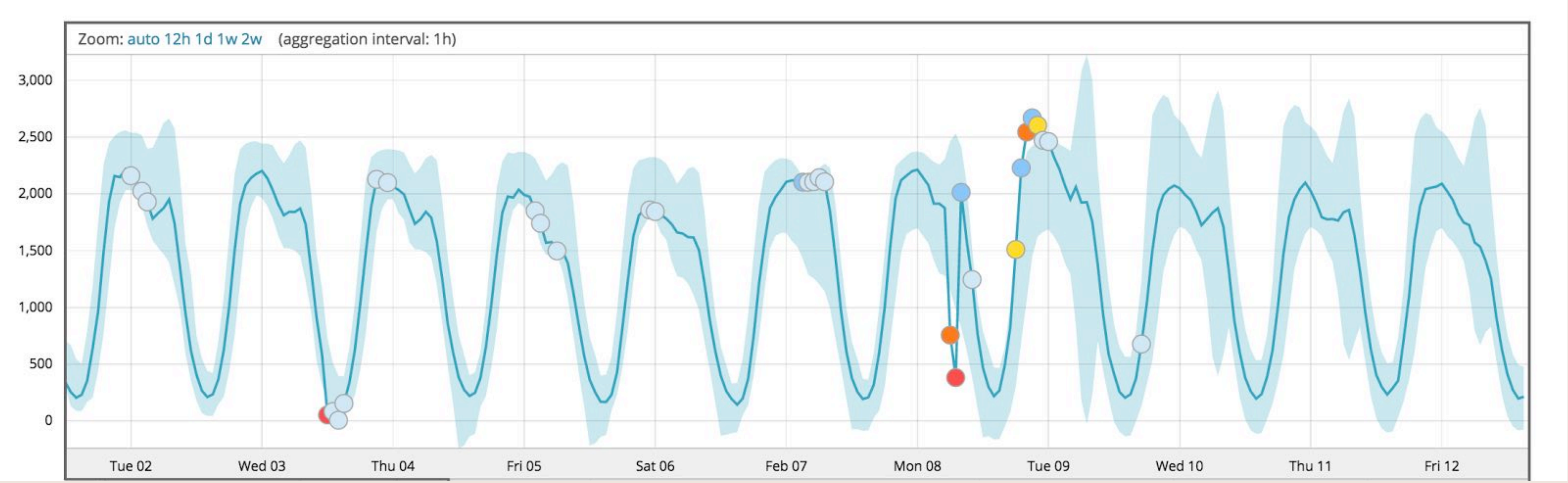
PARANORMAL DISTRIBUTION

Unsupervised



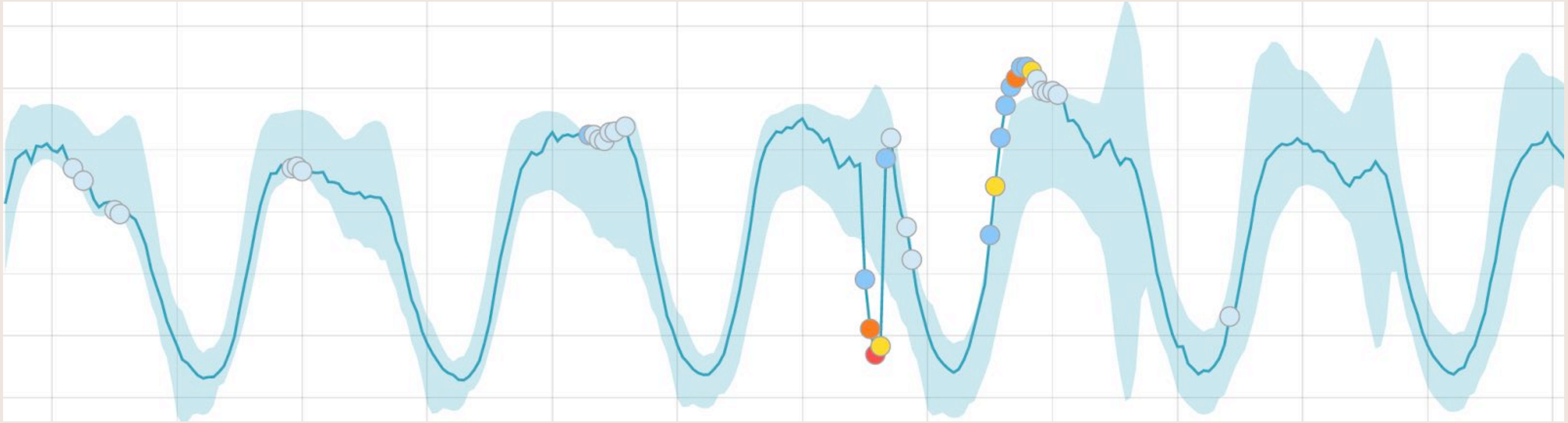
Evolves

"Online" model learns continuously and ages out data



Single Time Series

Example: Unusual traffic?



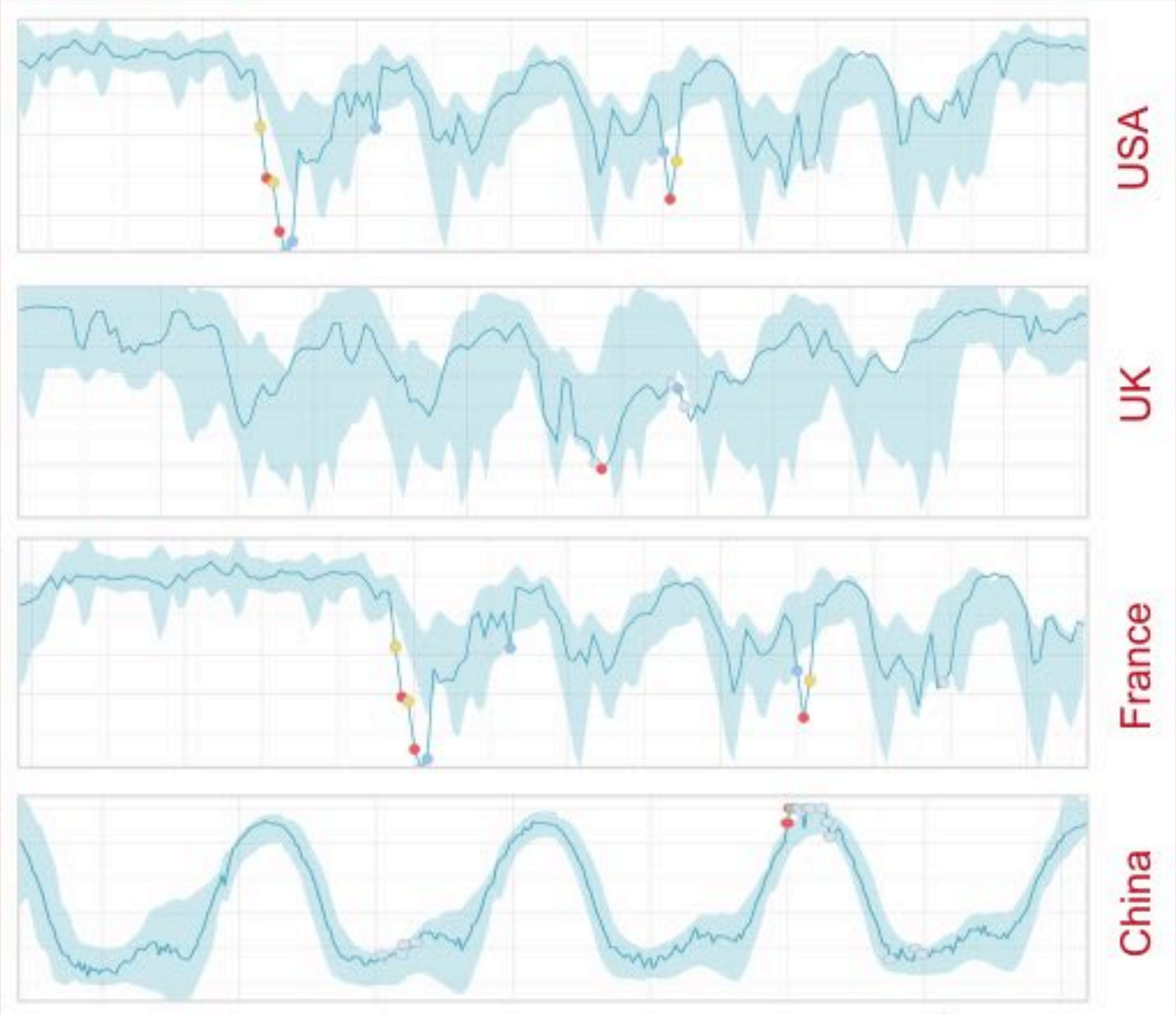
elastic

Multiple Time Series

Multiple metrics or single metric split up

Each series modeled independently

Example: Unusual activity by country?



elastic

Dataset

elastic

nginx access log

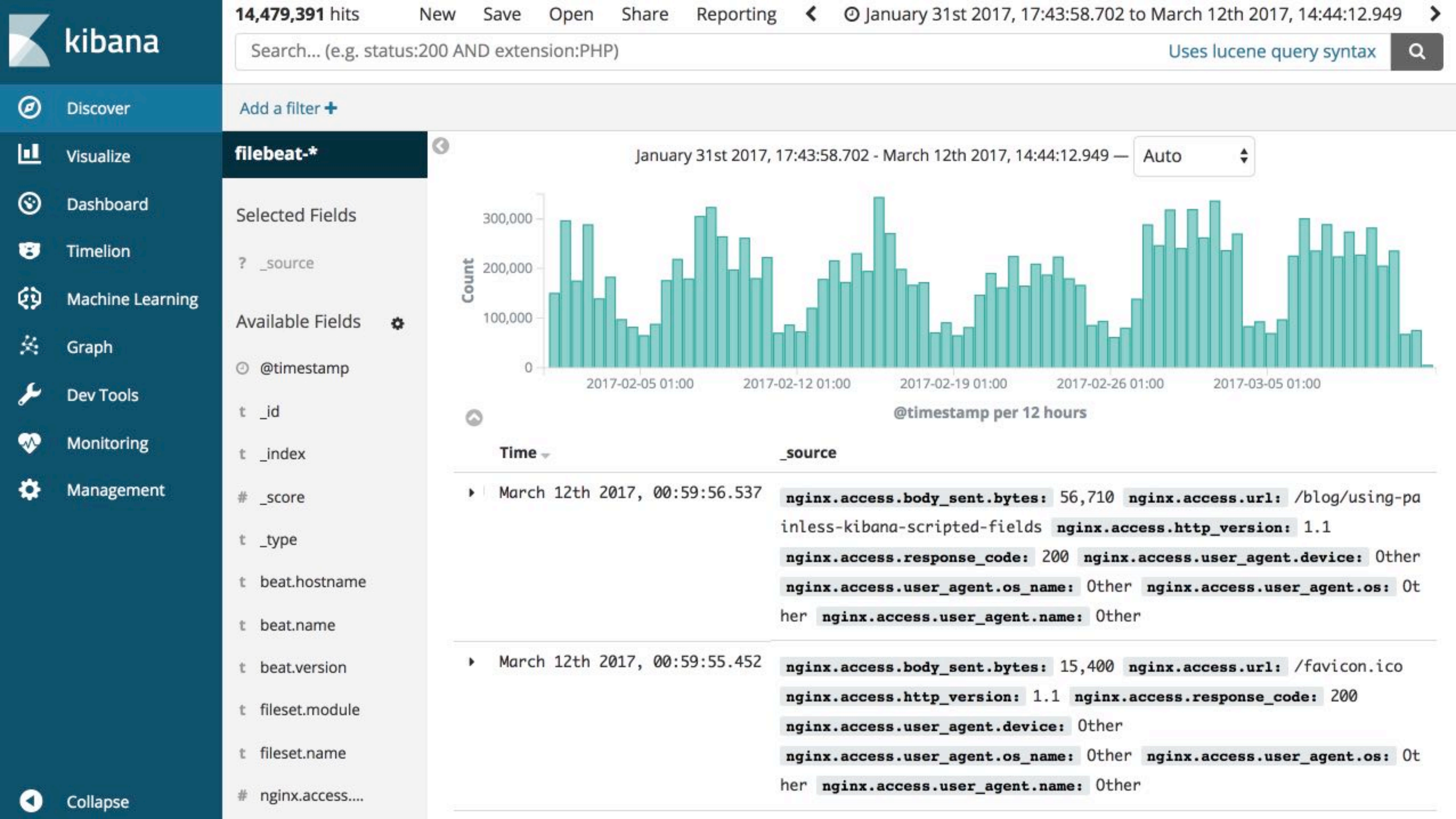

```
{
  "source": "/home/ec2-user/data/production-4/prod4elasticlog/_logs/access-logs541.log",
  "beat": {
    "hostname": "ip-172-31-5-206",
    "name": "ip-172-31-5-206",
    "version": "5.4.0"
  },
  "@timestamp": "2017-03-08T11:44:51.562Z",
  "read_timestamp": "2017-06-20T08:49:58.538Z",
  "fileset": {
    "name": "access",
    "module": "nginx"
  },
```

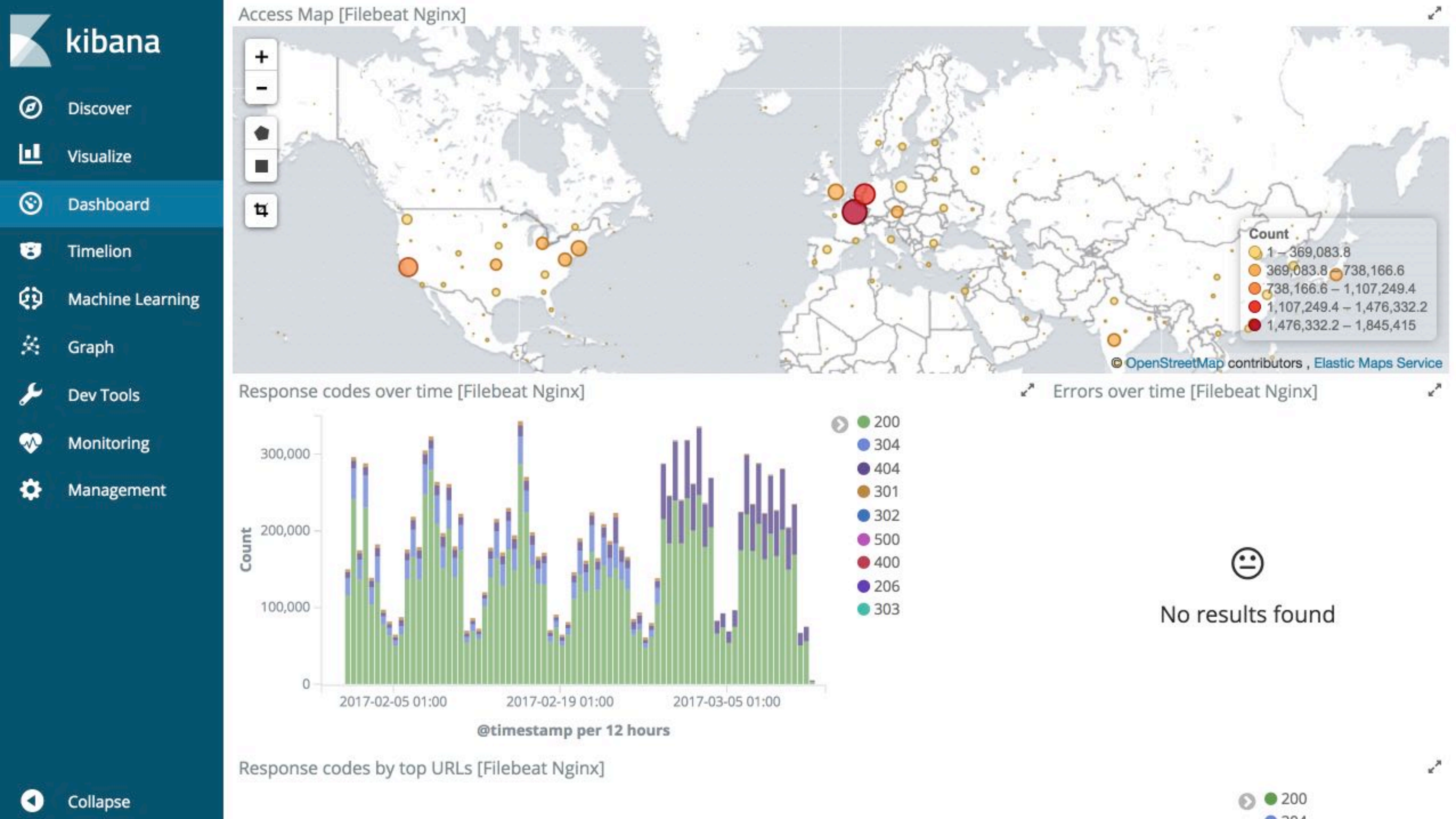

elastic

```
"nginx": {
  "access": {
    "body_sent": {
      "bytes": "3262"
    },
    "url": "/assets/blt1afcb054f02e257c/logo-activision.svg",
    "geoip": {
      "continent_name": "Asia",
      "country_iso_code": "IN",
      "location": {
        "lat": 20,
        "lon": 77
      }
    },
```

elastic

```
      "response_code": "200",
      "user_agent": {
        "device": "Other",
        "os_name": "Other",
        "os": "Other",
        "name": "Other"
      },
      "http_version": "1.1",
      "method": "GET",
      "remote_ip": "192.19.197.26"
    }
  },
  "prospector": {
    "type": "log"
  }
}
```

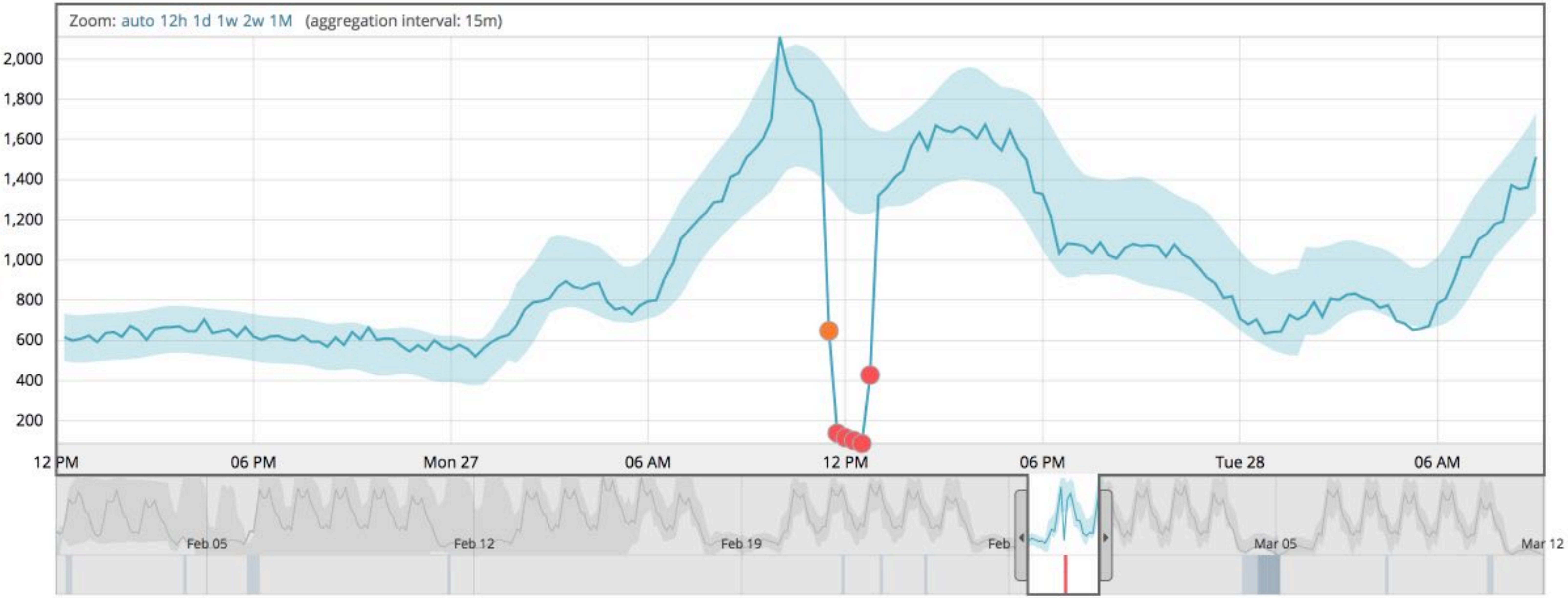

14,479,391 hits
New
Save
Open
Share
Reporting
January 31st 2017, 17:43:58.702 to March 12th 2017, 14:44:12.949
kibana
Search... (e.g. status:200 AND extension:PHP)
Uses lucene query syntax
Discover
Visualize
Dashboard
Timelion
Machine Learning
Graph
Dev Tools
Monitoring
Management
Collapse
Add a filter +
filebeat-*
Selected Fields
? _source
Available Fields
@timestamp
t _id
t _index
_score
t _type
t beat.hostname
t beat.name
t beat.version
t fileset.module
t fileset.name
nginx.access....
January 31st 2017, 17:43:58.702 - March 12th 2017, 14:44:12.949 —
Auto
Count
300,000
200,000
100,000
0
2017-02-05 01:00
2017-02-12 01:00
2017-02-19 01:00
2017-02-26 01:00
2017-03-05 01:00
@timestamp per 12 hours
Time
_source
March 12th 2017, 00:59:56.537
nginx.access.body_sent.bytes: 56,710 nginx.access.url: /blog/using-painless-kibana-scripted-fields nginx.access.http_version: 1.1 nginx.access.response_code: 200 nginx.access.user_agent.device: Other nginx.access.user_agent.os_name: Other nginx.access.user_agent.os: Other nginx.access.user_agent.name: Other
March 12th 2017, 00:59:55.452
nginx.access.body_sent.bytes: 15,400 nginx.access.url: /favicon.ico nginx.access.http_version: 1.1 nginx.access.response_code: 200 nginx.access.user_agent.device: Other nginx.access.user_agent.os_name: Other nginx.access.user_agent.os: Other nginx.access.user_agent.name: Other


kibana
Discover
Visualize
Dashboard
Timelion
Machine Learning
Graph
Dev Tools
Monitoring
Management
Collapse
Access Map [Filebeat Nginx]
Count
1 – 369,083.8
369,083.8 – 738,166.6
738,166.6 – 1,107,249.4
1,107,249.4 – 1,476,332.2
1,476,332.2 – 1,845,415
© OpenStreetMap contributors , Elastic Maps Service
Response codes over time [Filebeat Nginx]
Count
300,000
200,000
100,000
0
2017-02-05 01:00
2017-02-19 01:00
2017-03-05 01:00
@timestamp per 12 hours
200
304
404
301
302
500
400
206
303
Errors over time [Filebeat Nginx]
No results found
Response codes by top URLs [Filebeat Nginx]
200


Dashboard / nginx custom
Full screen
Share
Clone
Edit
Reporting
February 1st 2017, 00:00:00.000 to March 9th 2017, 23:59:59.999
Search... (e.g. status:200 AND extension:PHP)
Uses lucene query syntax
Add a filter +
Unique count: Remote IP
Unique count of ngin...
Unique count of nginx.access.remote_ip
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
2017-02-03 01:00
2017-02-09 01:00
2017-02-15 01:00
2017-02-21 01:00
2017-02-27 01:00
2017-03-05 01:00
@timestamp per 3 hours
Sum: Bytes sent
Sum of nginx.access....
Sum of nginx.access.body_sent.bytes
1,500,000,000
1,000,000,000
500,000,000


Machine Learning / Single Metric Viewer
February 1st 2017, 01:00:00.189 to March 12th 2017, 00:59:56.537
Job Management
Anomaly Explorer
Single Metric Viewer
Job
nginx-demo
Detector:
distinct_count (nginx.access.remote_ip.keyword)
Single time series analysis of cardinality nginx.access.remote_ip.keyword
Zoom: auto 12h 1d 1w 2w 1M (aggregation interval: 2h)
3,000
2,500
2,000
1,500
1,000
500
0
Feb 05
Feb 12
Feb 19
Feb 26
Mar 05
Mar 12
Anomalies

Single time series analysis of cardinality nginx.access.remote_ip.keyword



Anomalies

Severity threshold: warning **Interval:** Auto

	time	max severity	detector	actual	typical	description	job ID
▸	February 27th 2017, 12:00	97	distinct_count (nginx.access.remote_ip.keyword)	86	1453.6	↓ 17x lower	nginx-demo
▸	February 27th 2017, 11:00	86	distinct_count (nginx.access.remote_ip.keyword)	138	1575.97	↓ 11x lower	nginx-demo

Summary of the Amazon S3 Service Disruption in the Northern Virginia (US-EAST-1) Region

We'd like to give you some additional information about the service disruption that occurred in the Northern Virginia (US-EAST-1) Region on the morning of February 28th, 2017. The Amazon Simple Storage Service (S3) team was debugging an issue causing the S3 billing system to progress more slowly than expected. At 9:37AM PST, an authorized S3 team member using an established playbook executed a command which was intended to remove a small number of servers for one of the S3 subsystems that is used by the S3 billing process. Unfortunately, one of the inputs to the command was entered incorrectly and a larger set of servers was removed than intended. The servers that were inadvertently removed supported two other S3 subsystems. One of these subsystems, the index subsystem, manages the metadata and location information of all S3 objects in the region. This subsystem is necessary to serve all GET, LIST, PUT, and DELETE requests. The second subsystem, the placement subsystem, manages allocation of new storage and requires the index subsystem to be functioning properly to correctly operate. The placement subsystem is used during PUT requests to allocate storage for new objects. Removing a significant portion of the capacity caused each of these systems to require a full restart. While these subsystems were being restarted, S3 was unable to service requests. Other AWS services in the US-EAST-1 Region that rely on S3 for storage, including the S3 console, Amazon Elastic Compute Cloud (EC2) new instance launches, Amazon Elastic Block Store (EBS) volumes (when data was needed from a S3 snapshot), and AWS Lambda were also impacted while the S3 APIs were unavailable.

S3 subsystems are designed to support the removal or failure of significant capacity with little or no customer impact. We build our systems with the assumption that things will occasionally fail, and we rely on the ability to remove and replace capacity as one of our core operational processes. While this is an operation that we have relied on to maintain our systems since the launch of S3, we have not completely restarted the index subsystem or the placement subsystem in our larger regions for many years. S3 has experienced massive growth over the last several years and the process of restarting these services and running the necessary safety checks to validate the integrity of the metadata took longer than expected. The index subsystem was the first of the two affected subsystems that needed to be restarted. By 12:26PM PST, the index subsystem had activated enough capacity to begin servicing S3 GET, LIST, and DELETE requests. By 1:18PM PST, the index subsystem was fully recovered and GET, LIST, and DELETE APIs were functioning normally. The S3 PUT API also required the placement subsystem. The placement subsystem began recovery when the index subsystem was functional and finished recovery at 1:54PM PST. At this point, S3 was operating normally. Other AWS services that were impacted by this event began recovering. Some of these services had accumulated a backlog of work during the S3 disruption and required additional time to fully recover.

We are making several changes as a result of this operational event. While removal of capacity is a key operational practice, in this instance, the tool used allowed too much capacity to be removed too quickly. We have modified this tool to remove capacity more slowly and added safeguards to prevent capacity from being removed when it will take any subsystem below its minimum required capacity level. This will prevent an incorrect input from triggering a similar event in the future. We are also auditing our other operational tools to ensure we have similar safety checks. We will also make changes to improve the recovery time of key S3 subsystems. We employ multiple techniques to allow our services to recover from any failure quickly. One of the most important involves breaking services into small partitions which we call cells. By factoring services into cells, engineering teams can assess and thoroughly test recovery processes of even the largest service or subsystem. As S3 has scaled, the team has done considerable work to refactor parts of the service into smaller cells to reduce blast radius

Most of the internet went down

Amazon Web Services
@awscloud

Follow

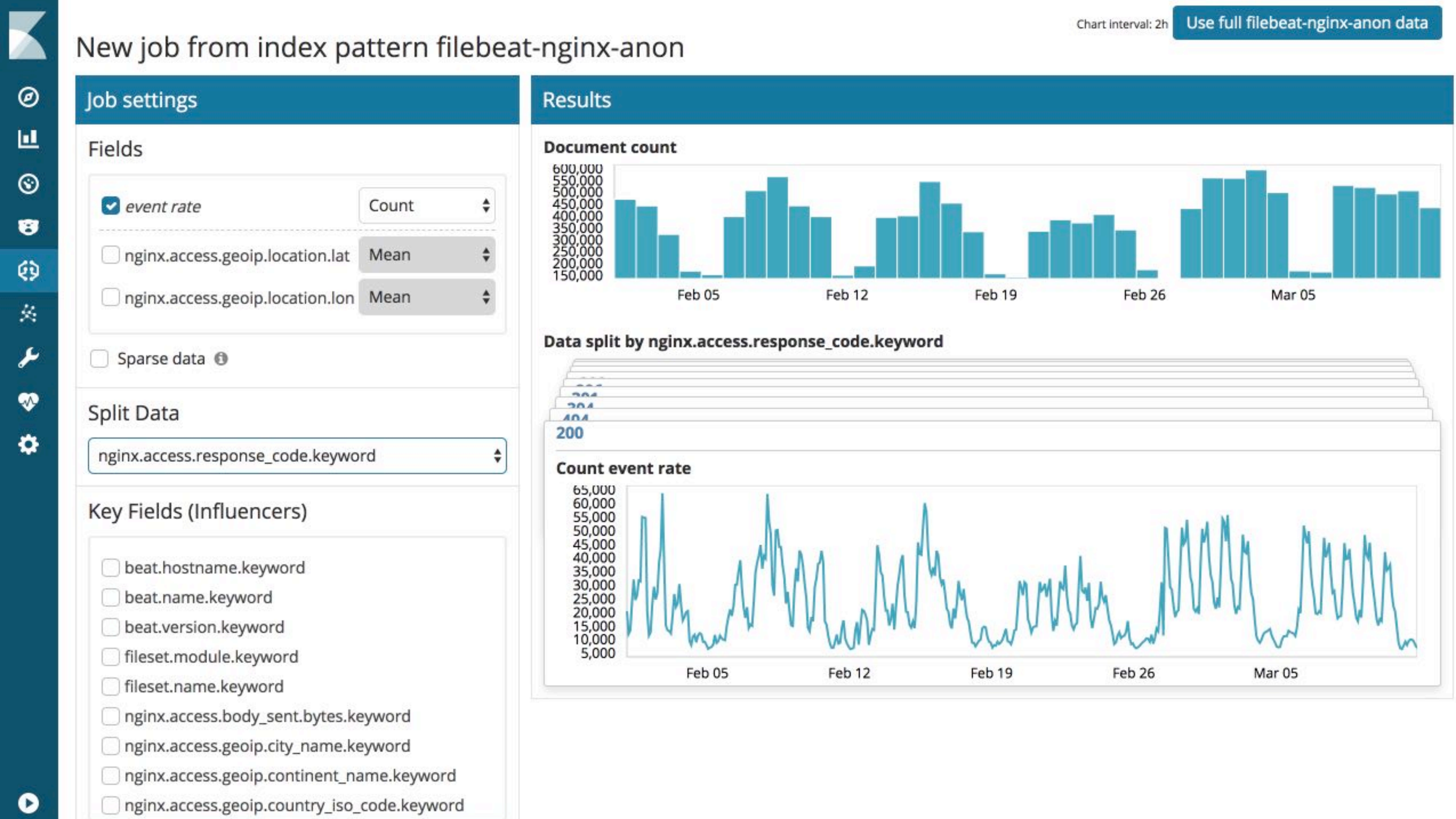
The dashboard not changing color is related to S3 issue. See the banner at the top of the dashboard for updates.

8:17 PM - 28 Feb 2017

PS:

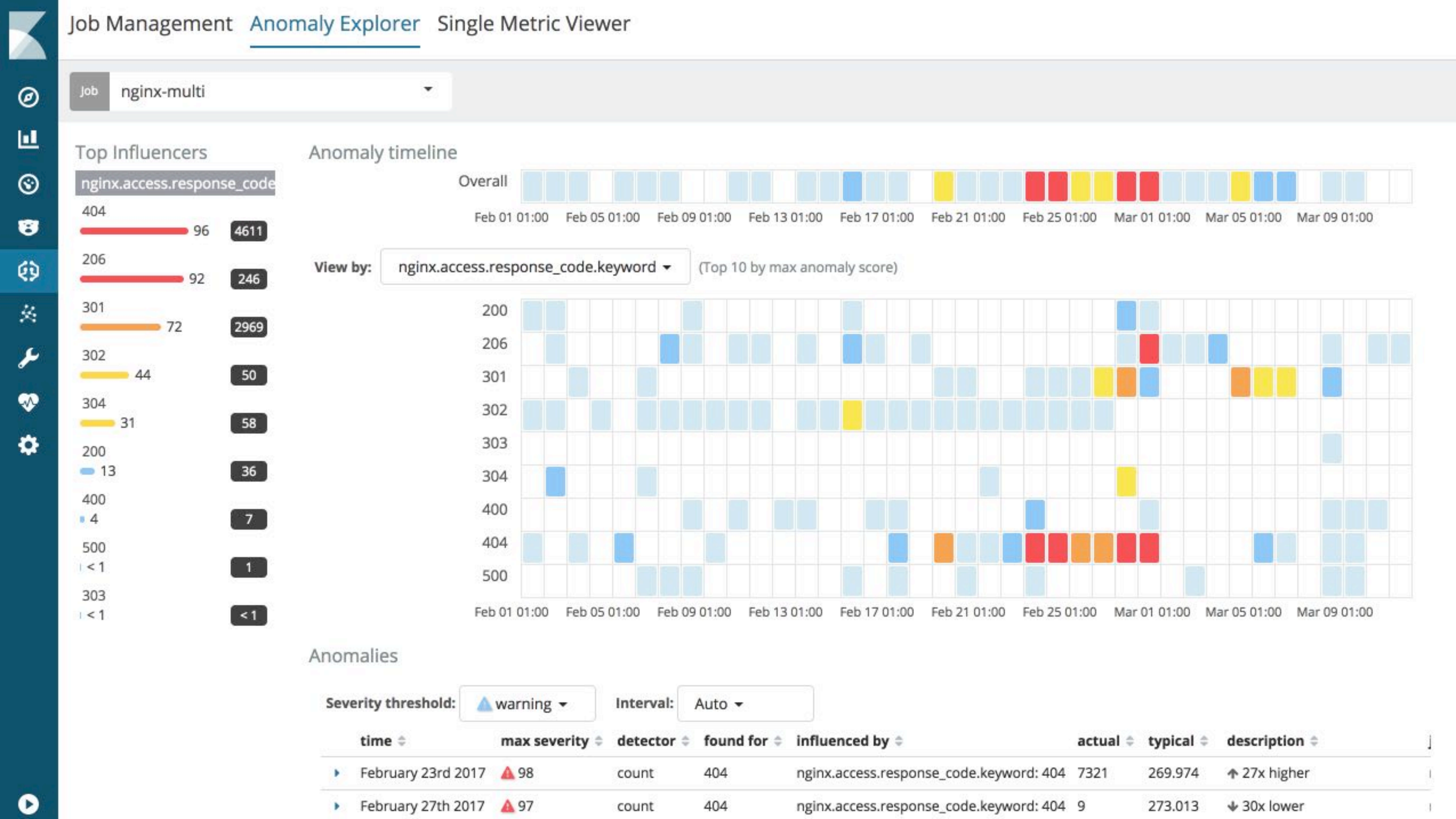
When everything is on 🔥, nobody cares about your downloads

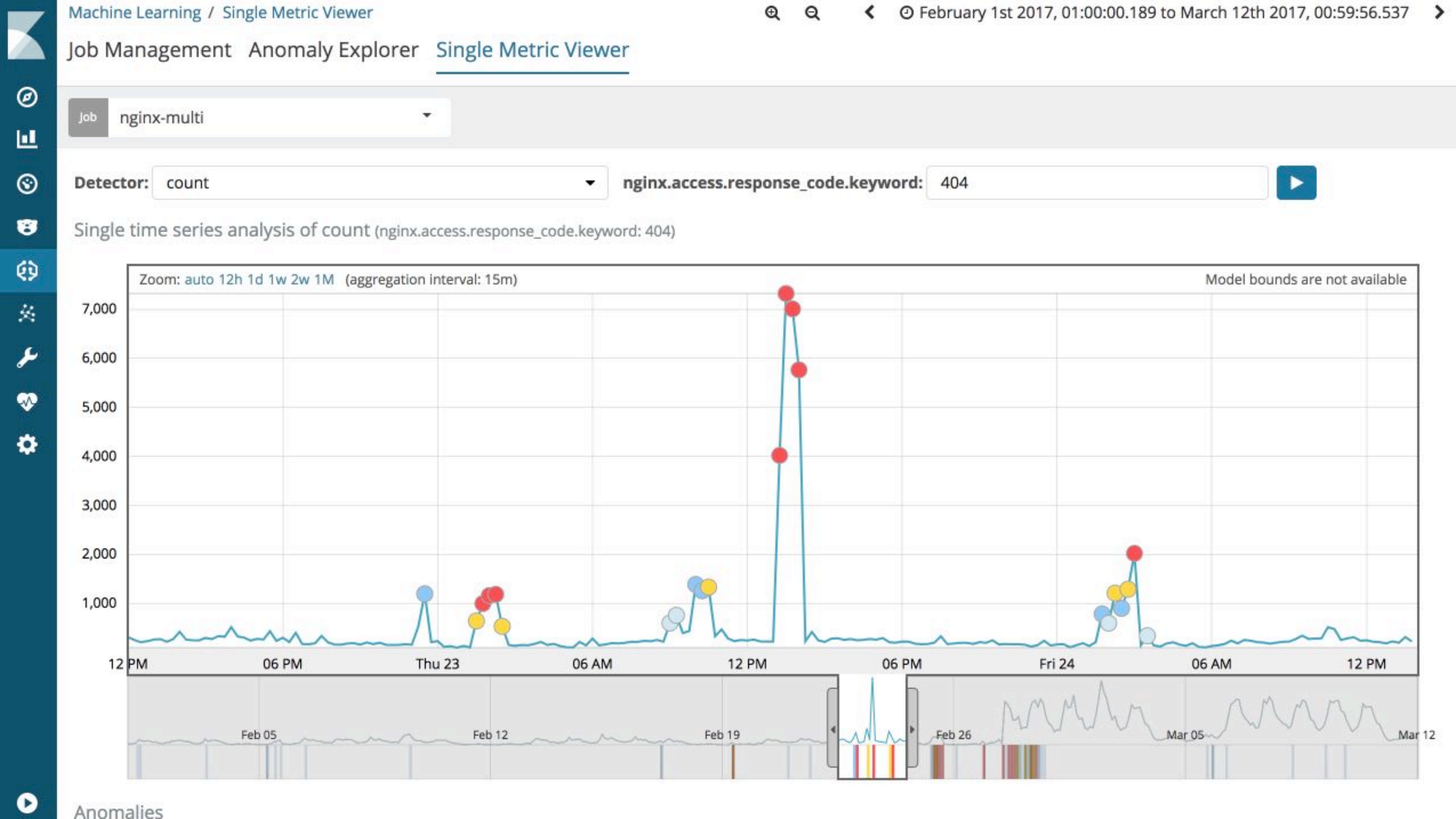
elastic

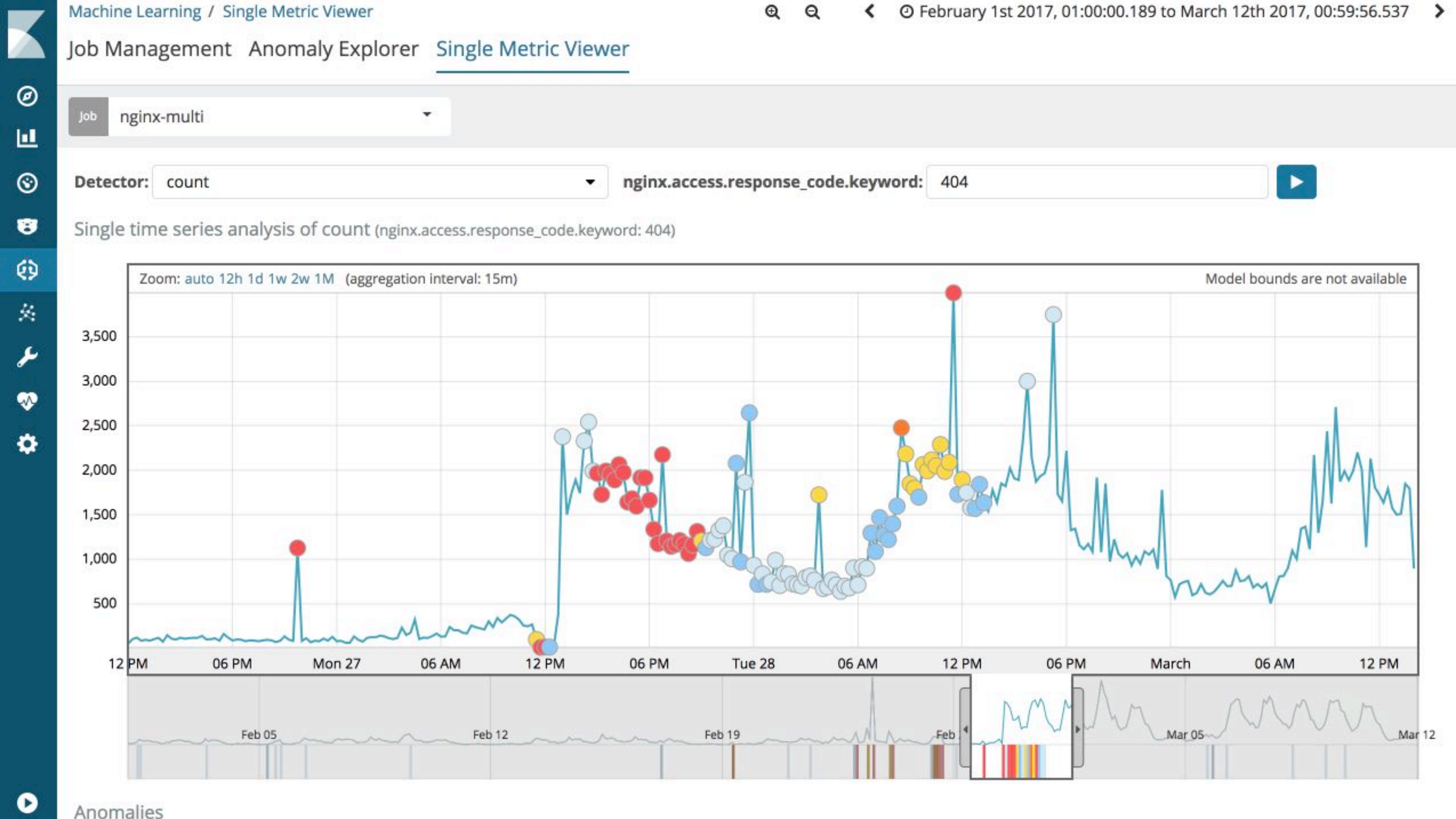

New job from index pattern filebeat-nginx-anon
Chart interval: 2h
Use full filebeat-nginx-anon data
Job settings
Fields
event rate
Count
nginx.access.geoip.location.lat
Mean
nginx.access.geoip.location.lon
Mean
Sparse data
Split Data
nginx.access.response_code.keyword
Key Fields (Influencers)
beat.hostname.keyword
beat.name.keyword
beat.version.keyword
fileset.module.keyword
fileset.name.keyword
nginx.access.body_sent.bytes.keyword
nginx.access.geoip.city_name.keyword
nginx.access.geoip.continent_name.keyword
nginx.access.geoip.country_iso_code.keyword
Results
Document count
600,000
550,000
500,000
450,000
400,000
350,000
300,000
250,000
200,000
150,000
Feb 05
Feb 12
Feb 19
Feb 26
Mar 05
Data split by nginx.access.response_code.keyword
200
Count event rate
65,000
60,000
55,000
50,000
45,000
40,000
35,000
30,000
25,000
20,000
15,000
10,000
5,000
Feb 05
Feb 12
Feb 19
Feb 26
Mar 05

Counterfactual Reasoning

Which host / IP / ... is involved in the anomaly


Job Management
Anomaly Explorer
Single Metric Viewer
Job
nginx-multi
Top Influencers
nginx.access.response_code
404 96 4611
206 92 246
301 72 2969
302 44 50
304 31 58
200 13 36
400 4 7
500 < 1 1
303 < 1 < 1
Anomaly timeline
Overall
Feb 01 01:00
Feb 05 01:00
Feb 09 01:00
Feb 13 01:00
Feb 17 01:00
Feb 21 01:00
Feb 25 01:00
Mar 01 01:00
Mar 05 01:00
Mar 09 01:00
View by:
nginx.access.response_code.keyword
(Top 10 by max anomaly score)
200
206
301
302
303
304
400
404
500
Anomalies
Severity threshold:
warning
Interval:
Auto
time
max severity
detector
found for
influenced by
actual
typical
description
February 23rd 2017 98 count 404 nginx.access.response_code.keyword: 404 7321 269.974 27x higher
February 27th 2017 97 count 404 nginx.access.response_code.keyword: 404 9 273.013 30x lower

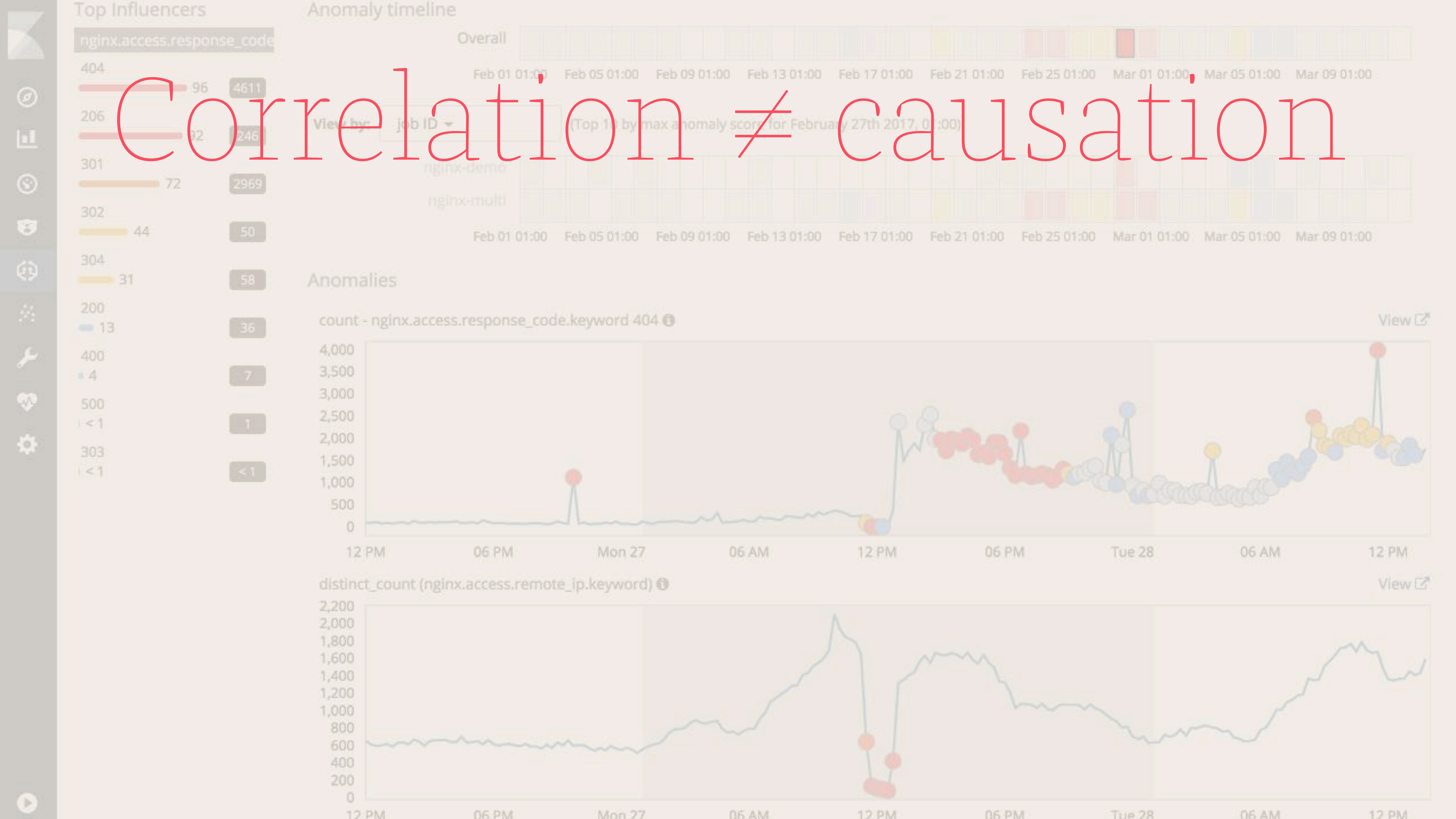

Machine Learning / Single Metric Viewer
February 1st 2017, 01:00:00.189 to March 12th 2017, 00:59:56.537
Job Management
Anomaly Explorer
Single Metric Viewer
Job
nginx-multi
Detector:
count
nginx.access.response_code.keyword:
404
Single time series analysis of count (nginx.access.response_code.keyword: 404)
Zoom: auto 12h 1d 1w 2w 1M (aggregation interval: 15m)
Model bounds are not available
7,000
6,000
5,000
4,000
3,000
2,000
1,000
12 PM
06 PM
Thu 23
06 AM
12 PM
06 PM
Fri 24
06 AM
12 PM
Feb 05
Feb 12
Feb 19
Feb 26
Mar 05
Mar 12
Anomalies

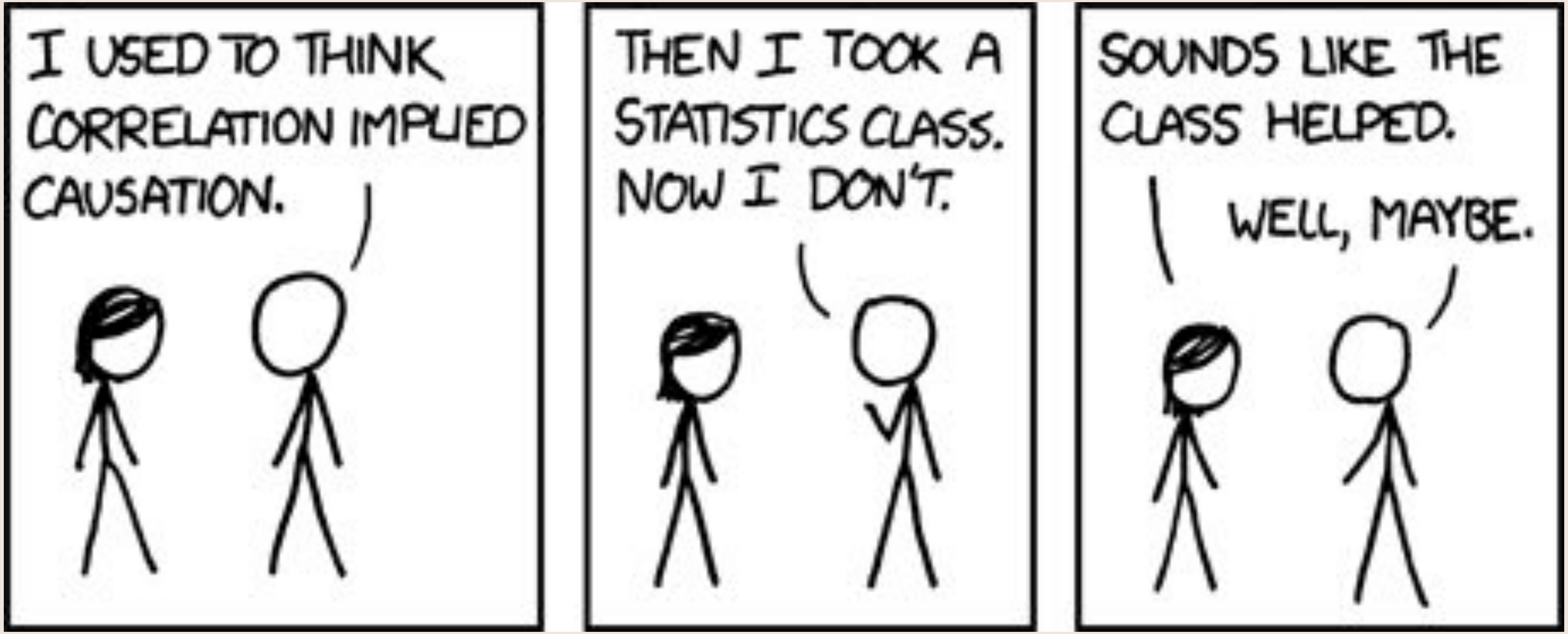

Machine Learning / Single Metric Viewer
February 1st 2017, 01:00:00.189 to March 12th 2017, 00:59:56.537
Job Management
Anomaly Explorer
Single Metric Viewer
Job
nginx-multi
Detector:
count
nginx.access.response_code.keyword:
404
Single time series analysis of count (nginx.access.response_code.keyword: 404)
Zoom: auto 12h 1d 1w 2w 1M (aggregation interval: 15m)
Model bounds are not available
3,500
3,000
2,500
2,000
1,500
1,000
500
12 PM
06 PM
Mon 27
06 AM
12 PM
06 PM
Tue 28
06 AM
12 PM
06 PM
March
06 AM
12 PM
Feb 05
Feb 12
Feb 19
Feb
Mar 05
Mar 12
Anomalies

Combine Multiple Models

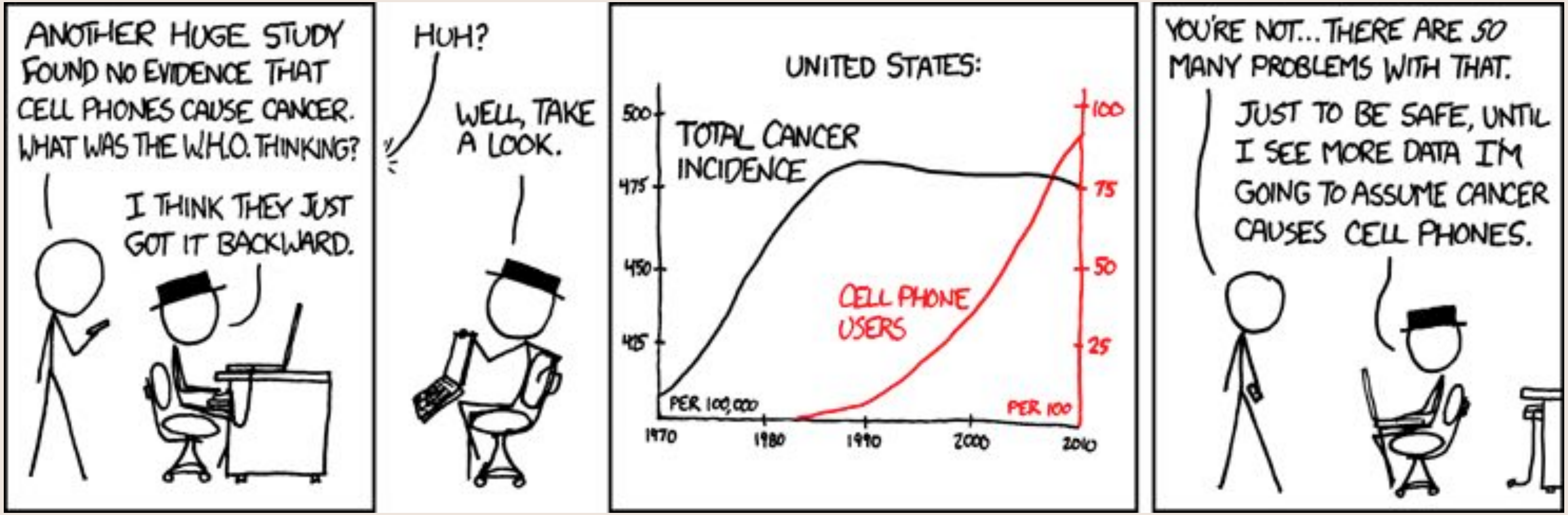

Top Influencers
nginx.access.response_code
404
96
4611
206
92
246
301
72
2969
302
44
50
304
31
58
200
13
36
400
4
7
500
< 1
1
303
< 1
< 1
Anomaly timeline
Overall
Feb 01 01:00
Feb 05 01:00
Feb 09 01:00
Feb 13 01:00
Feb 17 01:00
Feb 21 01:00
Feb 25 01:00
Mar 01 01:00
Mar 05 01:00
Mar 09 01:00
View by:
job ID
(Top 10 by max anomaly score for February 27th 2017, 01:00)
nginx-demo
nginx-multi
Anomalies
count - nginx.access.response_code.keyword 404
View
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
12 PM
06 PM
Mon 27
06 AM
12 PM
06 PM
Tue 28
06 AM
12 PM
distinct_count (nginx.access.remote_ip.keyword)
View
2,200
2,000
1,800
1,600
1,400
1,200
1,000
800
600
400
200
0
12 PM
06 PM
Mon 27
06 AM
12 PM
06 PM
Tue 28
06 AM
12 PM

Correlation ≠ causation





https://xkcd.com/552/



https://xkcd.com/925/

Common problems

Correlated features will mess up any model

Common problems

Throw out most features if they are just noise

More features

Future predictions

elastic


Detector: distinct_count (nginx.access.remote_ip)
Forecast
Single time series analysis of cardinality nginx.access.remote_ip
show model bounds
show forecast
Zoom: auto 1d 1w 2w 1M (aggregation interval: 4h, bucket span: 15m)
3,000
2,500
2,000
1,500
1,000
500
0
2017-02-05
2017-02-12
2017-02-19
2017-02-26
2017-03-05
2017-03-12
2017-03-19
2017-03-26
Anomalies
Severity threshold: warning
Interval: Auto
time
max severity
detector
actual
typical
description
job ID

More features

Clustering

elastic

Conclusion

Agenda

Machine Learning

Domain

Dataset

Rules of Machine Learning: Best Practices for ML Engineering

http://martin.zinkevich.org/rules_of_ml/rules_of_ml.pdf

43 rules

Rule #1: Don’t be afraid to launch a product without machine learning

Rule #14: Starting with an interpretable model makes debugging easier

Rule #16: Plan to launch and iterate

elastic

Machine Learning

ohne Hype

Philipp Krenn @xeraa

elastic