



Analysis and interpretation of monitoring data

Yury Yafimachau



Principal developer
SolarWinds Backup

Why is this worth talking about?







Monitoring metrics

- Common
 - Load average
 - CPU/memory/disk usage
 - IO/network bandwidth
- Application specific
 - Request count/timing
 - Version/States
 - Backup duration/status
 - Used/Selected /Processed/Changed size

Monitoring goals

- **Alerts & Notifications**

Fast, Precise, Reliable

- **Investigation tool** (dashboard)

Fast, Simple & Self-sufficient, High definition, Medium retention

- **Status report** (dashboard)

Big retention, Hard to misinterpret, Annotated

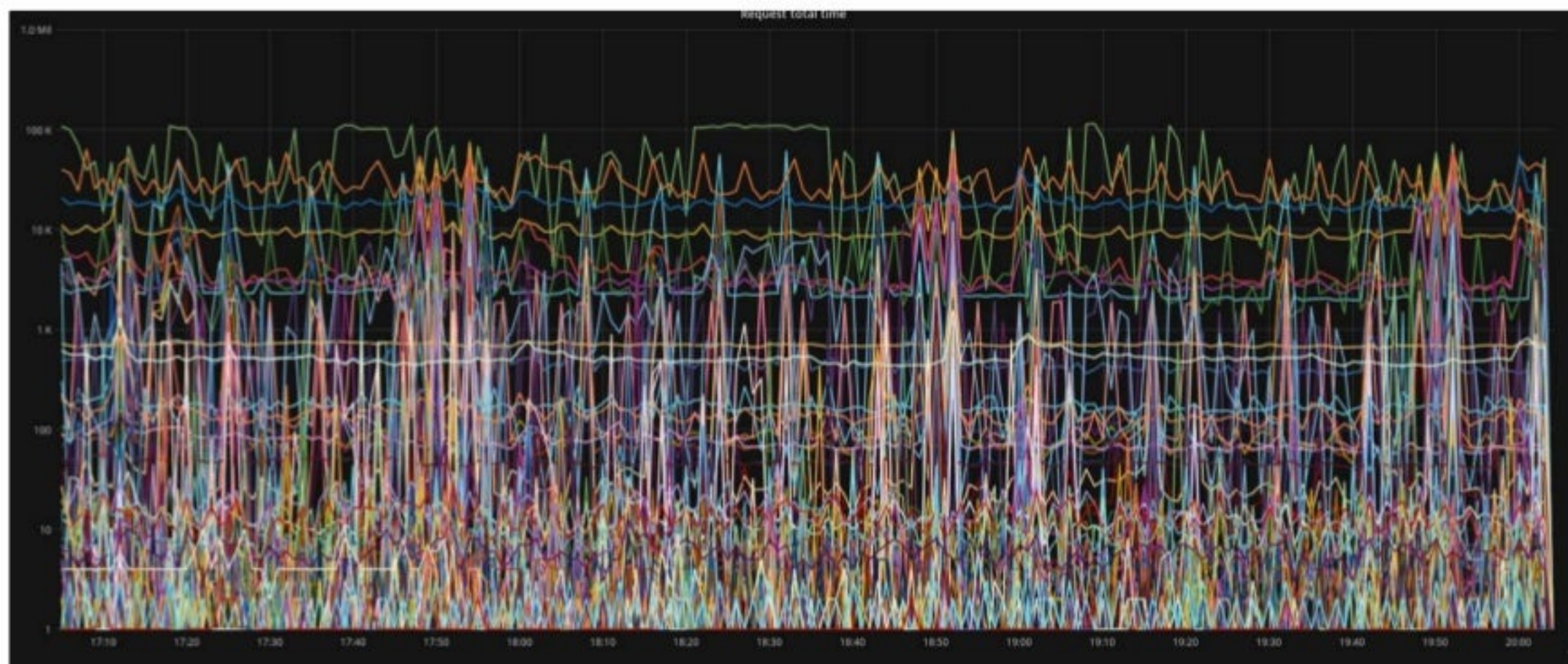
- **Exploration tool**

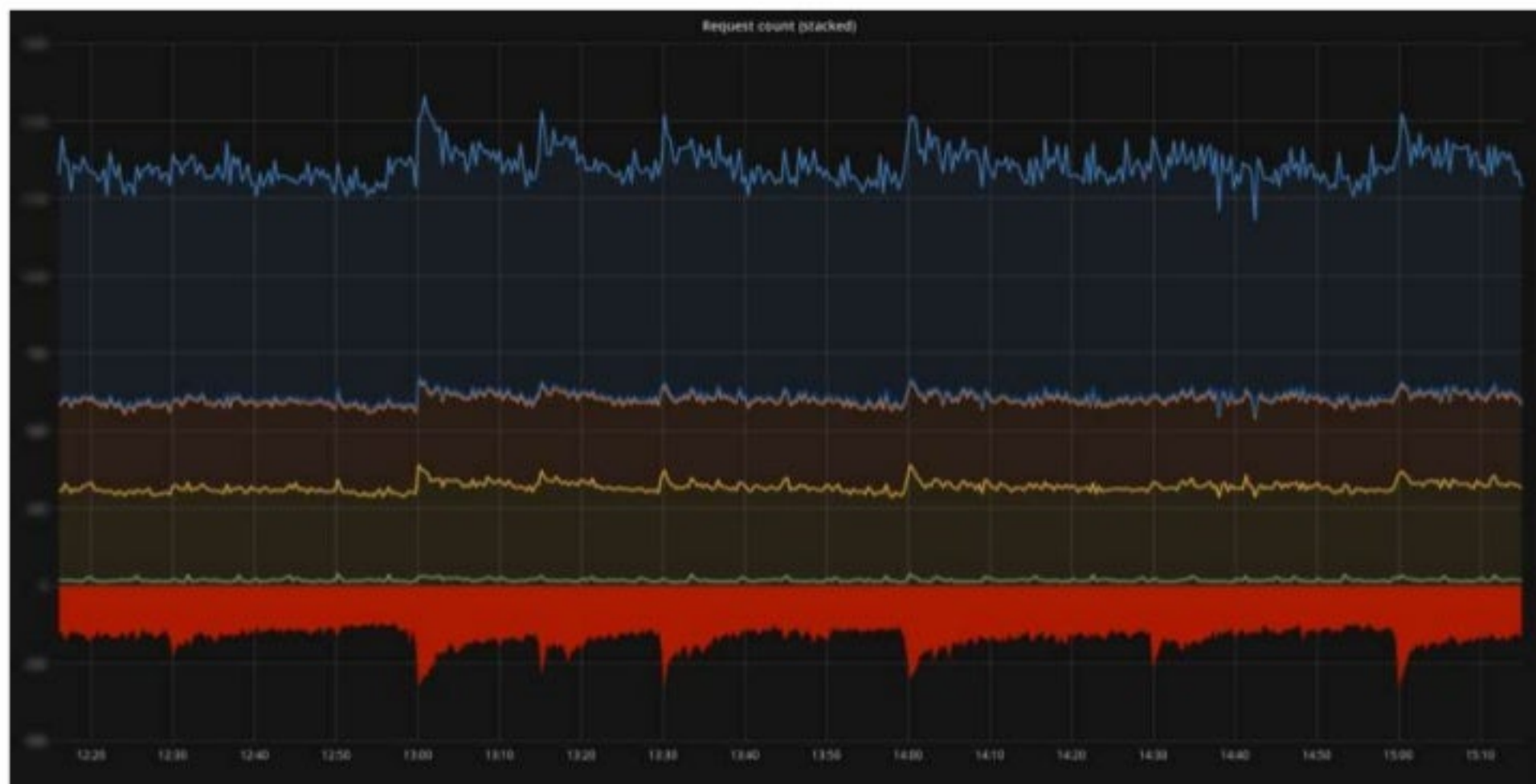
Plenty of historical data, Flexible

Example:

Status report & Investigation tool







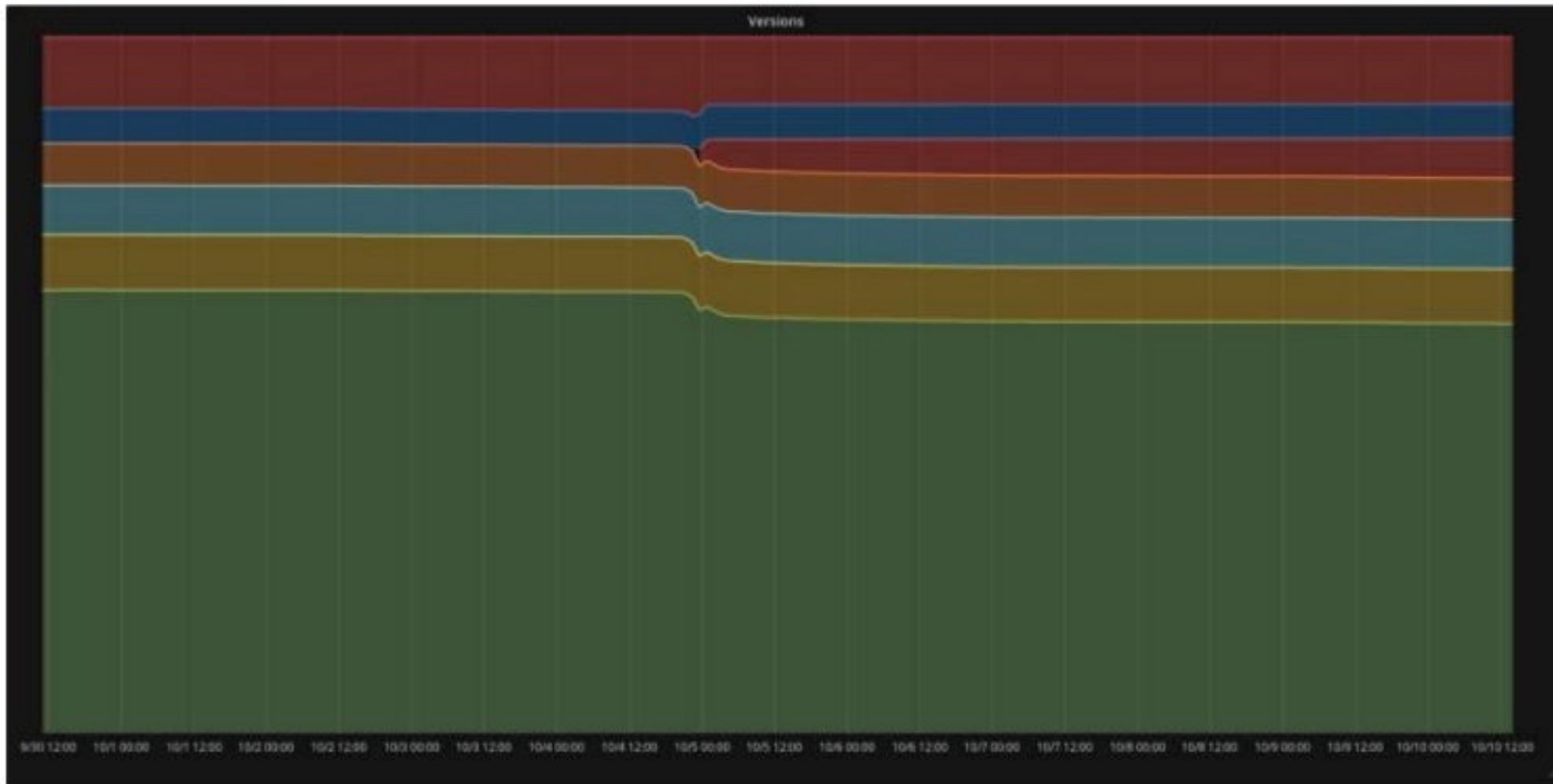
There are three kinds of lies: lies, damned lies and statistics.

- Mark Twain

Lies, damned lies and statistics



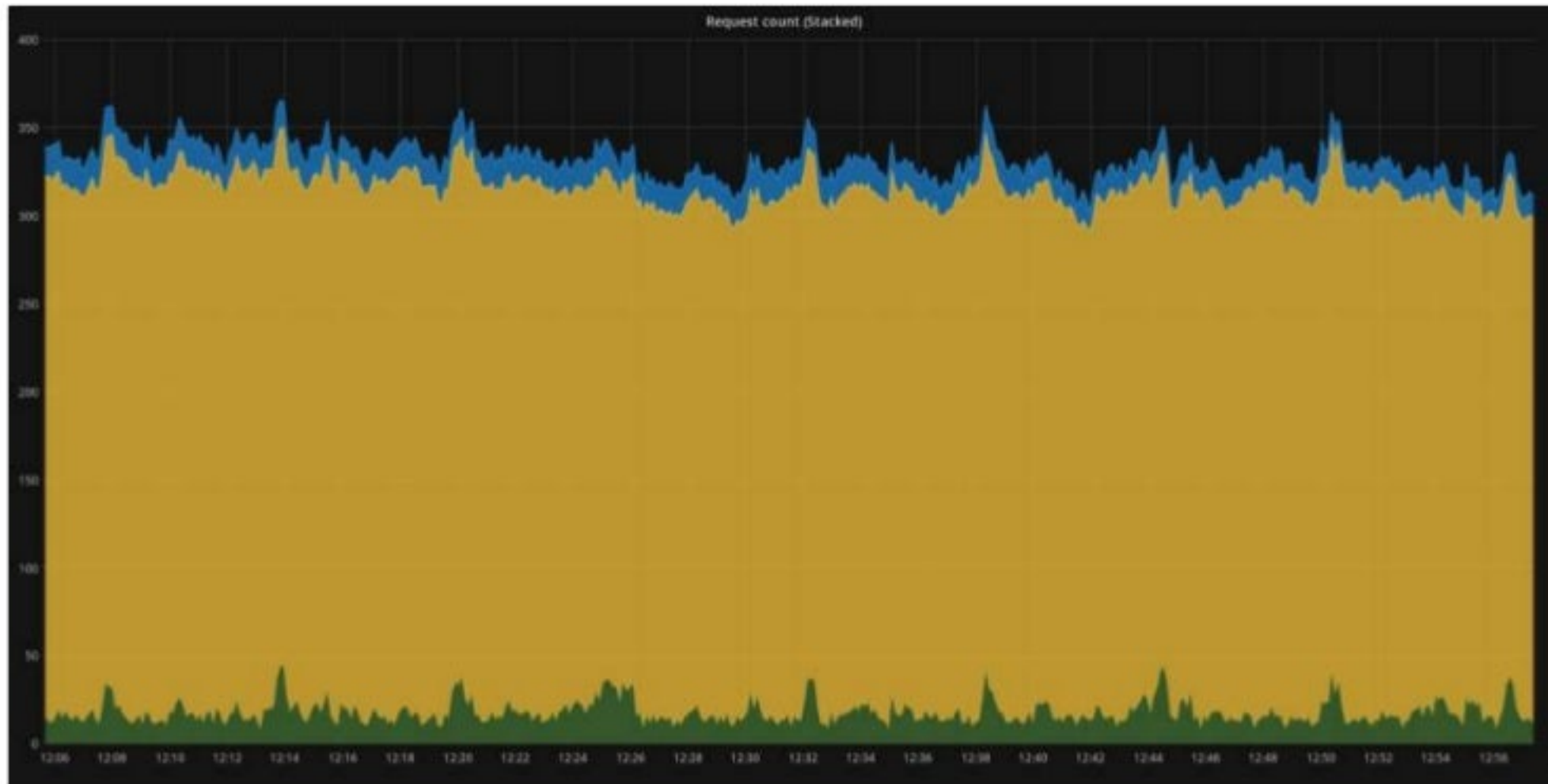
Stacked graph



Stacked graph



Stacked graph



Logarithmic scale



Non-zero y-min



Non-zero y-min



Exploration tool

How to analyze data?

- **Collect data**

logs, SQL, ansible, ...

- **Prepare data**

grep, awk, sed, SQL, ...

- **Visualize data**

Excel, SQL, R, Matlab/Octave, ...

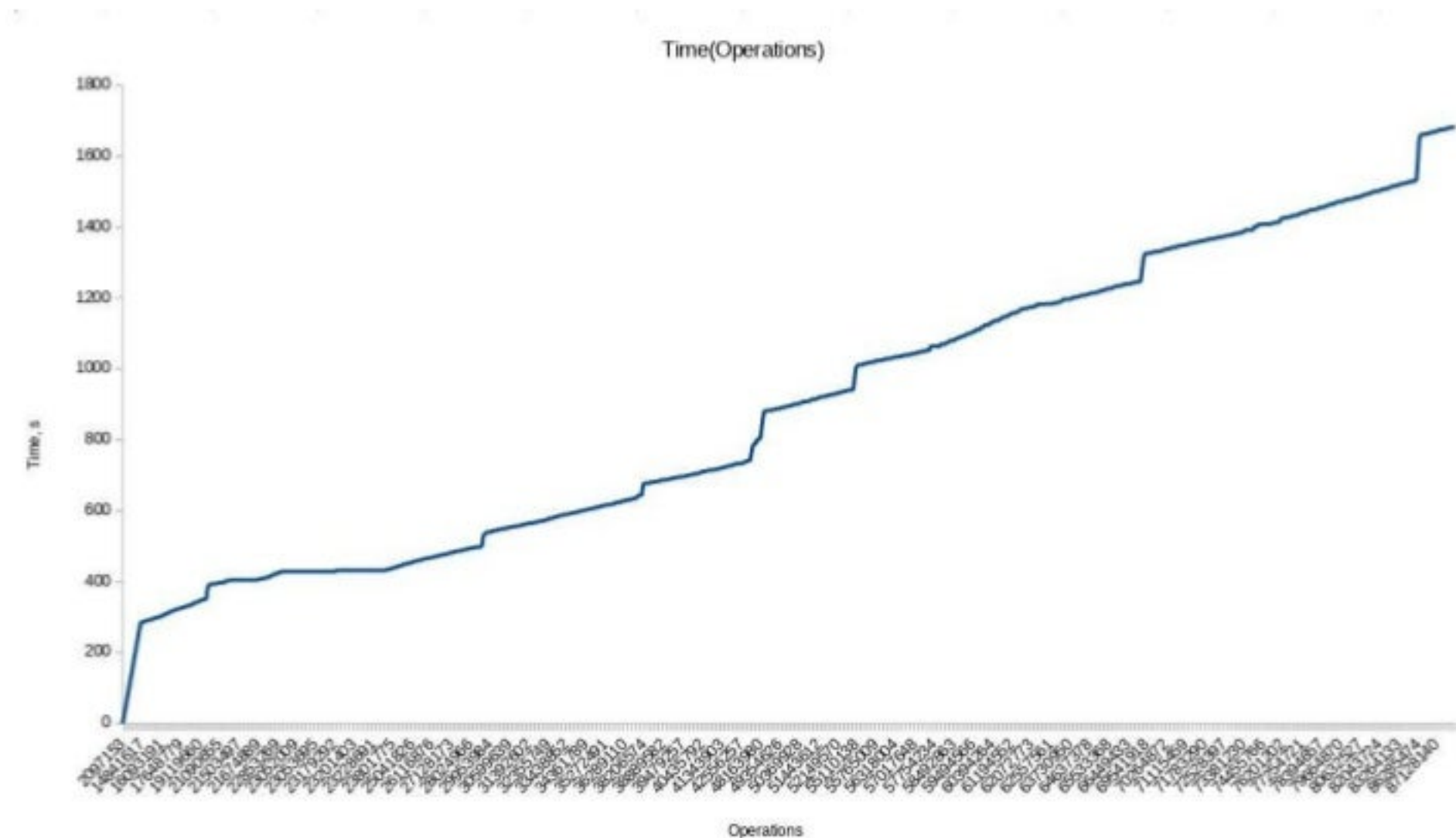
Visualization

```
[2017-10-10 18:26:09.971335] Processing chunk [1, 1048577)
[2017-10-10 18:26:27.909452] Processing chunk [1048577, 2097153)
[2017-10-10 18:26:48.278841] Processing chunk [2097153, 3145729)
[2017-10-10 18:27:08.590043] Processing chunk [3145729, 4194305)
[2017-10-10 18:27:28.872476] Processing chunk [4194305, 5242881)
[2017-10-10 18:27:49.330426] Processing chunk [5242881, 6291457)
[2017-10-10 18:28:09.565986] Processing chunk [6291457, 7340033)
[2017-10-10 18:28:29.925240] Processing chunk [7340033, 8388609)
```

... 1000 lines

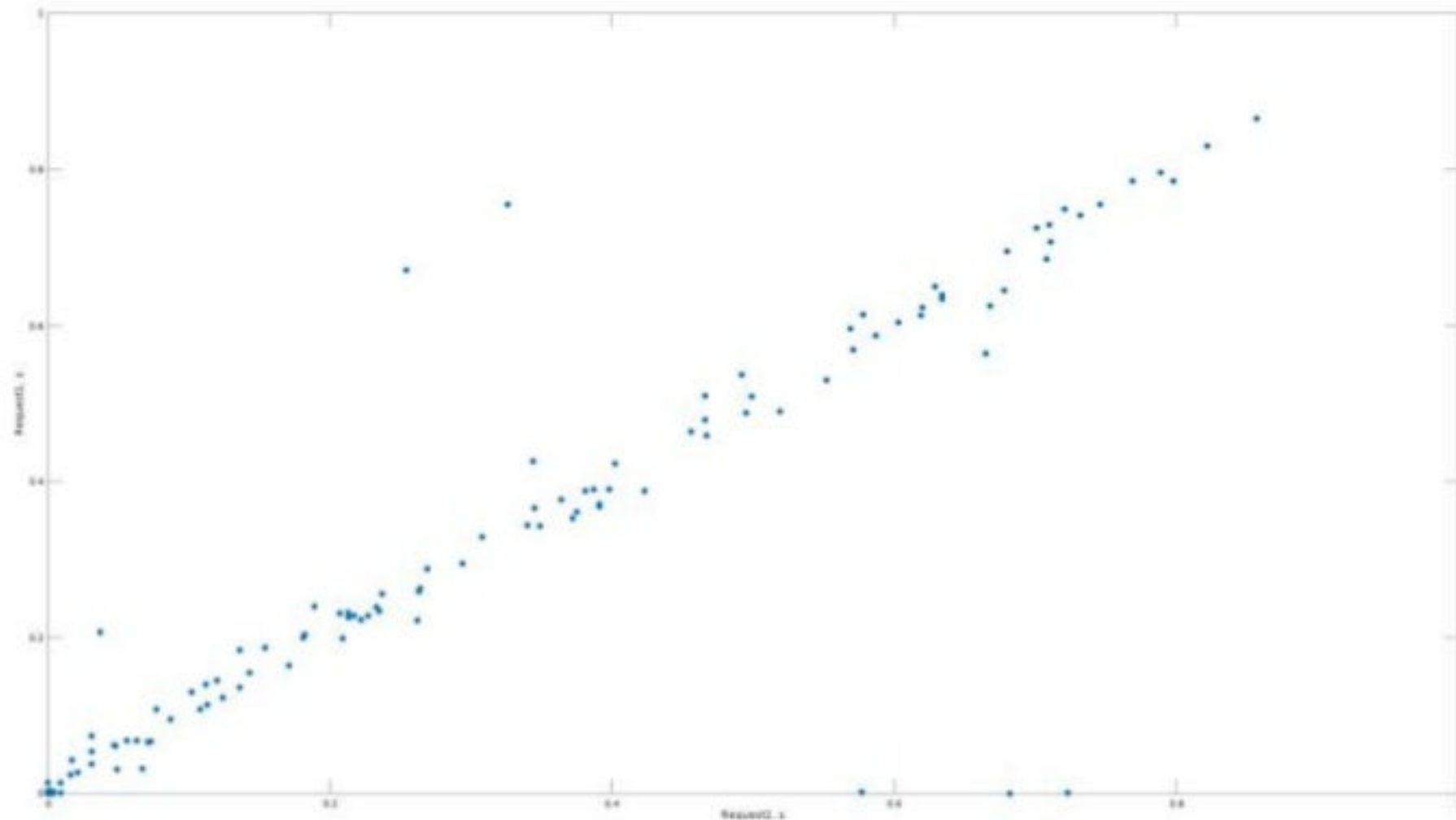
```
[2017-10-10 18:54:12.593898] Processing chunk [87576316, 87633909)
```

Restore DB from dump



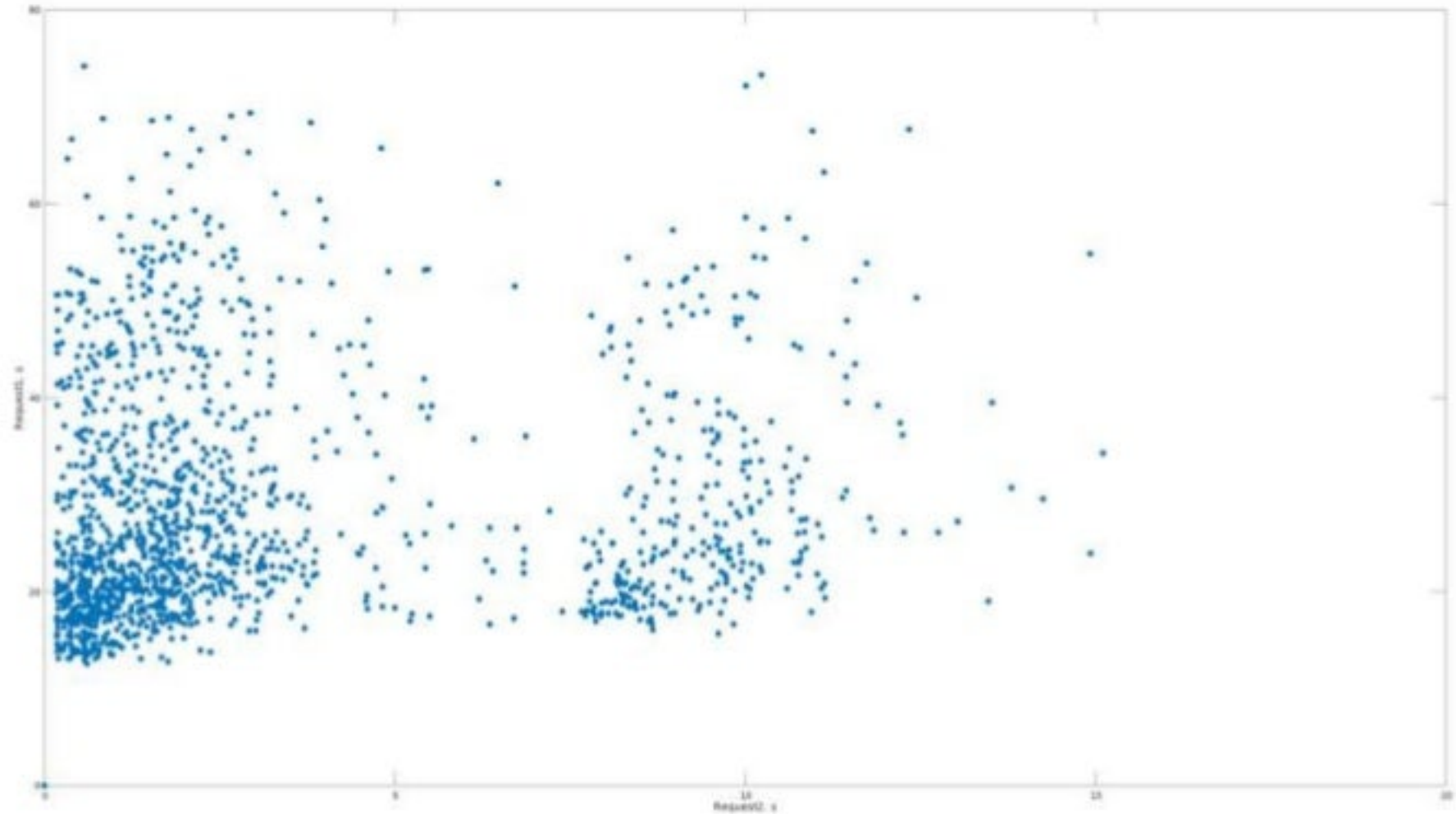
Entanglement detection

	Request1	Request2
1	0.00100	0.00100
2	0.00000	0.00100
3	0.00000	0.00100
4	0.00000	0.00000
5	0.39100	0.36800
6	0.00000	0.00100
7	0.00000	0.00100
8	0.00000	0.00000
9	0.00000	0.00100
10	0.00000	0.00000
11	0.00000	0.00000
12	0.00000	0.00000
13	0.00000	0.00100
14	0.00100	0.00100
15	0.00000	0.00000
16	0.00000	0.00000
17	0.00000	0.00000
18	0.00000	0.00100
...



Entanglement detection

	Request1	Request2
1	3.583	23.174
2	4.398	40.414
3	1.541	54.097
4	0.401	23.830
5	0.971	18.446
6	0.938	23.164
7	9.645	17.802
8	0.959	39.820
9	10.864	27.506
10	8.992	17.809
11	0.174	50.634
12	1.162	31.359
13	3.520	29.936
14	9.758	24.268
15	1.212	28.426
16	8.579	29.744
17	7.819	17.952
18	3.045	22.624
...




```
data = load("~/data/requests.csv");

[_ n] = size(data);

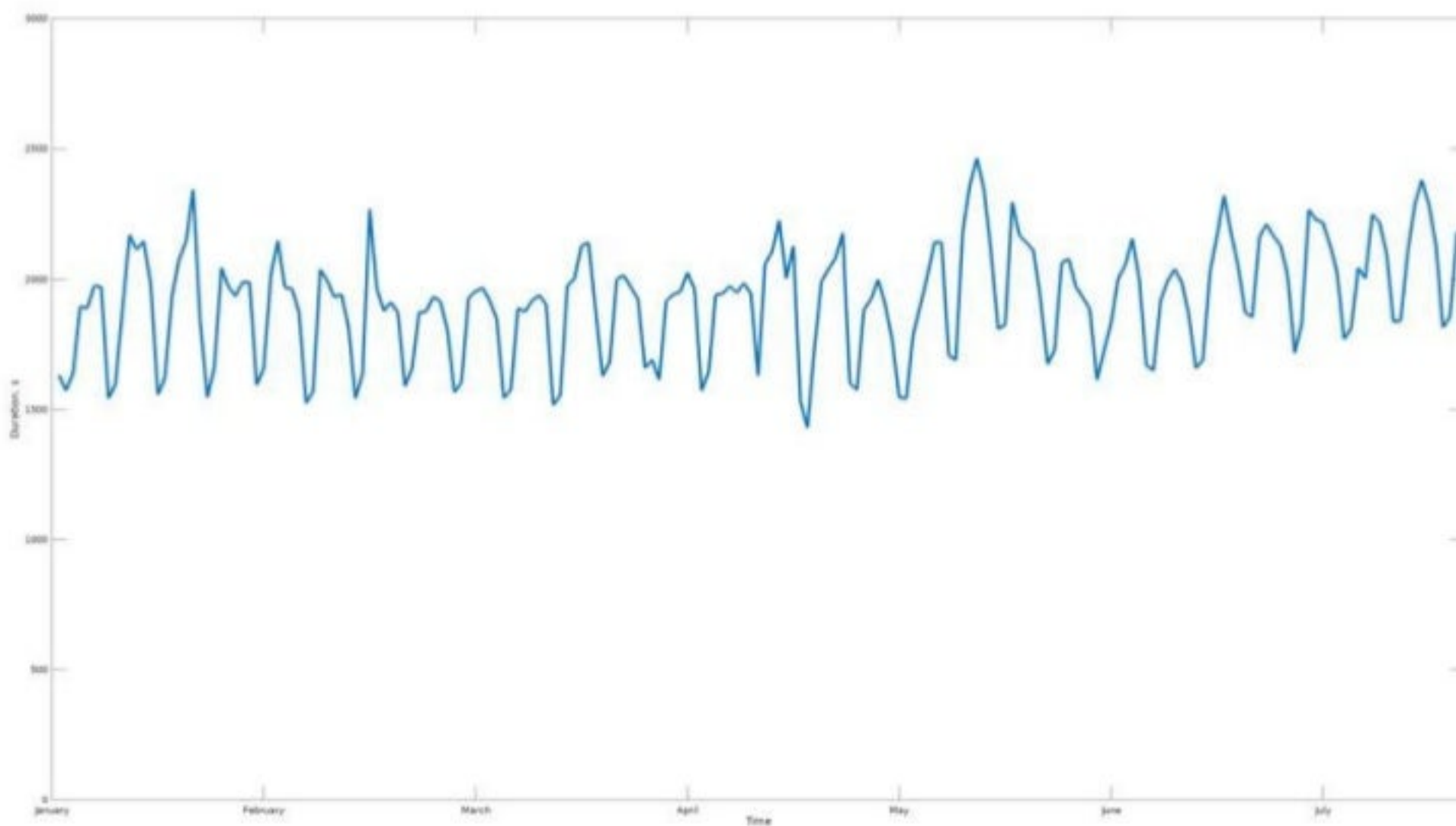
for i = 1:n
    for j = i + 1:n

        c = corr(data(:, [i j]))(1, 2);

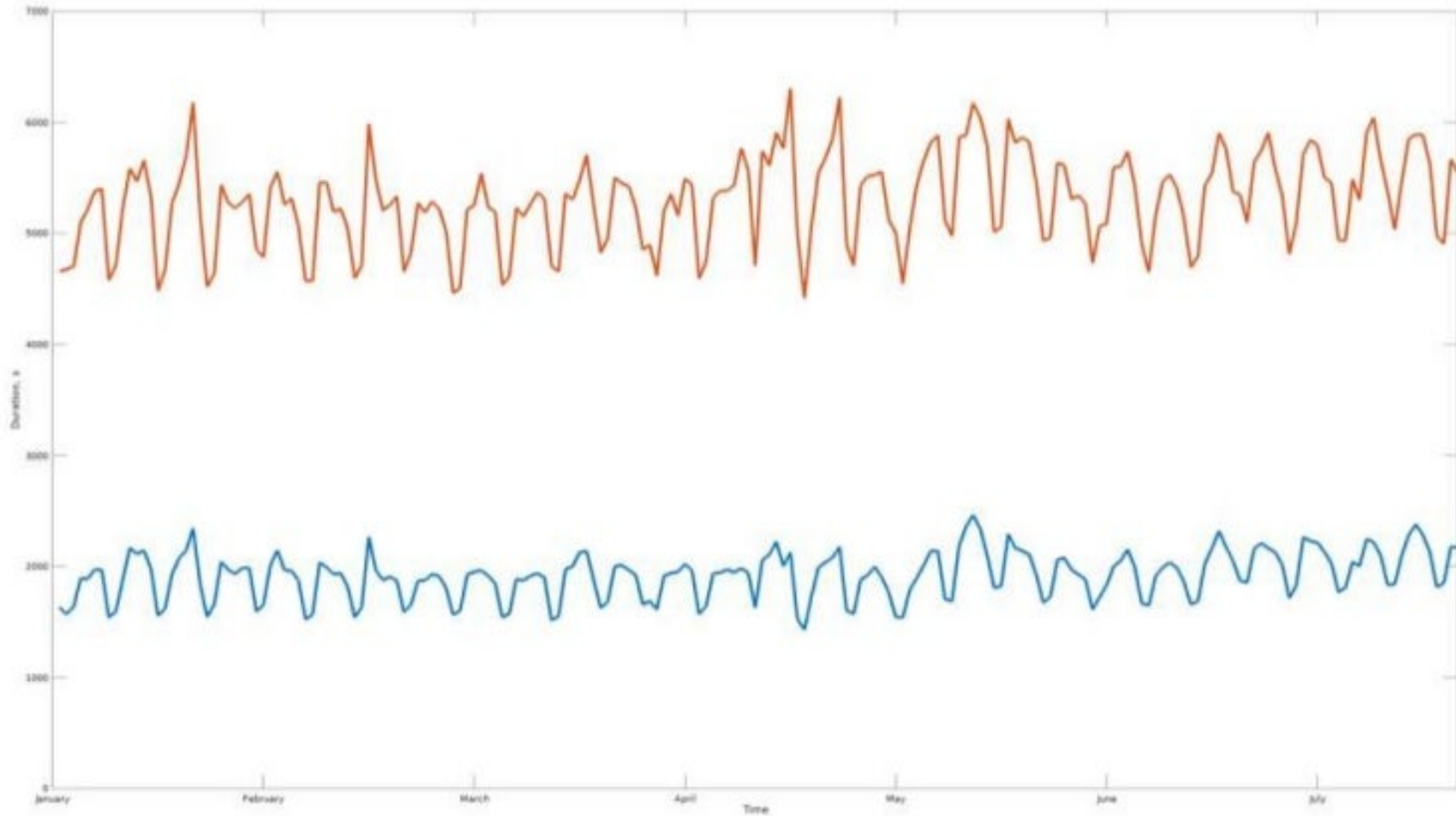
        if c > 0.9
            fprintf('request%d request%d correlation: %f\n', i, j, c);
            plot(data(:, i), data(:, j), "marker", "*", "linestyle", "none")
            pause();
        end

    end
end
```

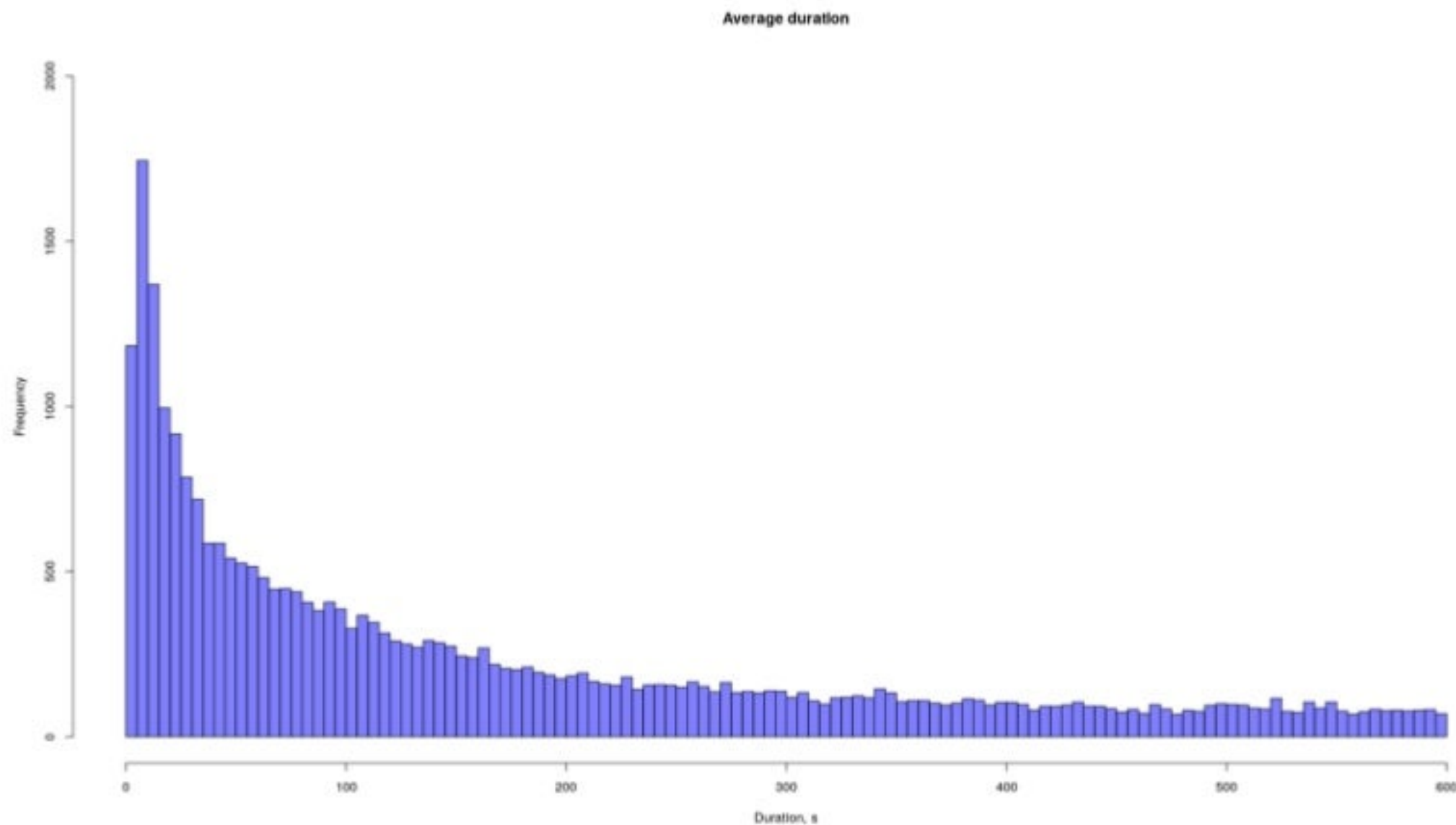
Backup duration: average



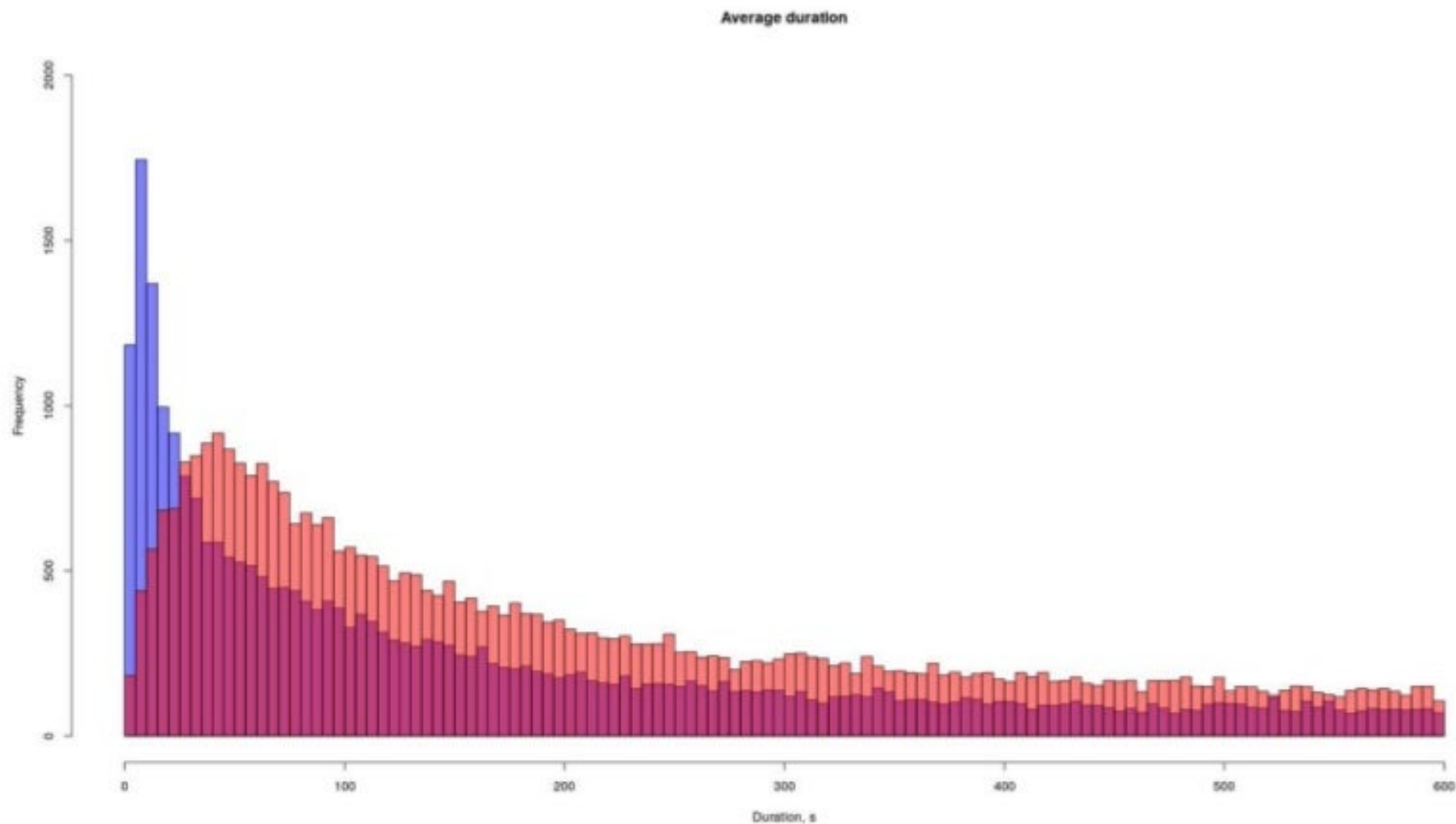
Backup duration: average + std



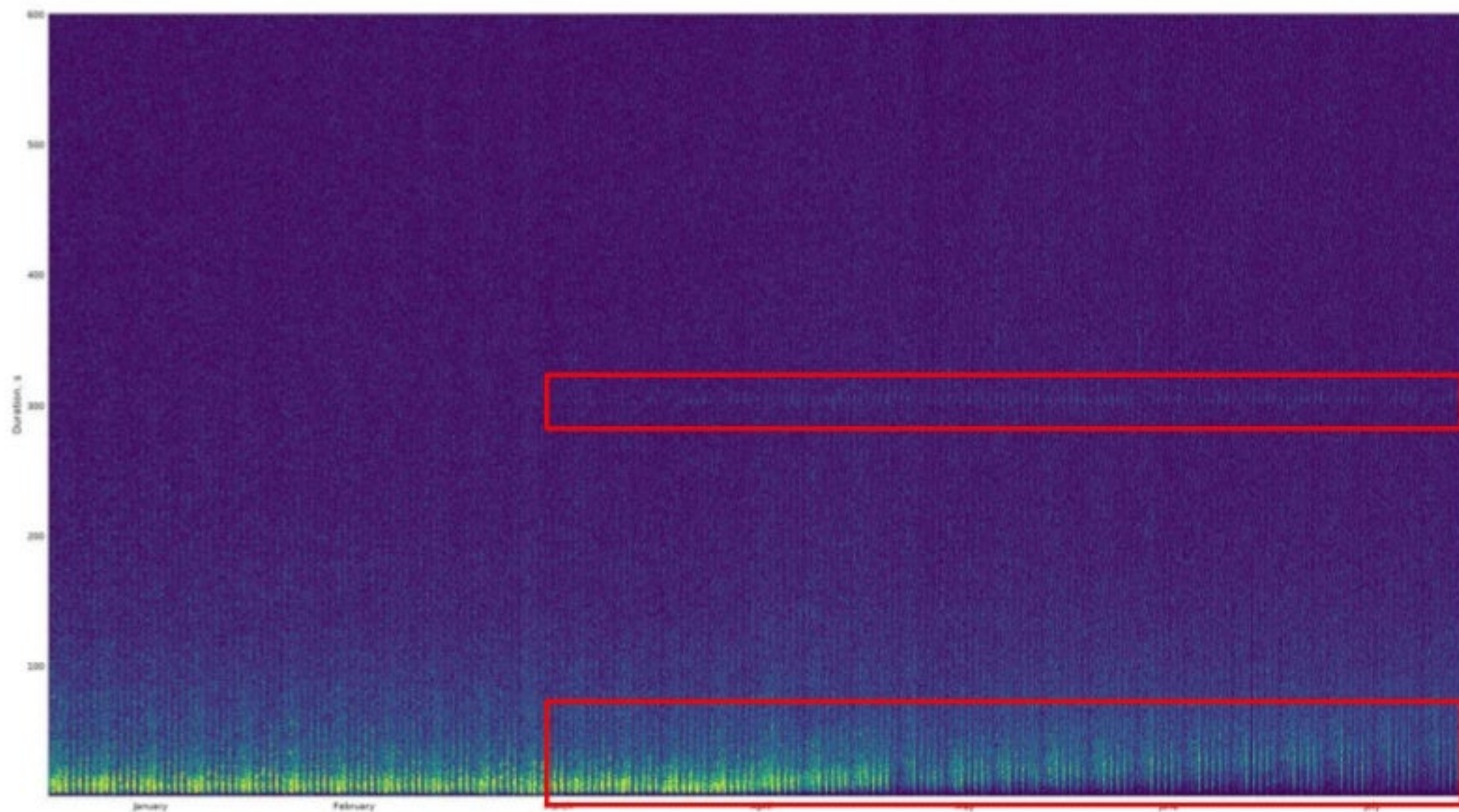
Backup duration: histogram



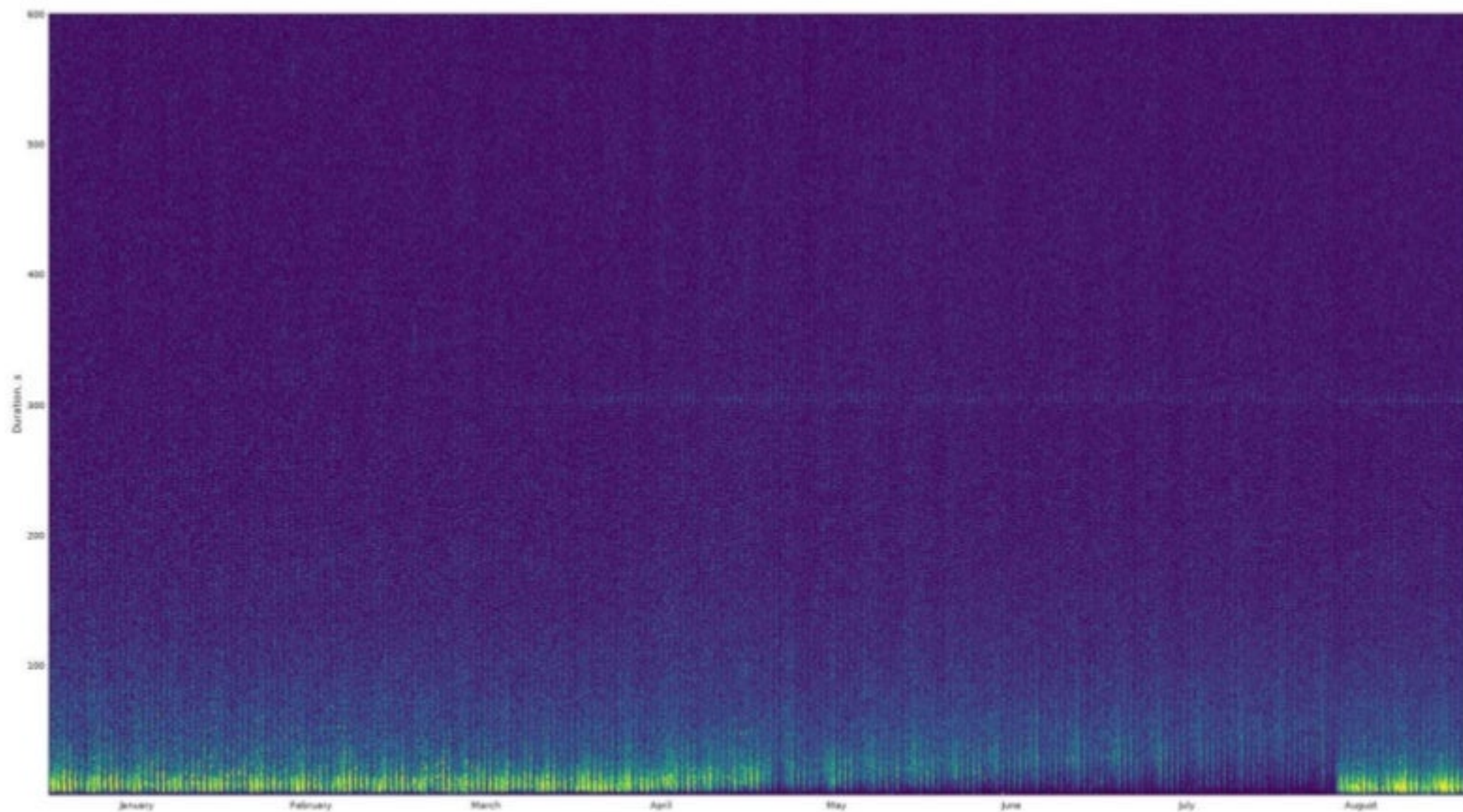
Backup duration: histogram



Backup duration: heatmap



Backup duration: heatmap



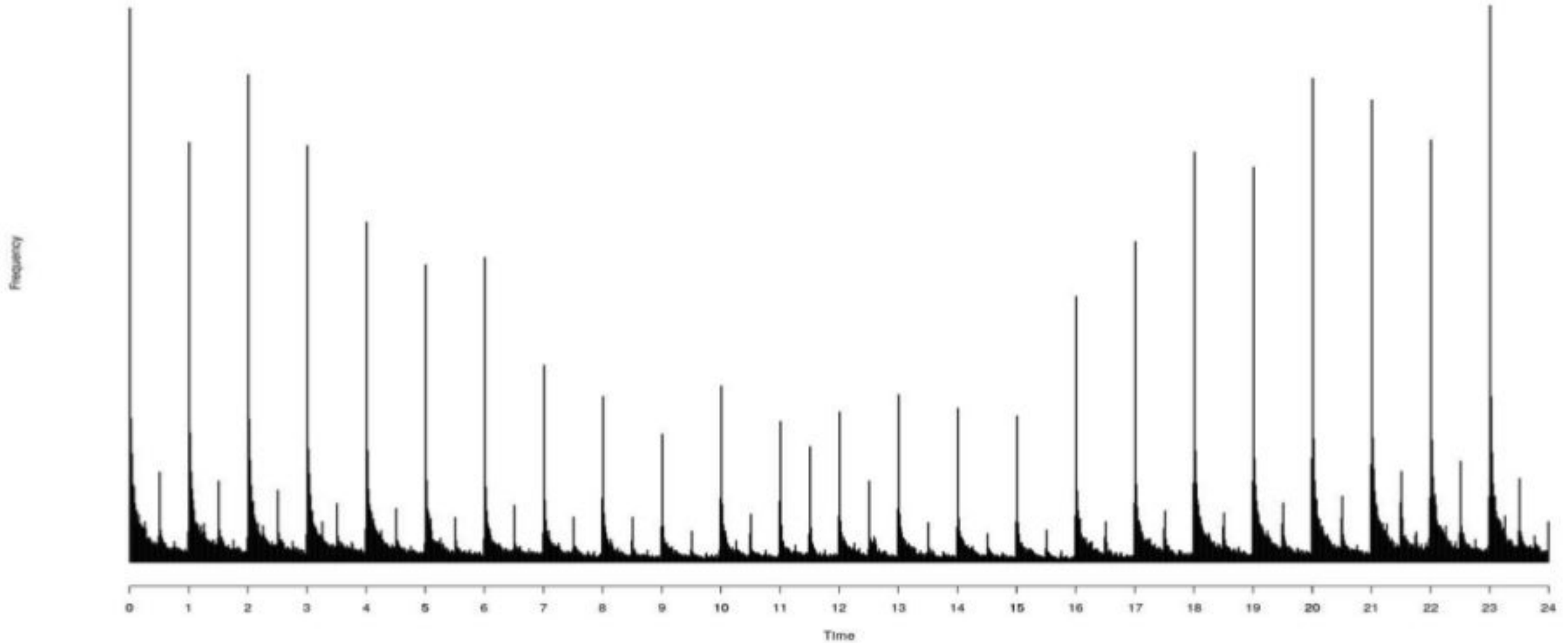
Lies, damned lies and statistics

It ain't what you don't know that gets you into trouble.
It's what you know for sure that just ain't so.

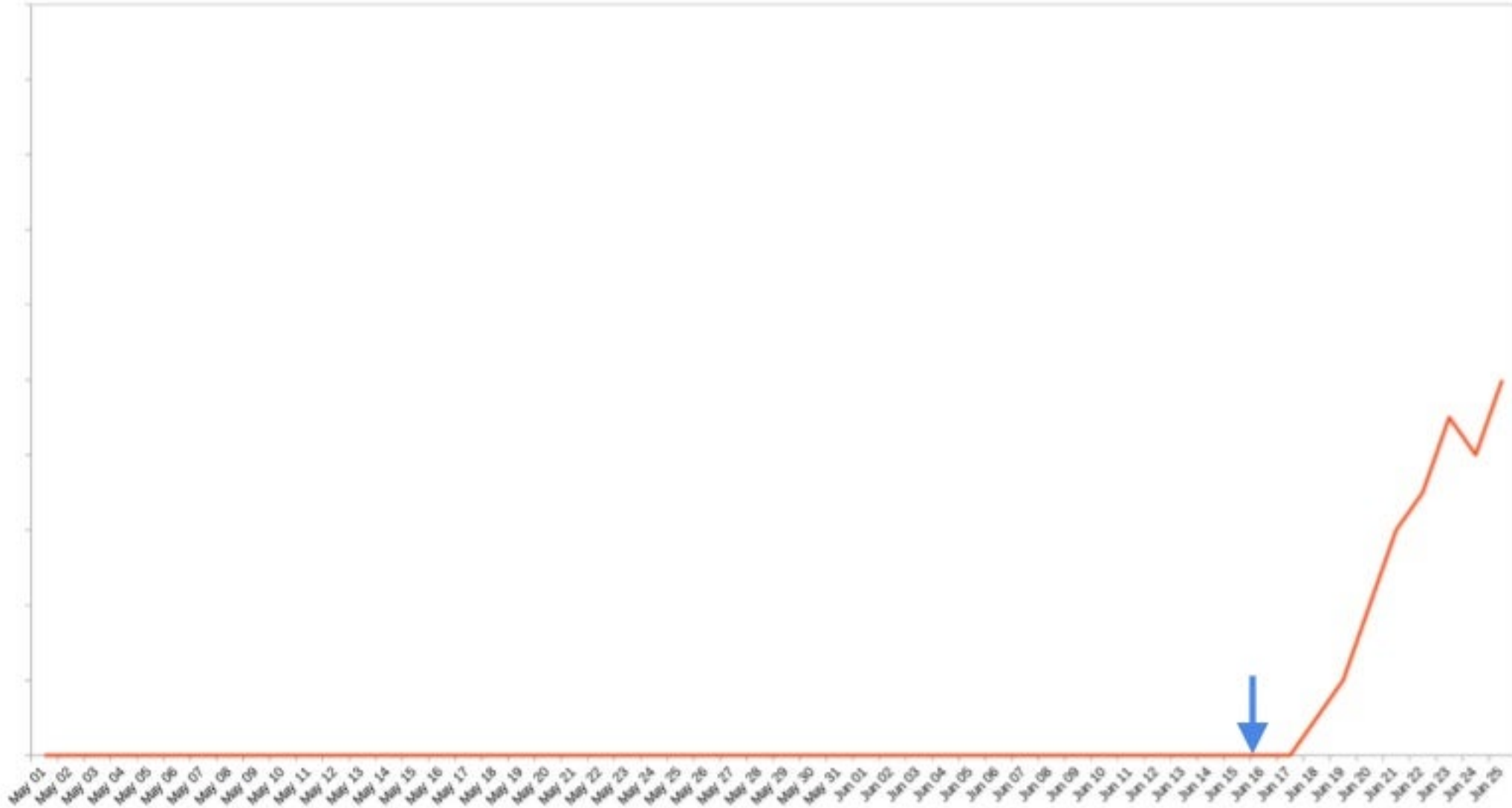
– Mark Twain

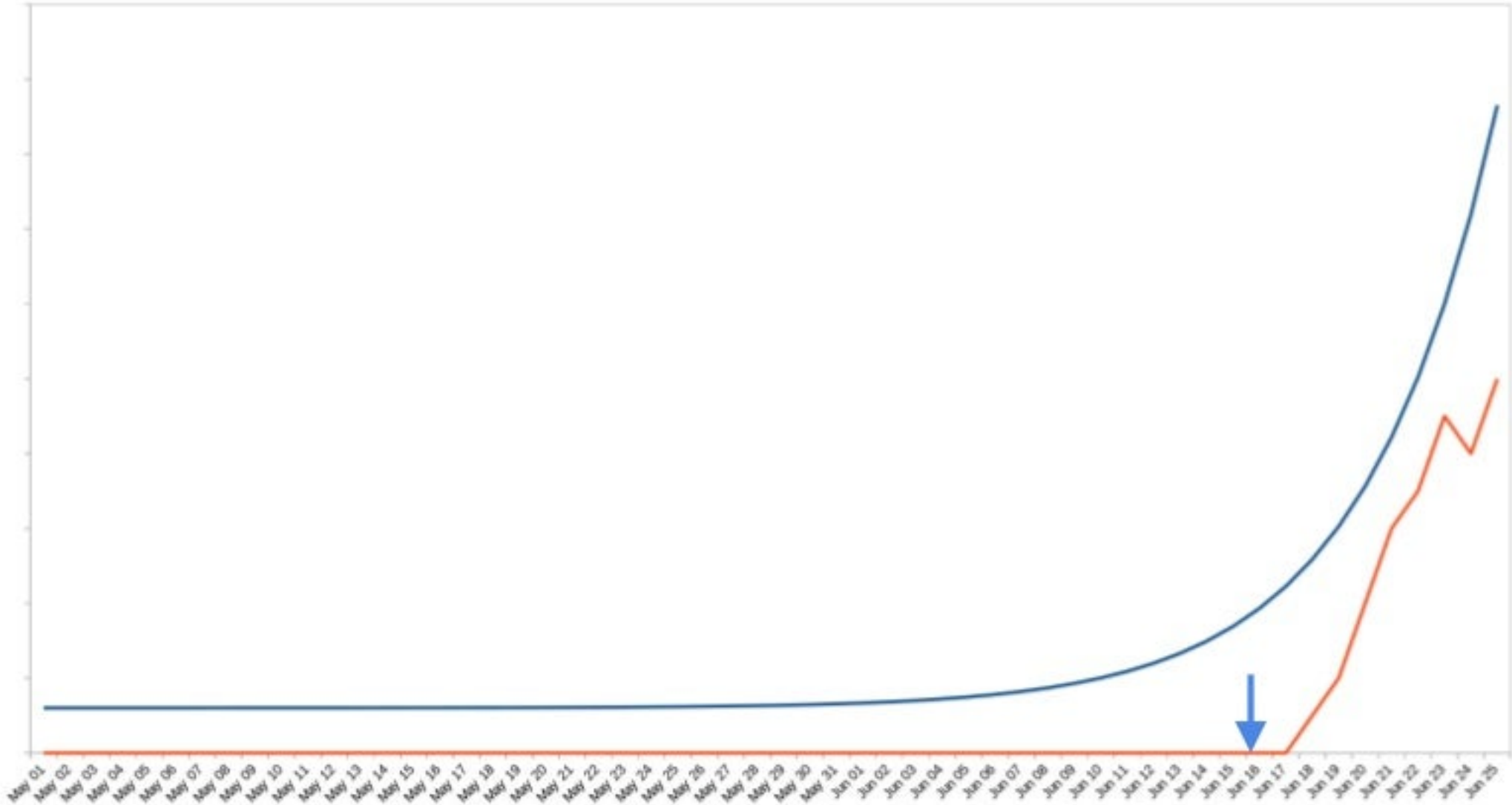
- Miss existing problem
- Find unexisting problem

Backup start time









- Grafana <https://github.com/grafana/grafana>
- Prometheus <https://github.com/prometheus/prometheus>
- Graphite <https://github.com/graphite-project>
- R <https://www.r-project.org/>
- Octave <https://www.gnu.org/software/octave/>

?

yury.efimochev@solarwinds.com
efimyury@gmail.com