In Graphic Detail

Data Science in Application Performance Management

Bill Kayser

- Founding engineer at New Relic
- Software architect specializing in web applications
- Working in APM for 10 years
- Data Science, Machine Learning, and Visualization
- Helped start New Relic's "Autism at Work" Program

Overview

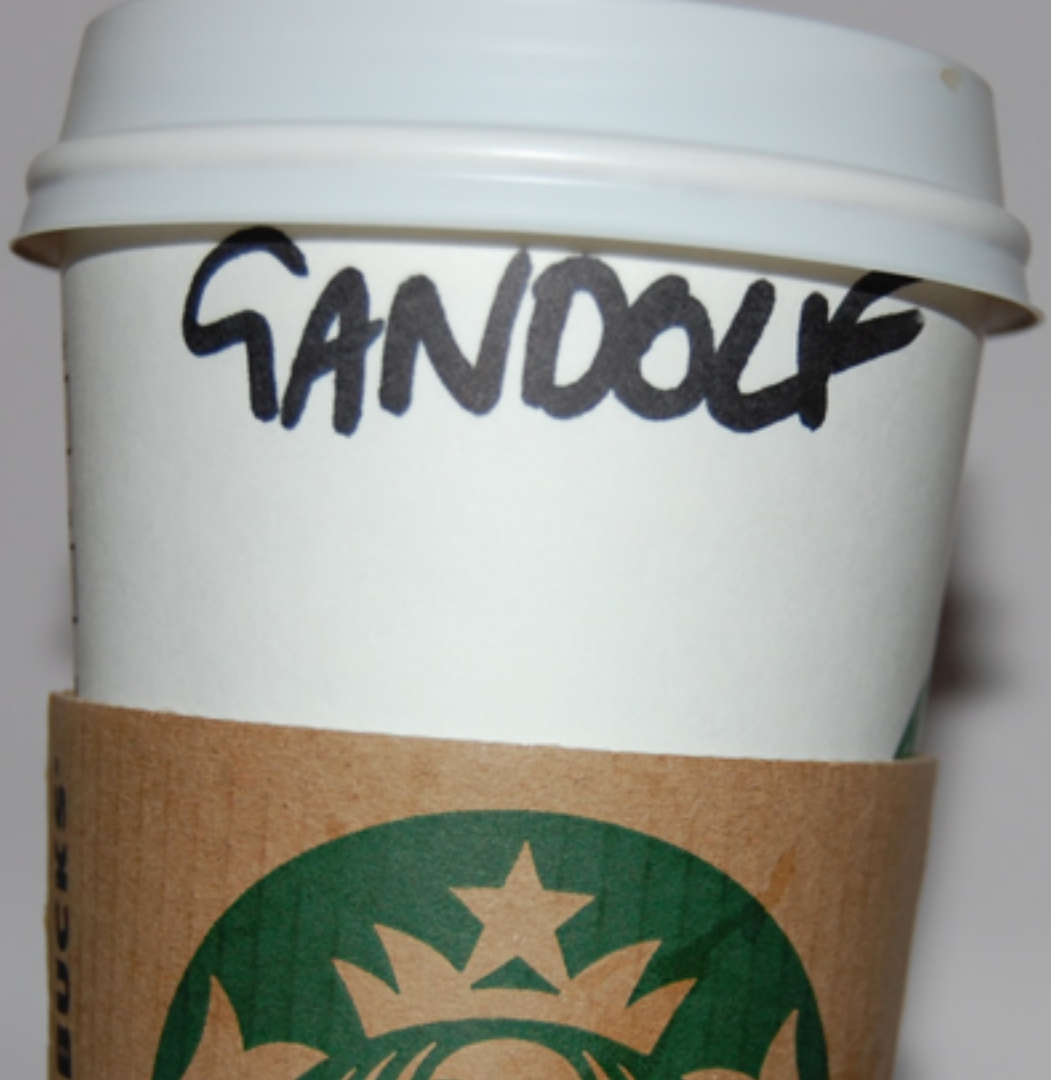
- Context: Understanding how our web applications are performing
- How we decide what data to collect and what visualizations to use
- Evaluate simple summary statistics
- Histograms, response time distributions and the geometric mean
- Understanding the Apdex measure for latency
- Using alternative visualizations


STARBUCKS COFFEE
STARBUCKS COFFEE


MAD-AH-LYNN
STARBUCKS
Madeline


KRISH
Kris
IRISH
Iris


Justin
Bieber
CM


GANDOLF


STARBUCKS
COFFEE


Felix
STARBUCKS
COFFEE

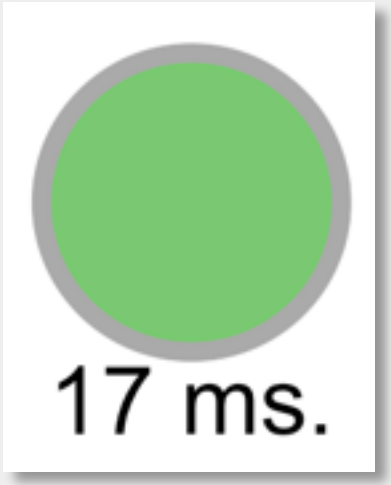
Example: Microservice for Stock Quotes

Congratulations on Your Launch!

How are we doing?

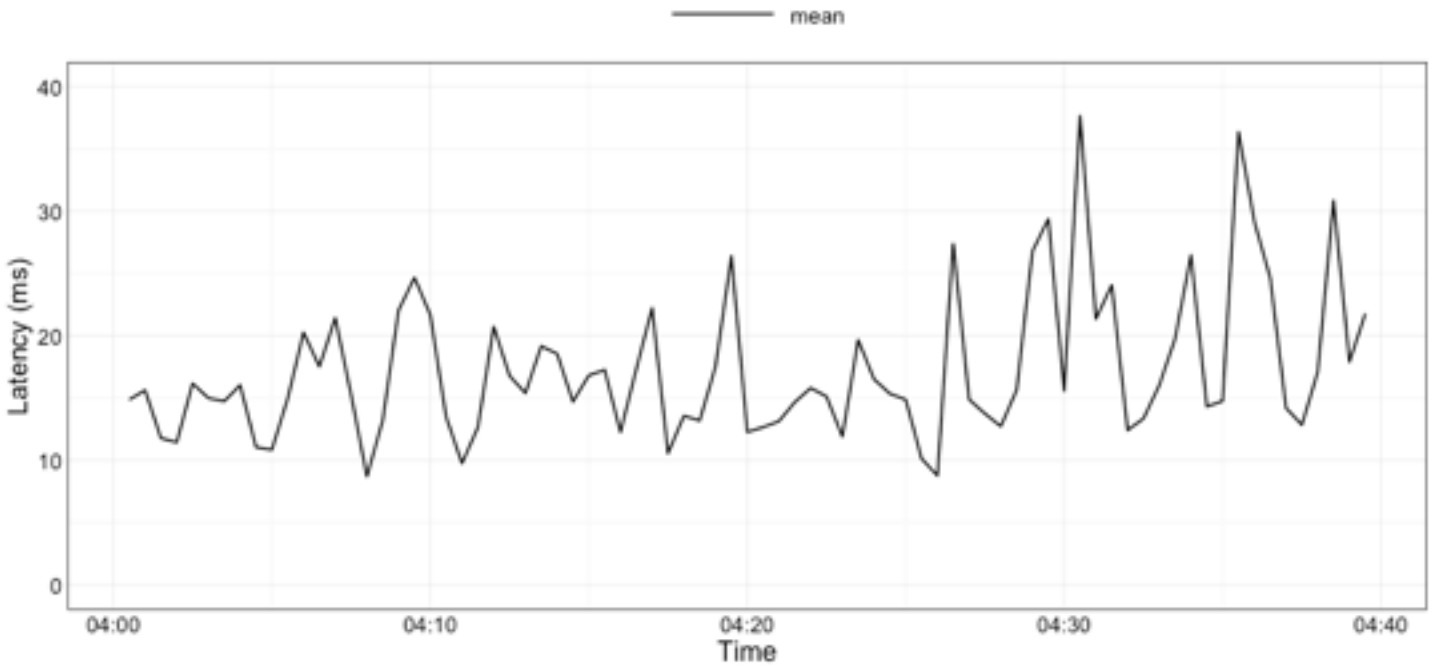
17 ms.

All good?

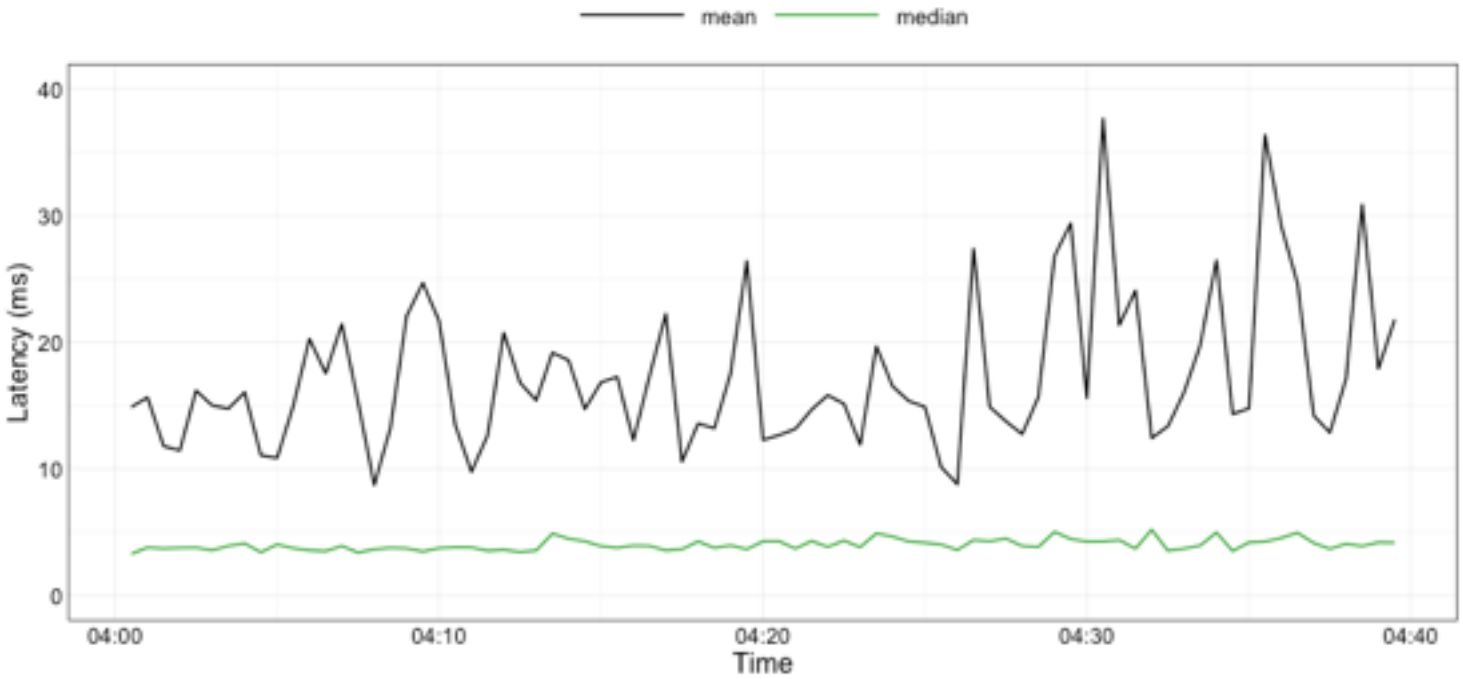


Better??

Average Response Time

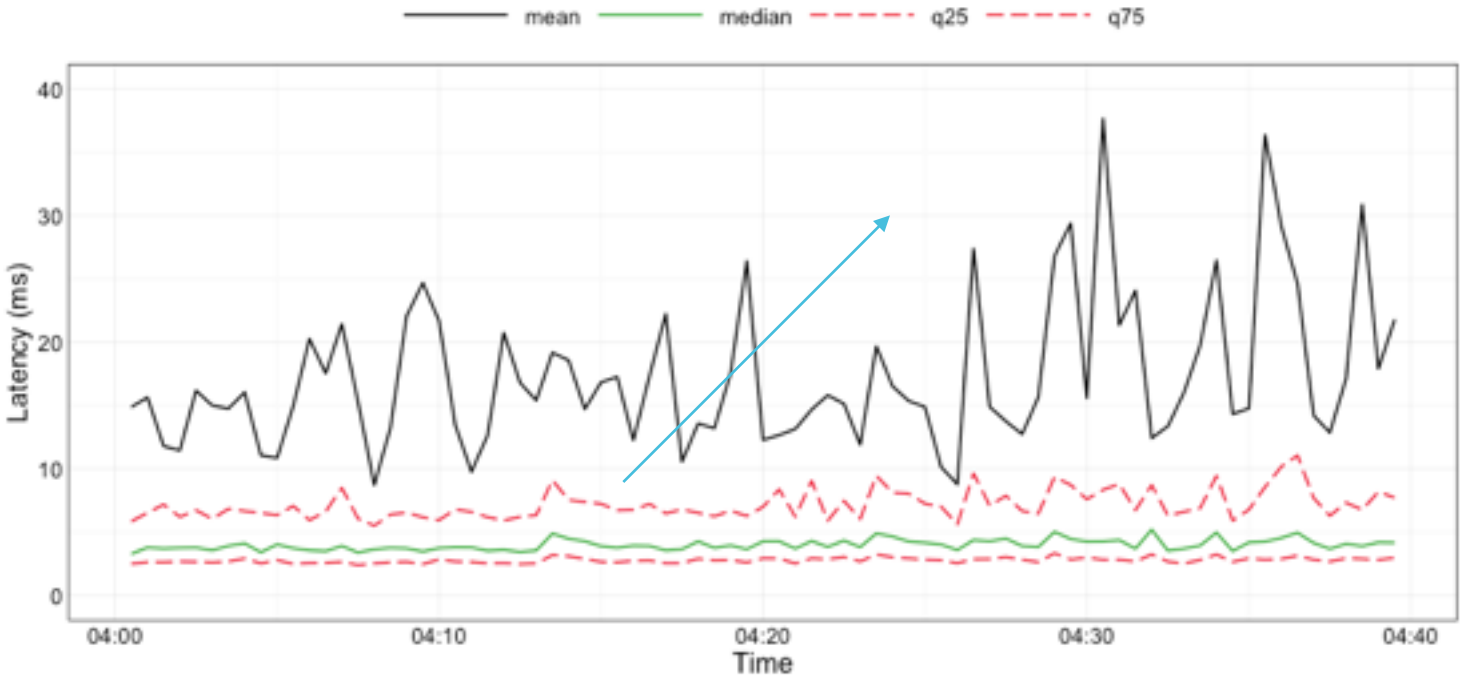


Median Response Time



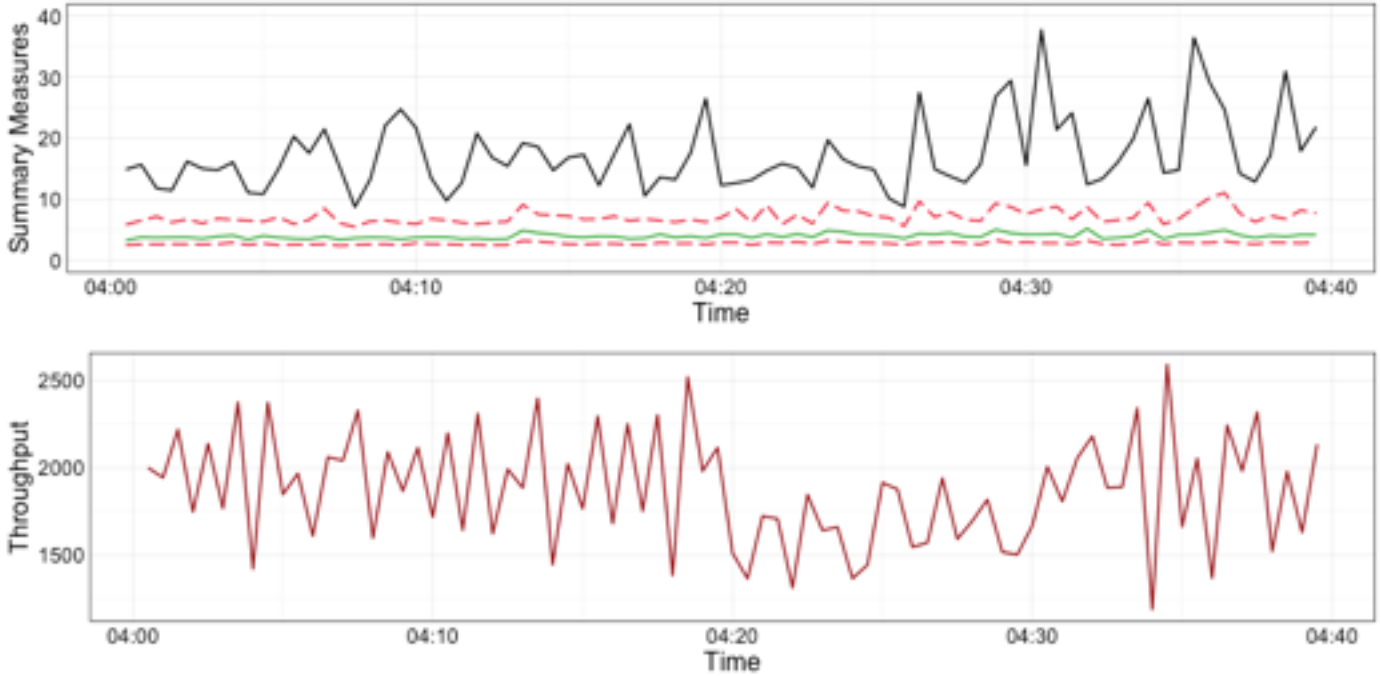
New Relic.

25% and 75% Bands (Quartiles)



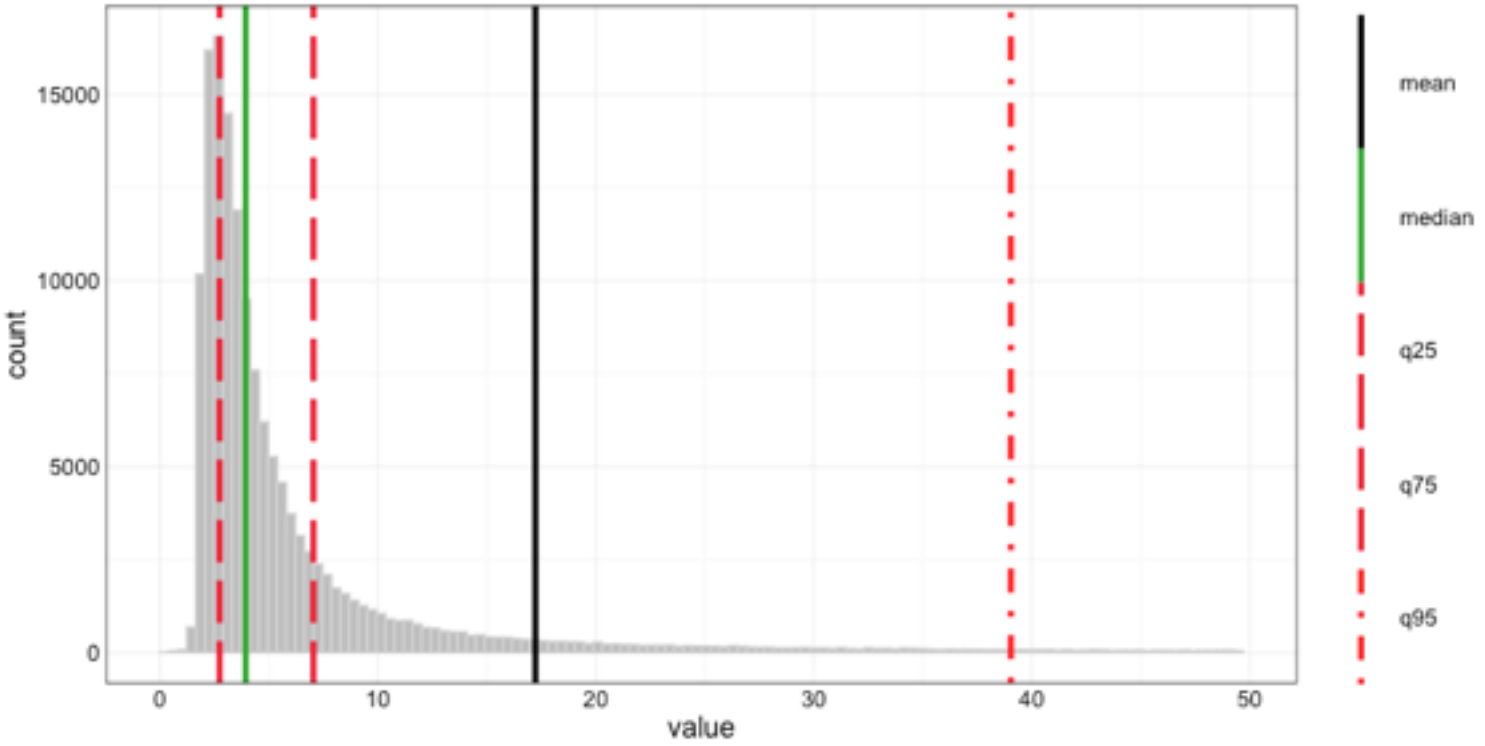
New Relic.

Throughput

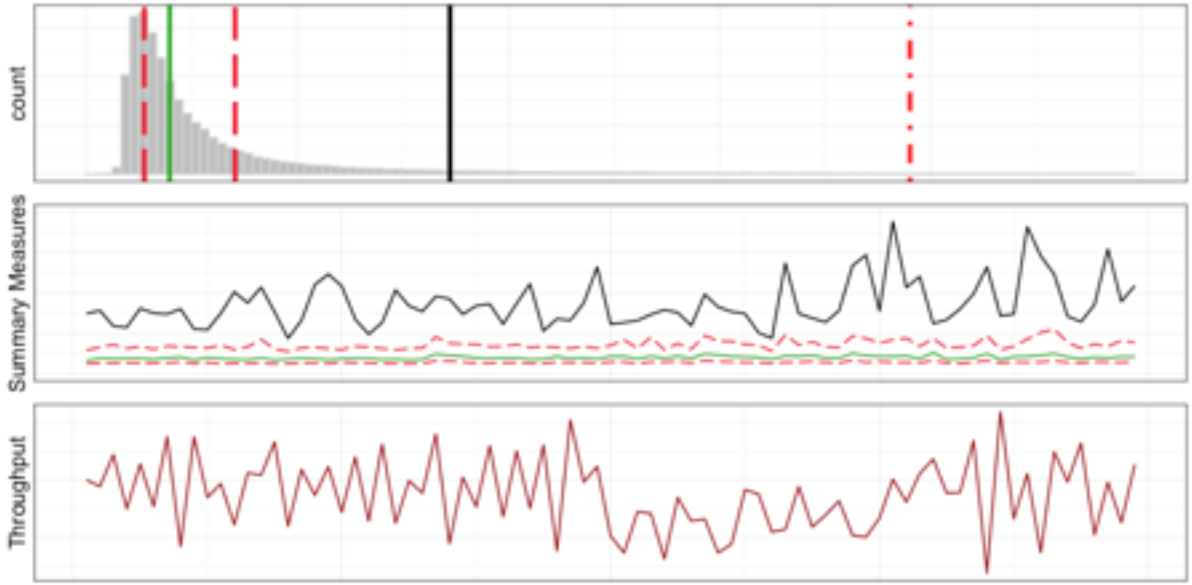


New Relic.

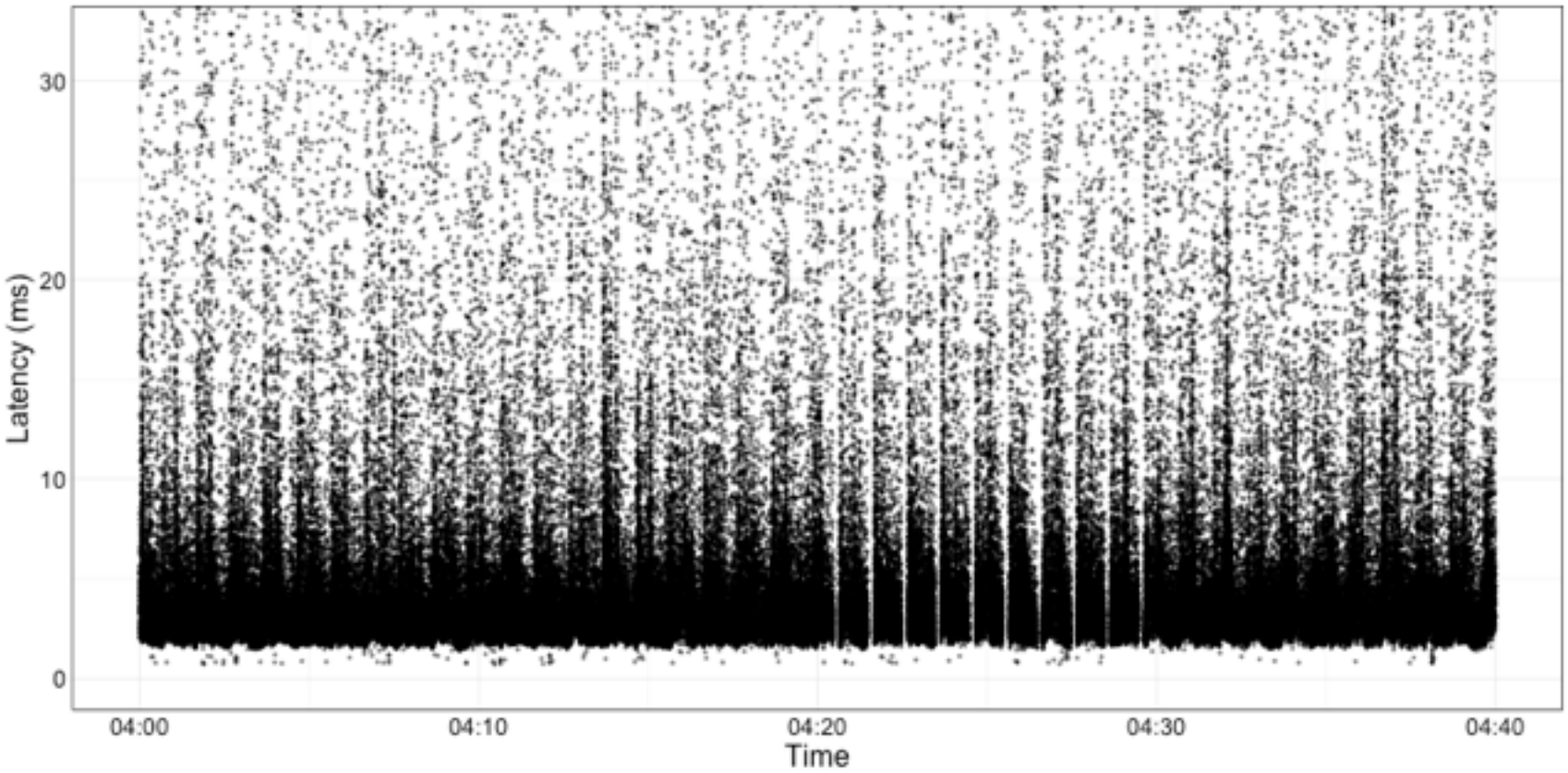
Histograms

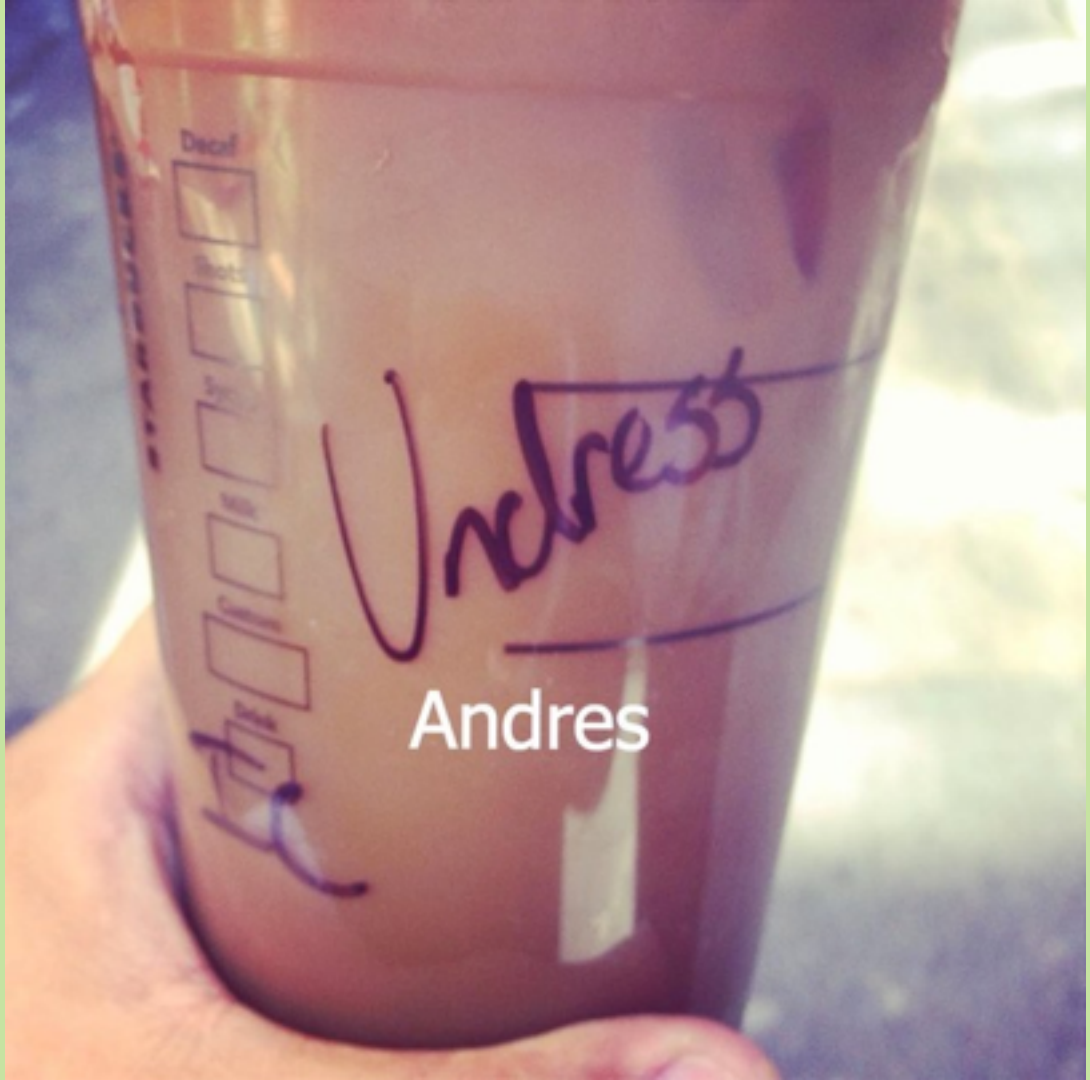


Combining Histograms with Timeseries

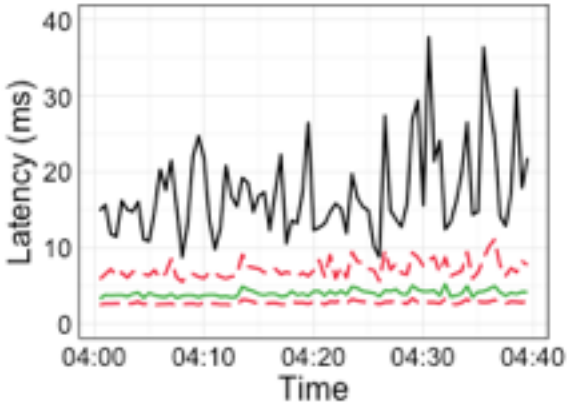


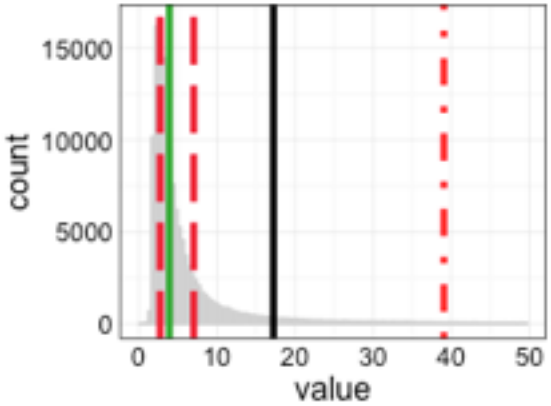
Got it all figured out?

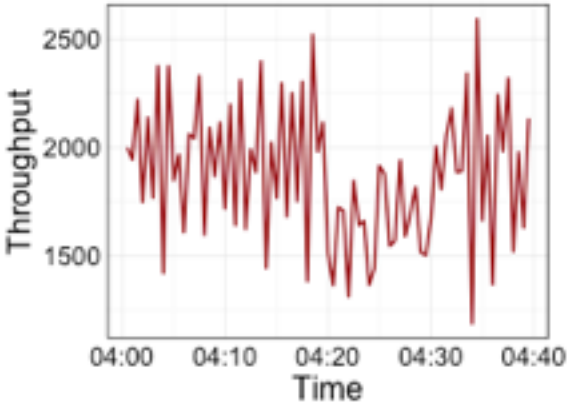


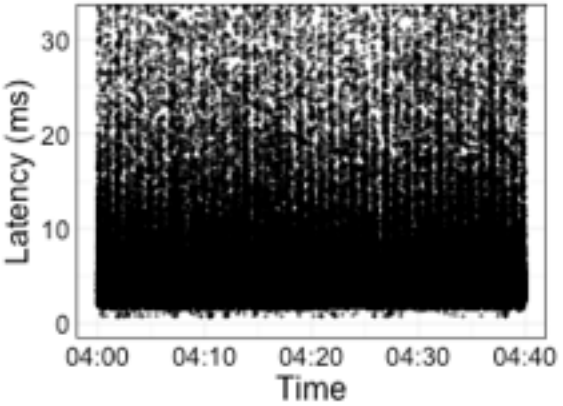

Andres

17 ms.









New Relic.

What Metrics Do We Care About?

Estimating the Value of a Metric

- Does it quantify something we care about?
- Does it give us a qualitative assessment of the current status?
- Does it help us measure differences across similar things, like applications or servers?
- Does it help us identify trends and anomalies?
- Does it reveal underlying patterns or relationships in the data?

Understanding the Cost of a Metric

- How cheaply can we collect the data?
  - Is space or bandwidth a premium?
  - Can it be reduced by combining values into a single value, or do we have to store every measurement?
  - How much screen real estate does it take up?
- The smaller the data the more things you can measure.
- The less space it takes the more options you have for screen layout.

Reducing Data

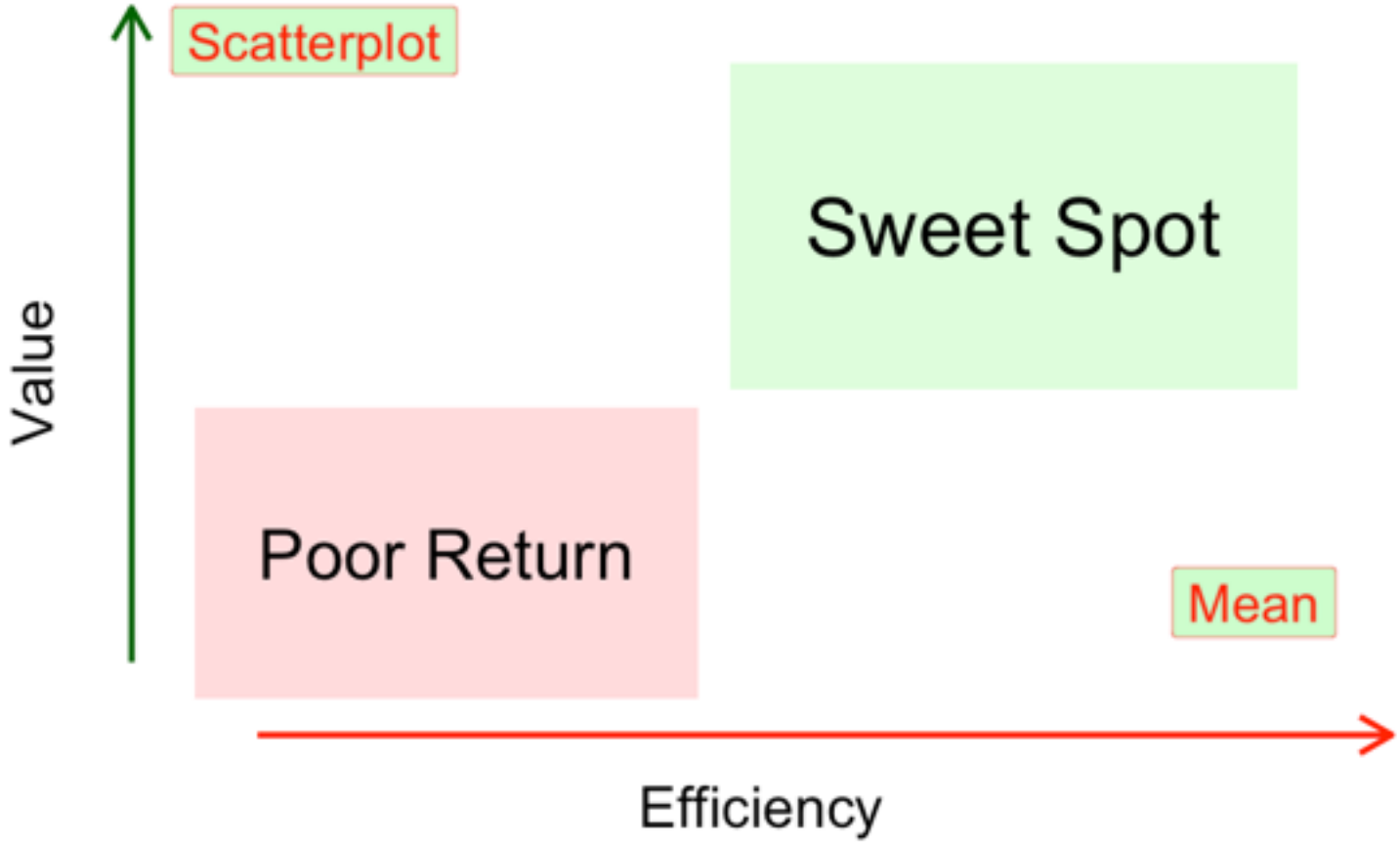
- Can you reduce your metric by combining values recursively?
- Space required grows linearly with length of history; a simple time series
- Examples:
  - Count, Sum
  - Mean, Standard Deviation
  - Min, Max

Hard to Reduce Data

- Examples:
  - Histograms (keep N buckets for each time period)
  - Median
  - Percentiles
- The amount of storage required depends on the length of history but also the granularity of the data.
- You need N buckets in each time period.

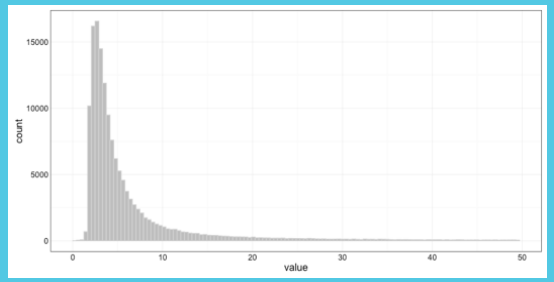
Irreducible Data (Mostly)

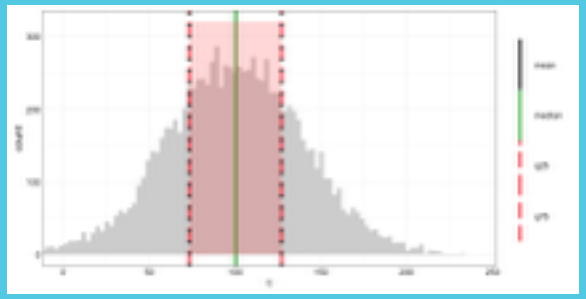
- Examples:
  - Scatterplots
  - Replays
- Requires keeping a complete history of every measurement.
- Space is proportional to the number of measures.

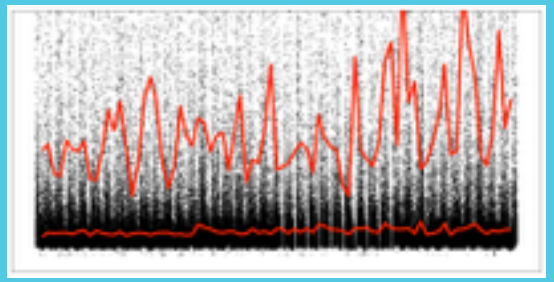

Scatterplot
Sweet Spot
Value
Poor Return
Mean
Efficiency

Evaluating Metrics

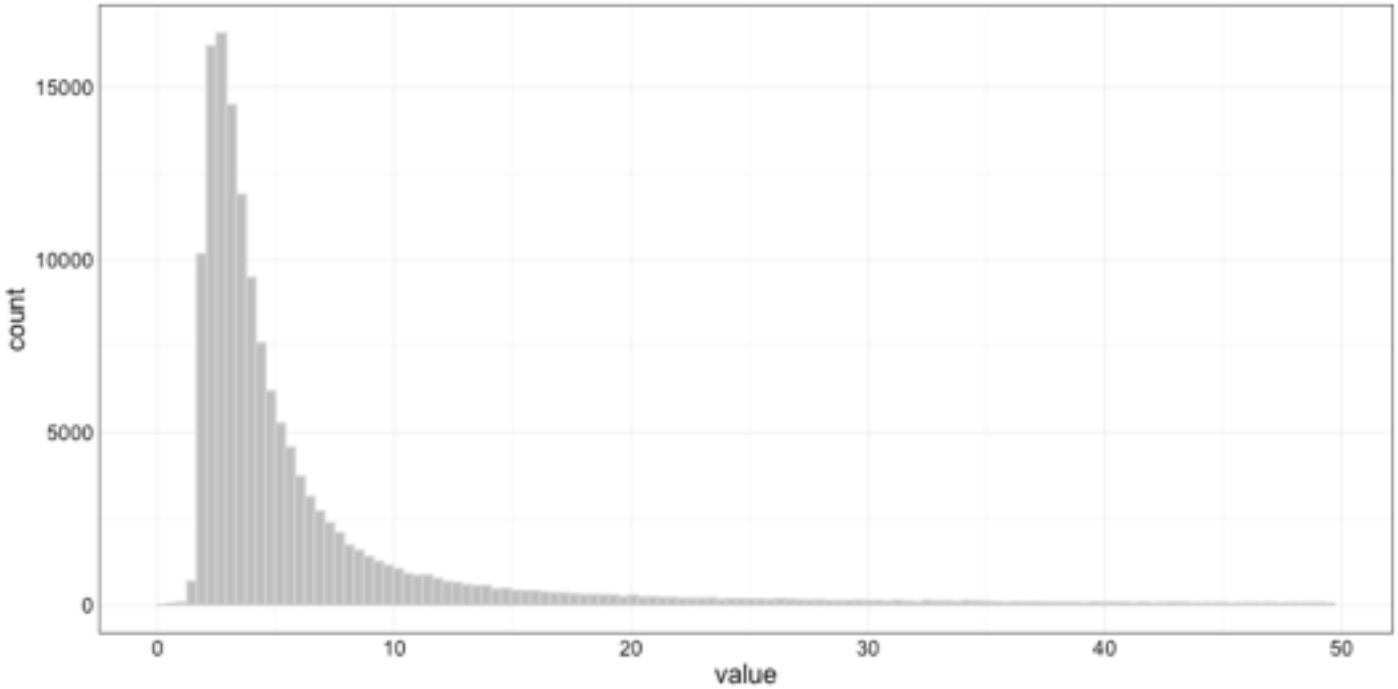
- Histograms
- Standard Deviation
- Median







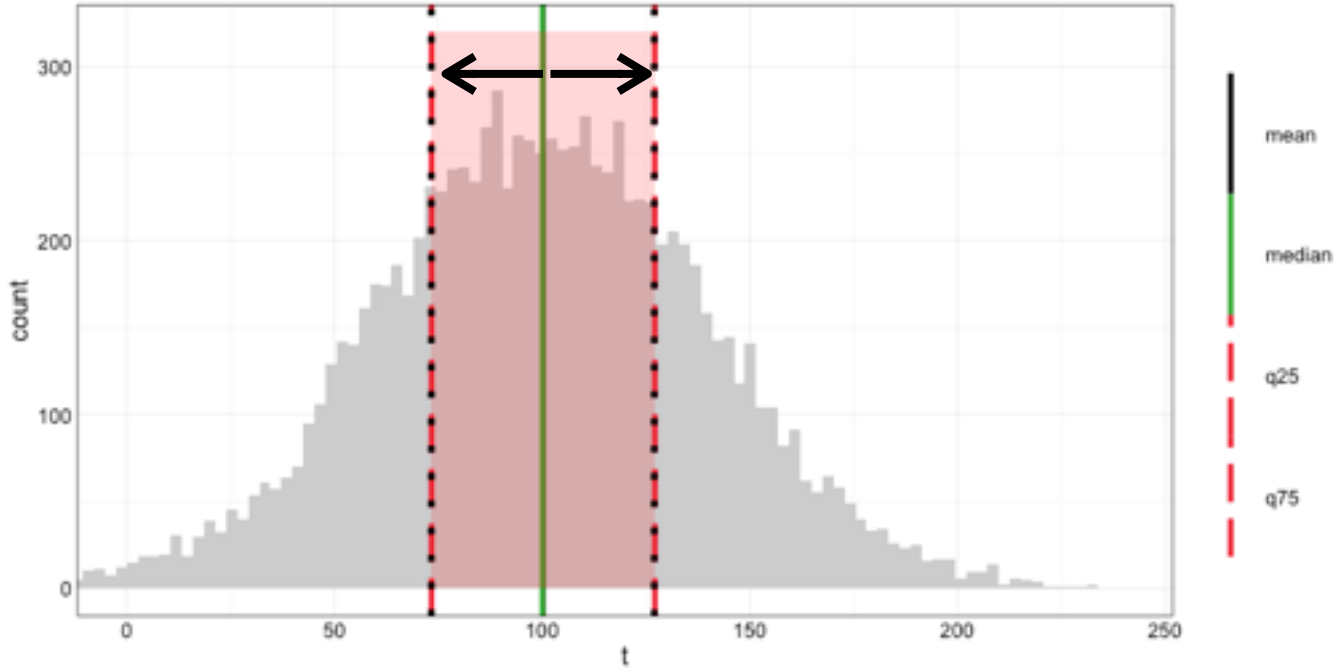
Cost/Benefit of Histograms?




Scatterplot
Histogram
Mean
Value
Efficiency

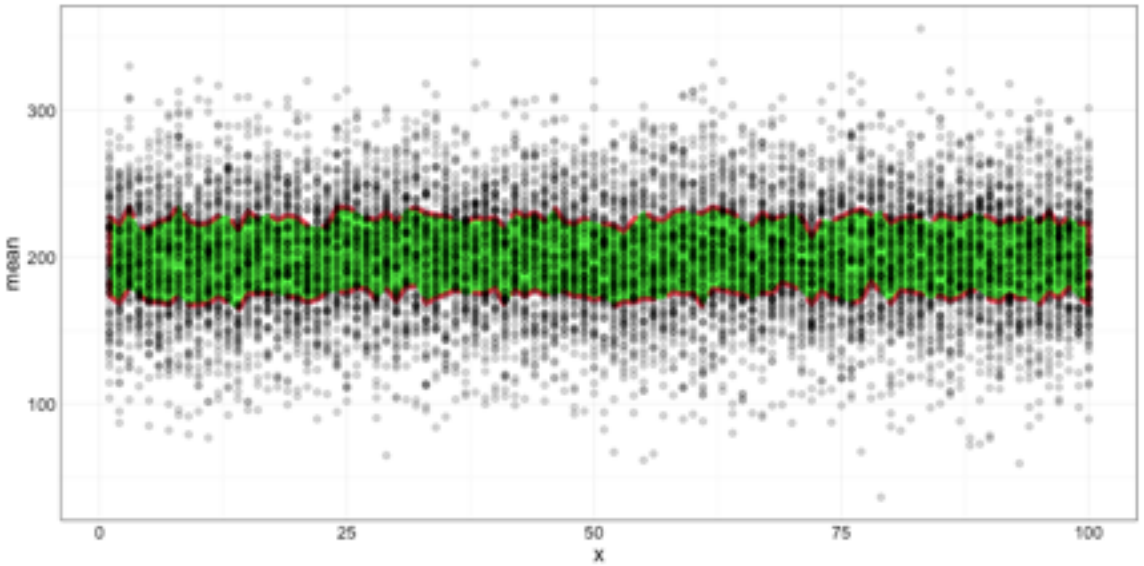
Can we make use of the Standard Deviation?

± 0.68 standard deviations



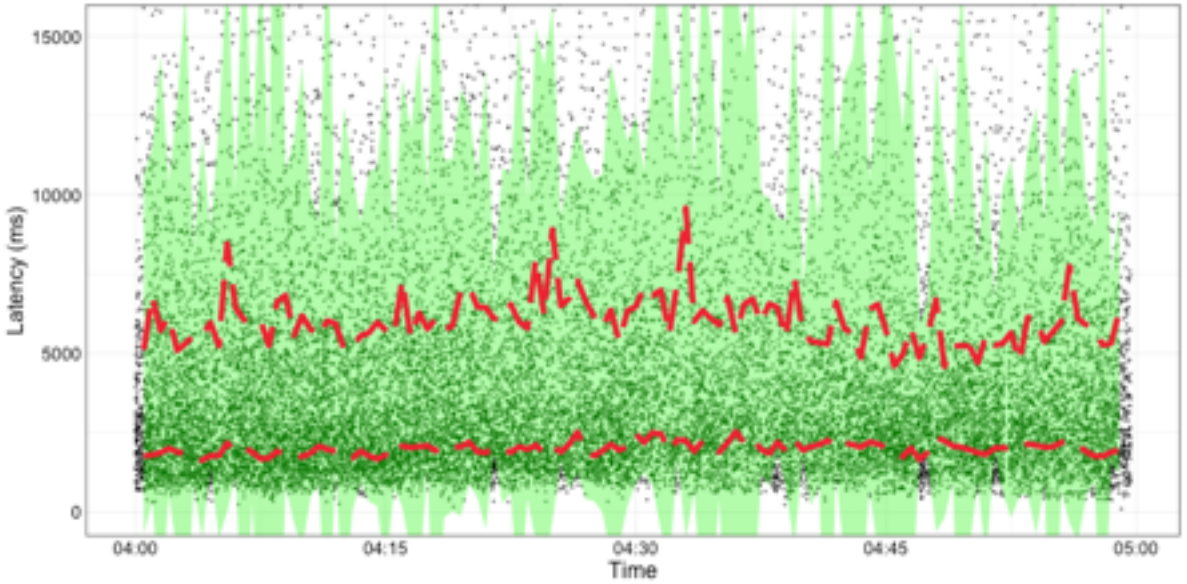
Estimating Quartiles Using Standard Deviation

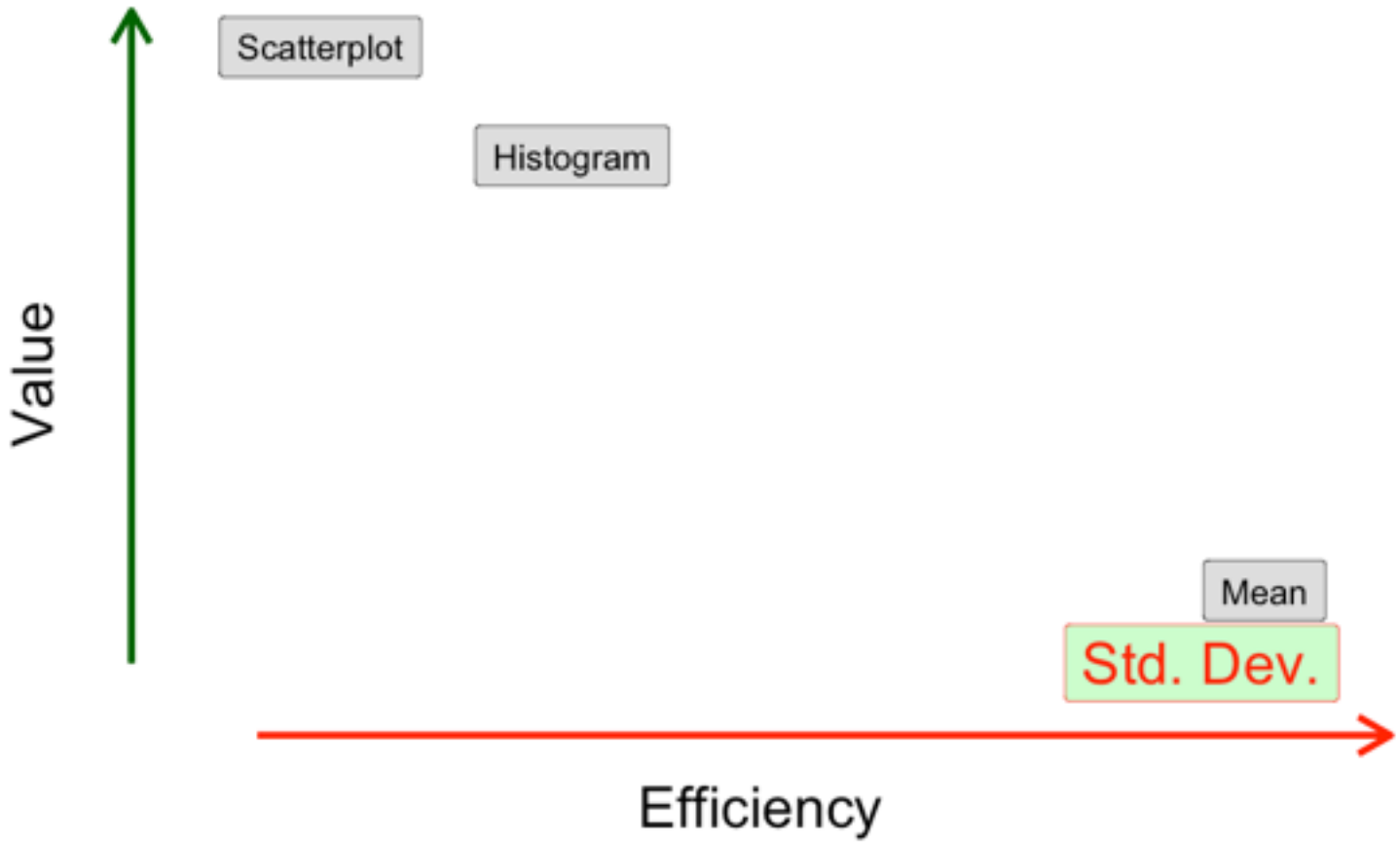
If response times had a normal (gaussian) distribution we could show everything using simple statistics and estimated percentiles.



Estimating Quartiles Using Standard Deviation

But instead it would end up looking like this:

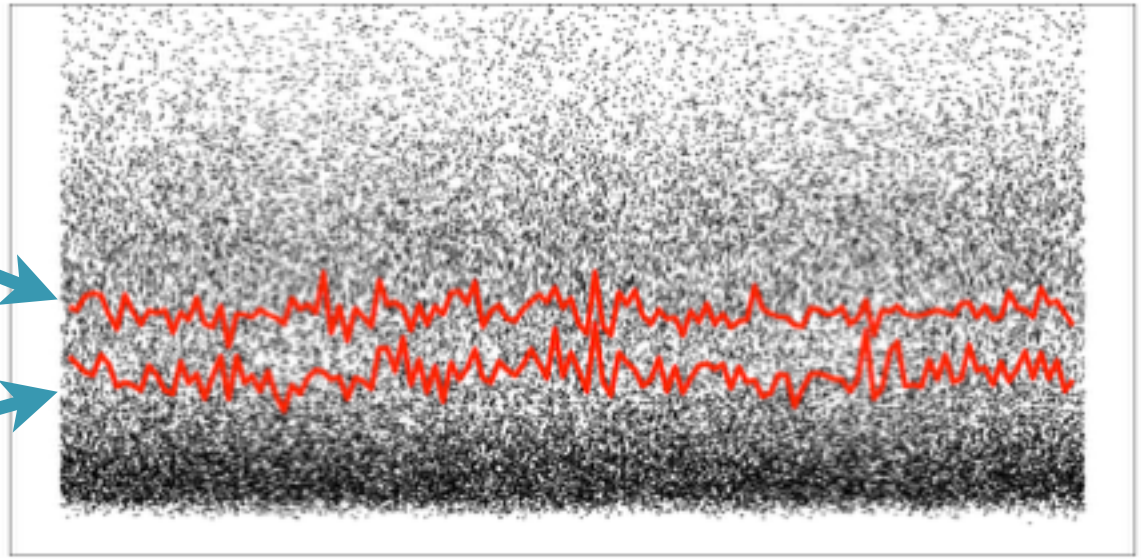



Scatterplot
Histogram
Mean
Std. Dev.
Value
Efficiency

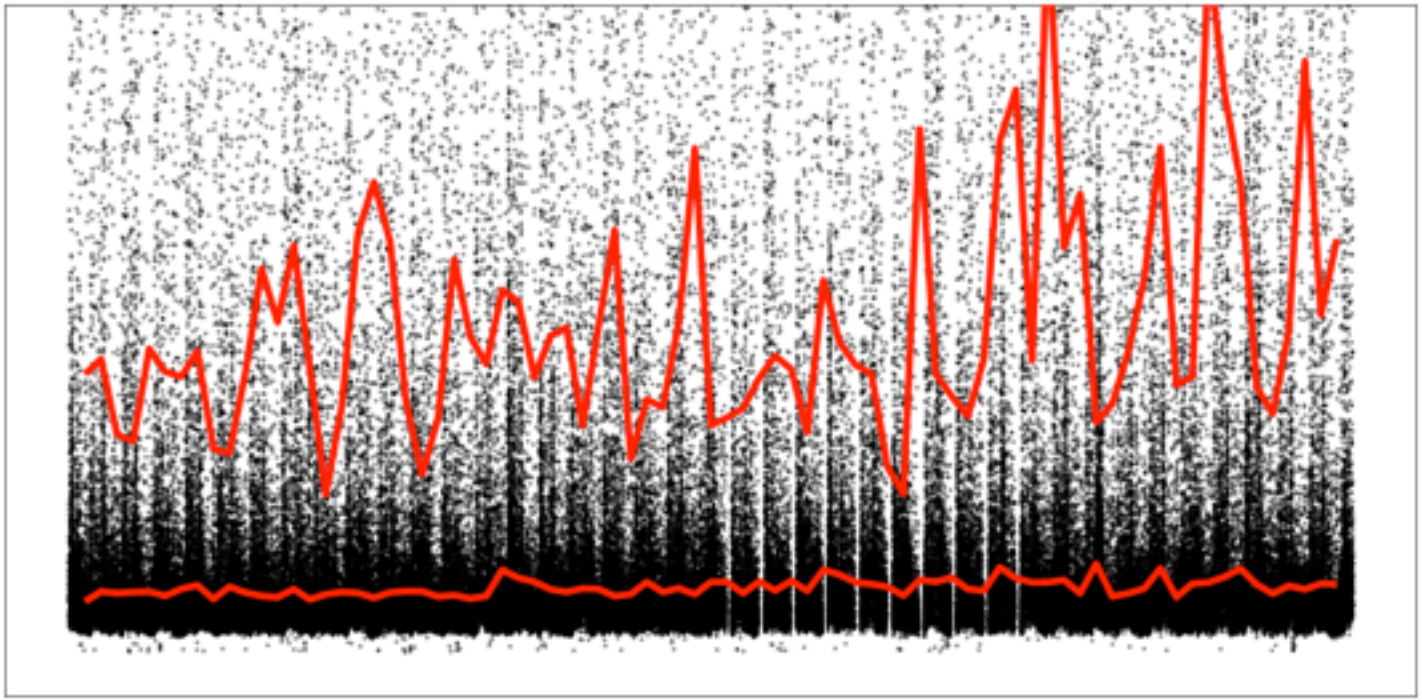
Mean vs Median

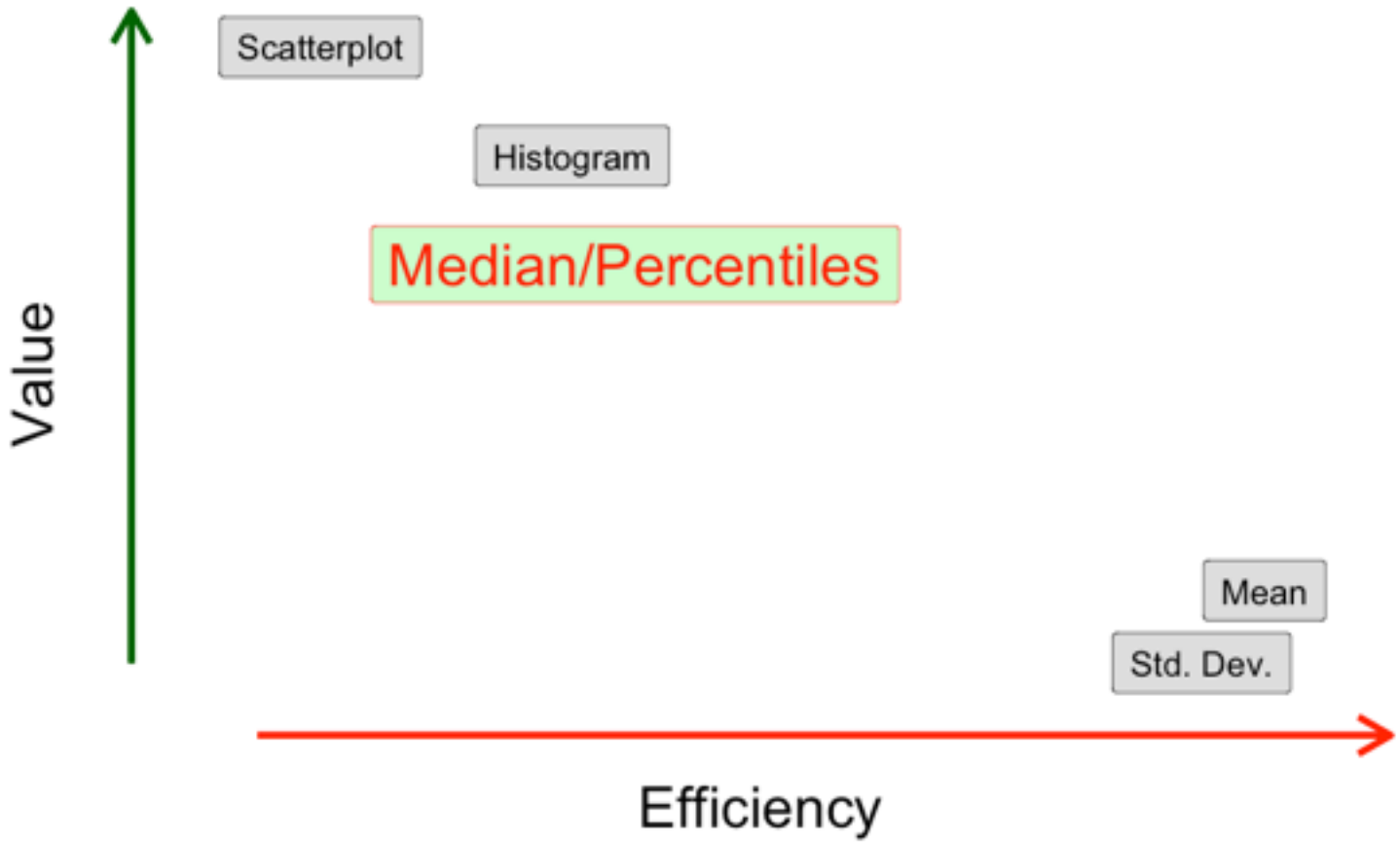
Mean

Median



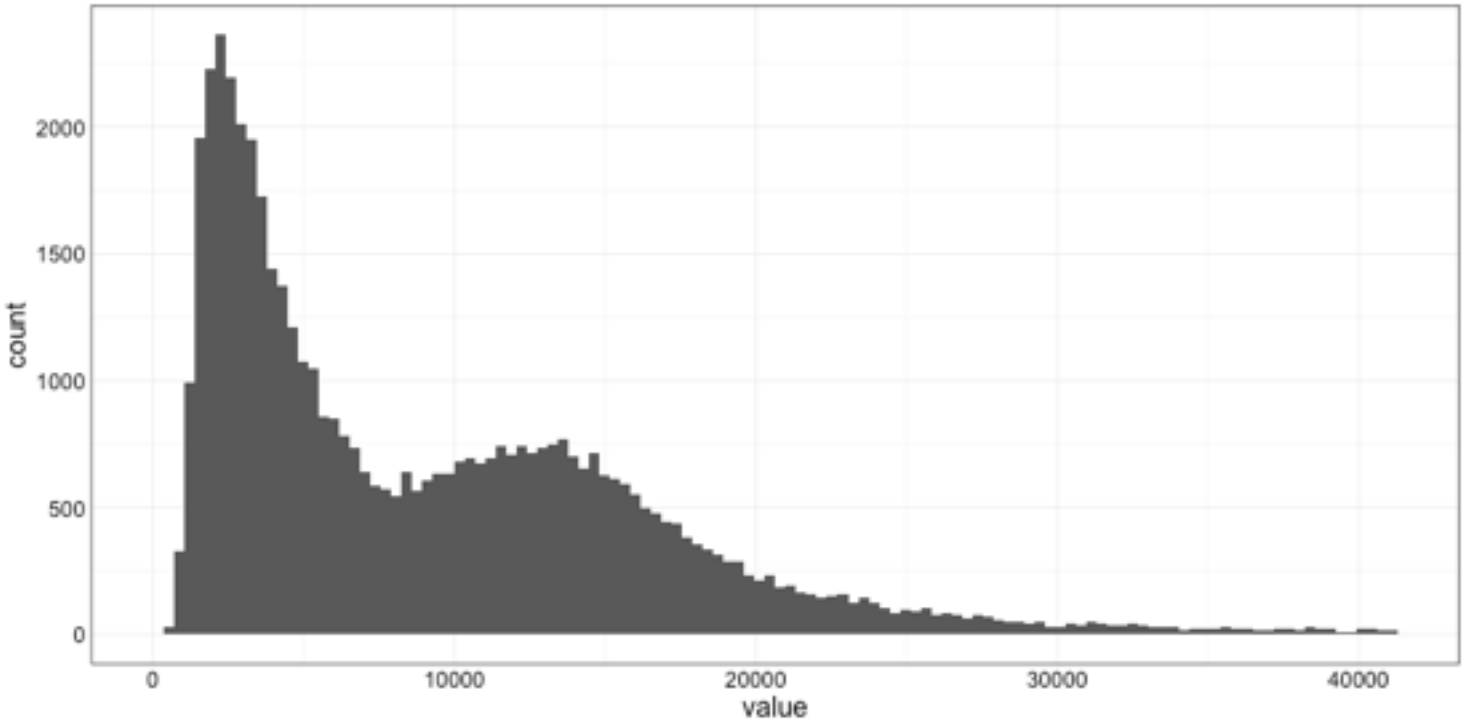
Mean vs Median



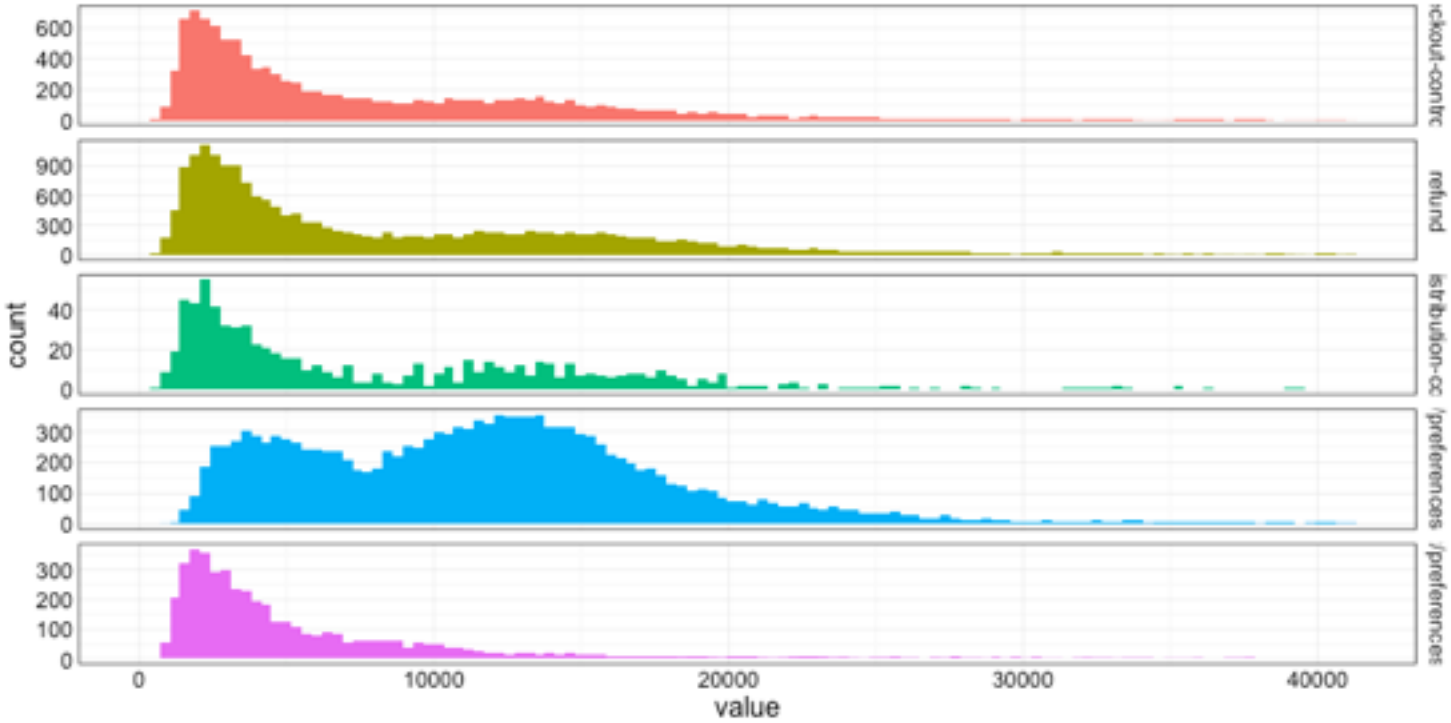

Scatterplot
Histogram
Median/Percentiles
Mean
Std. Dev.
Value
Efficiency

Exploring Histograms and Geometric Mean

Multimodal Histograms

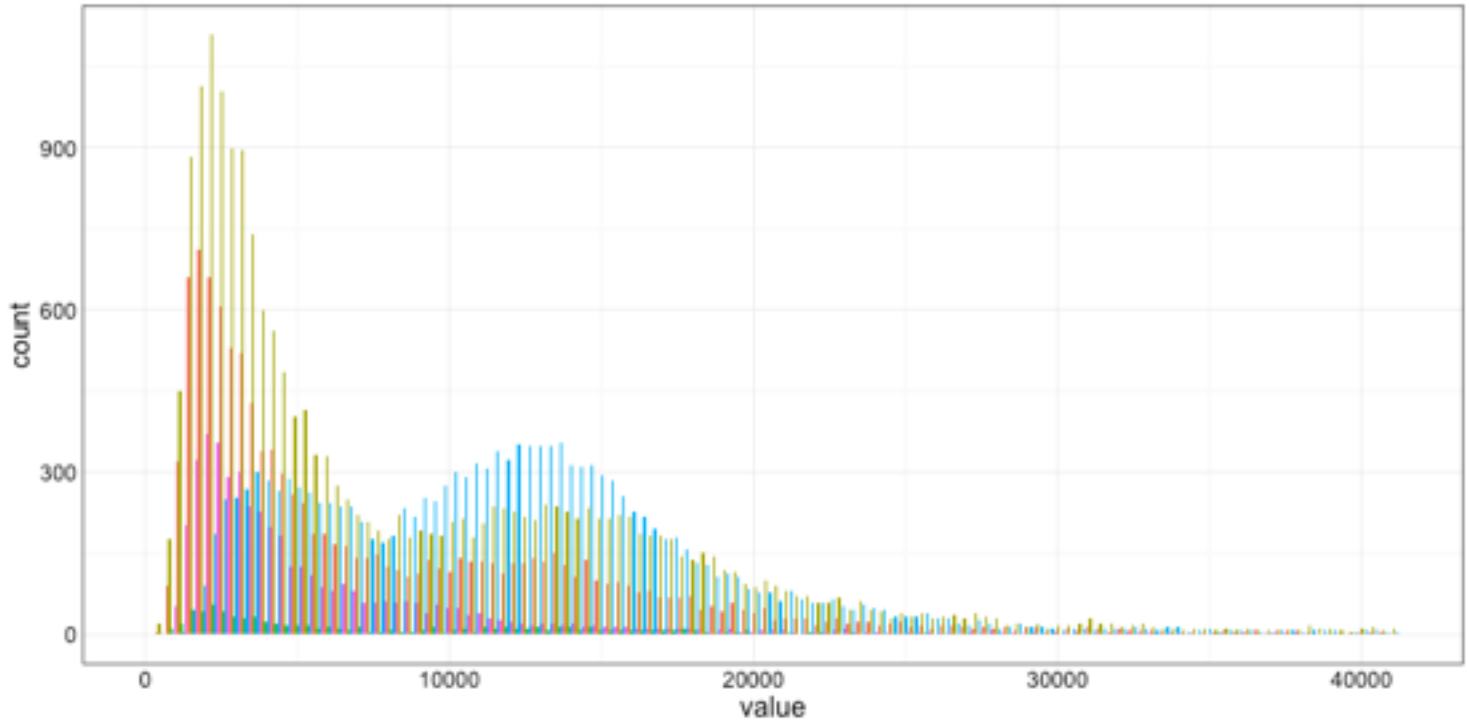


Multimodal Histograms

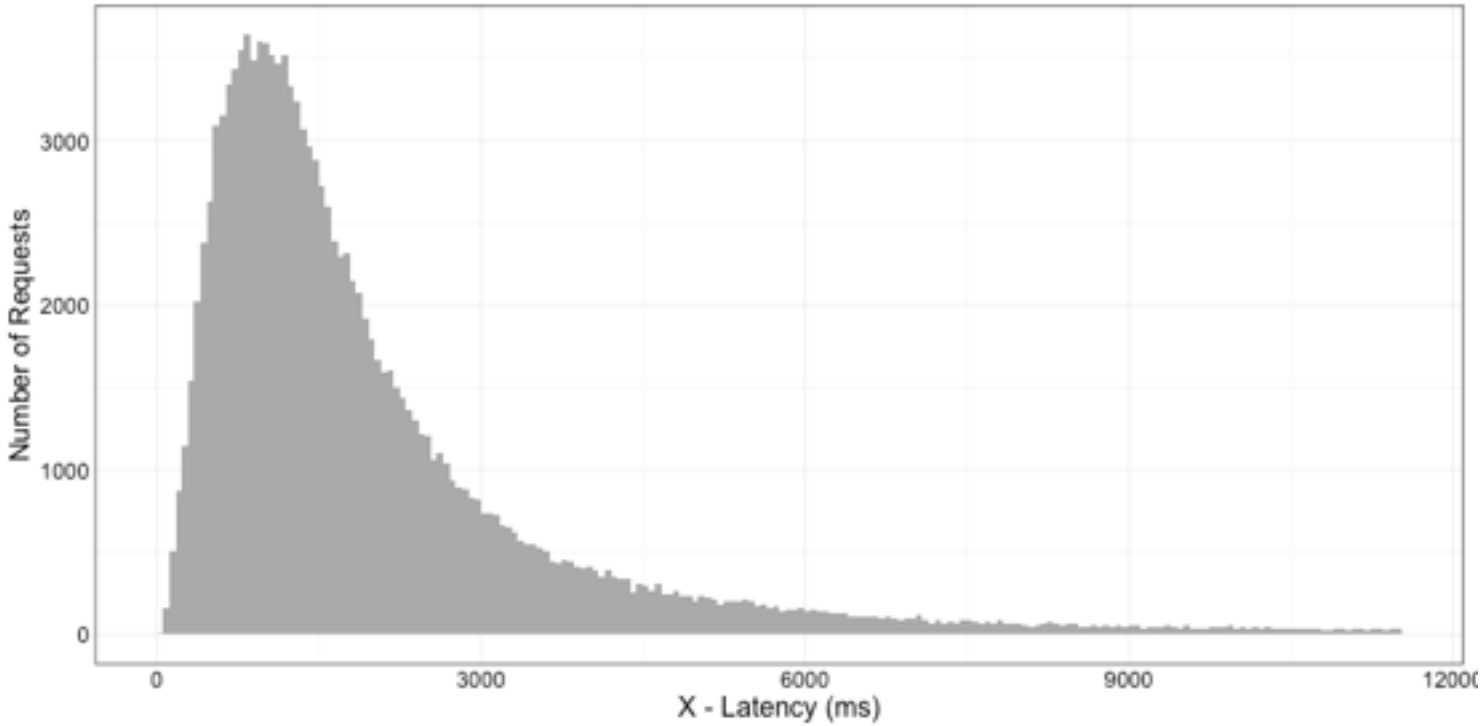


New Relic.

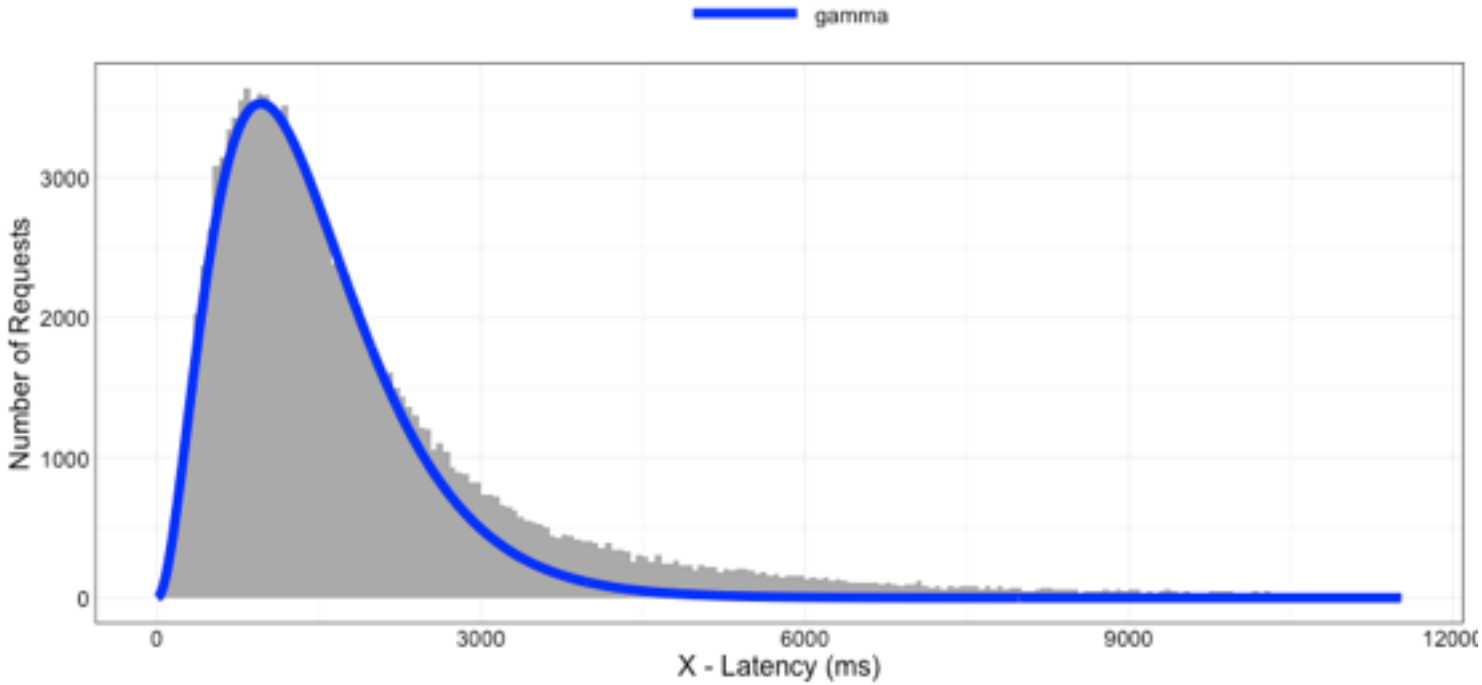
Multimodal Histograms



Histogram Example

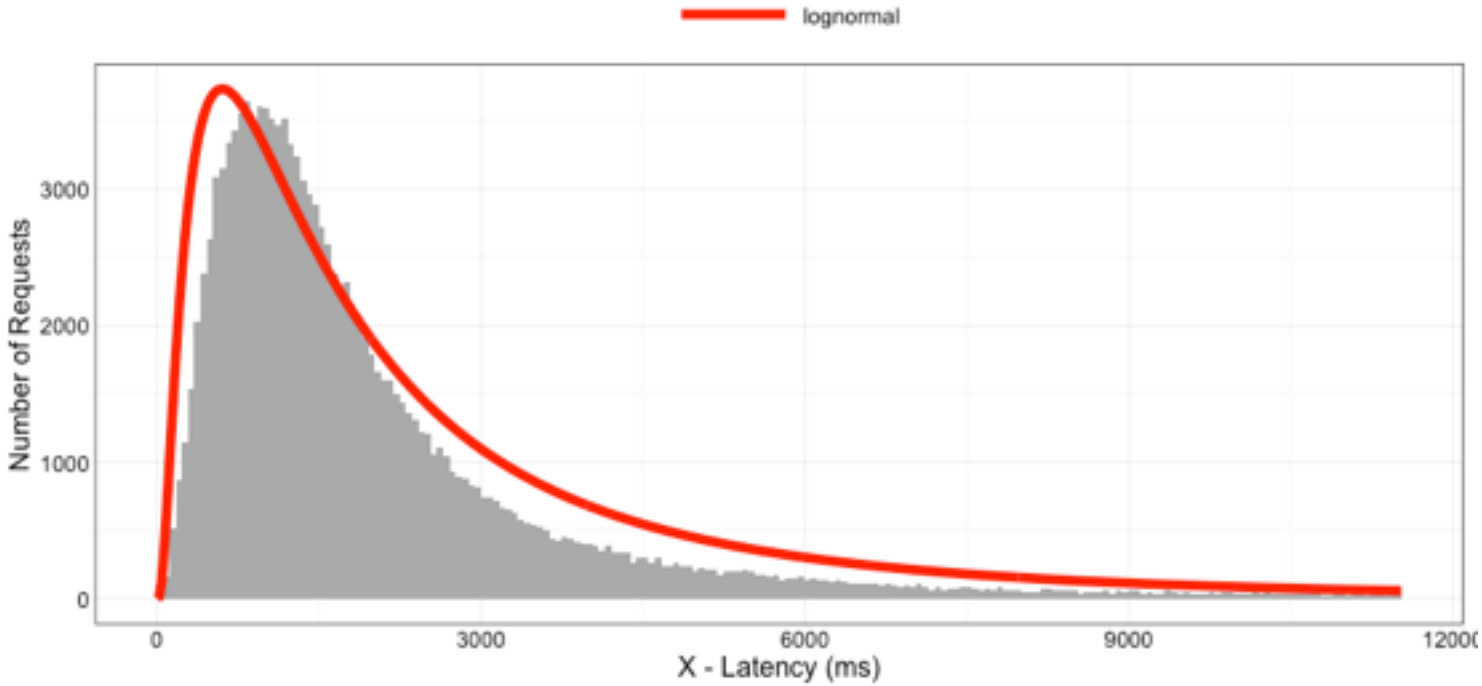


Gamma Distribution

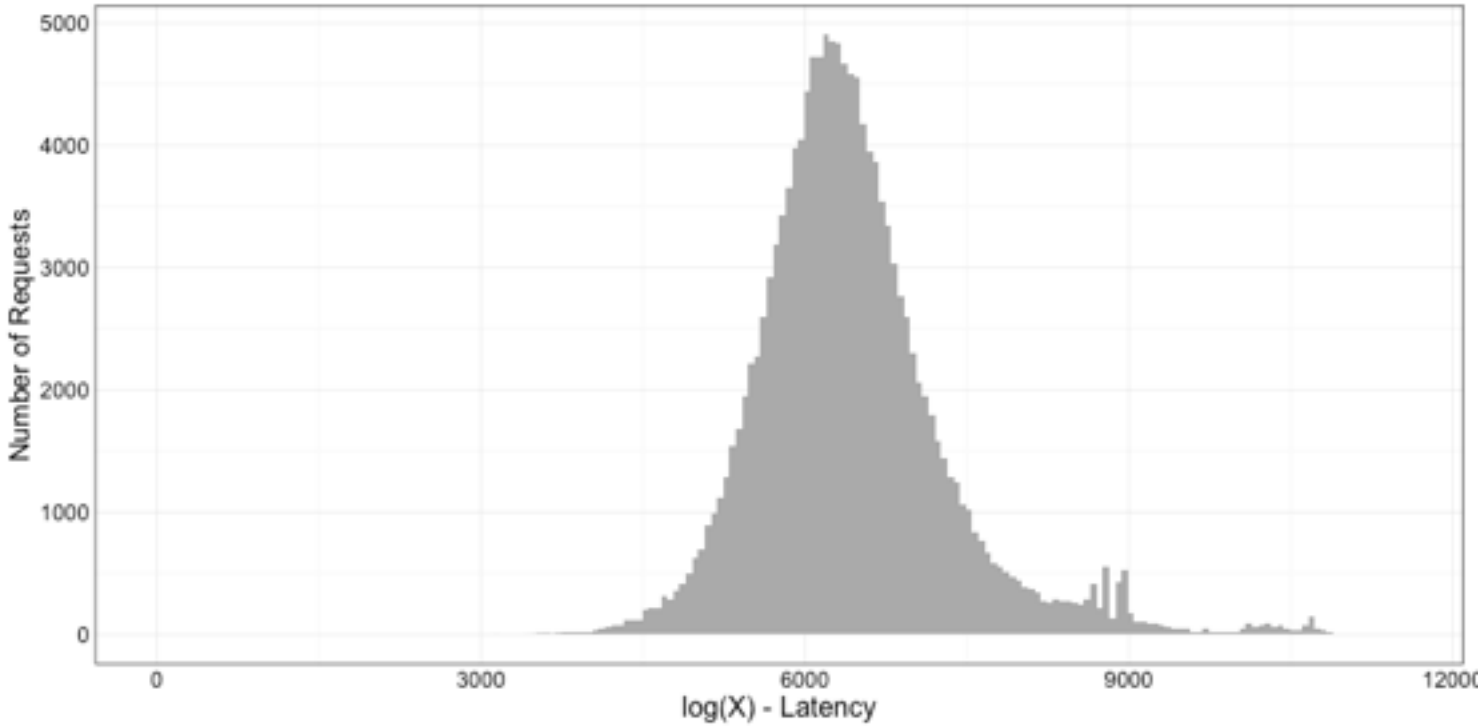


New Relic.

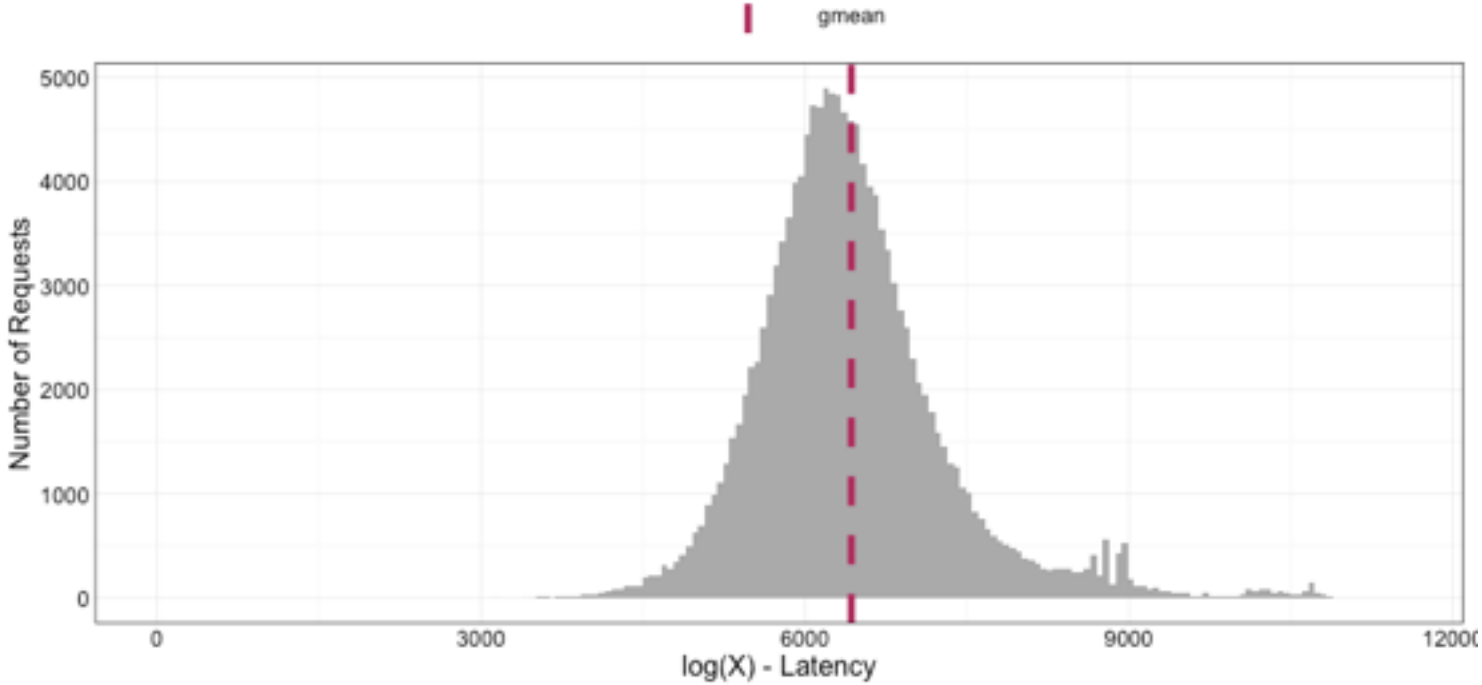
Log Normal Distribution



Response Times in Log Space

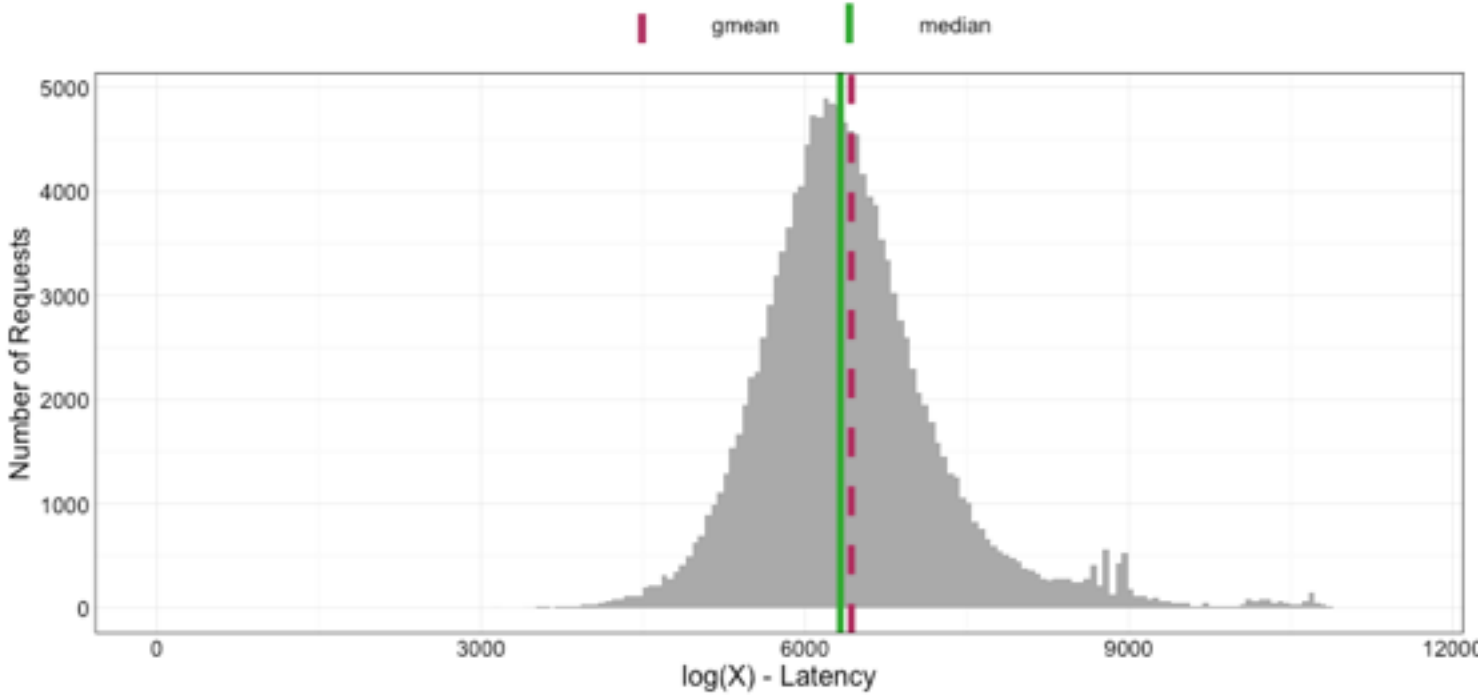


Response Times in Log Space

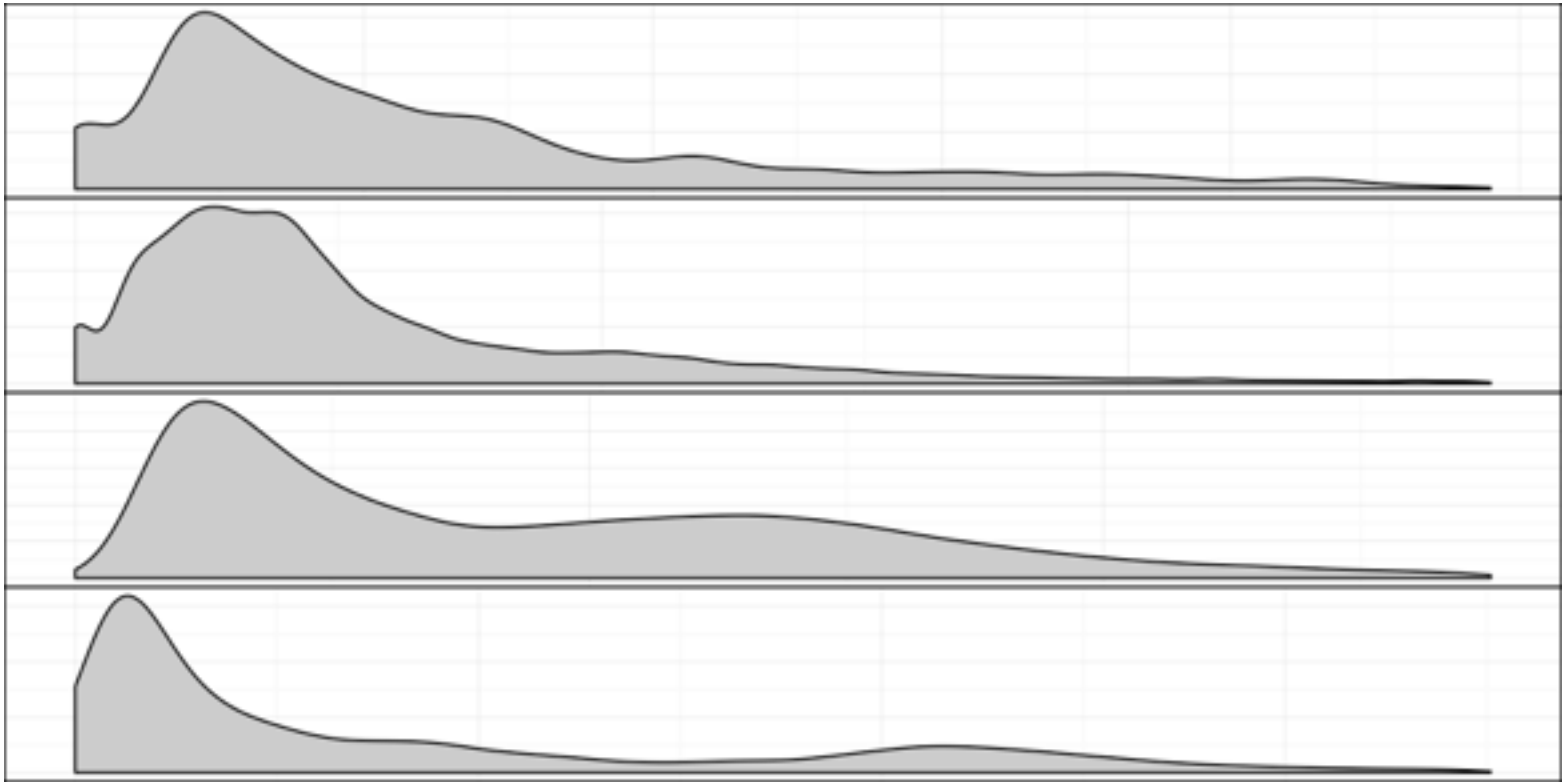


New Relic.

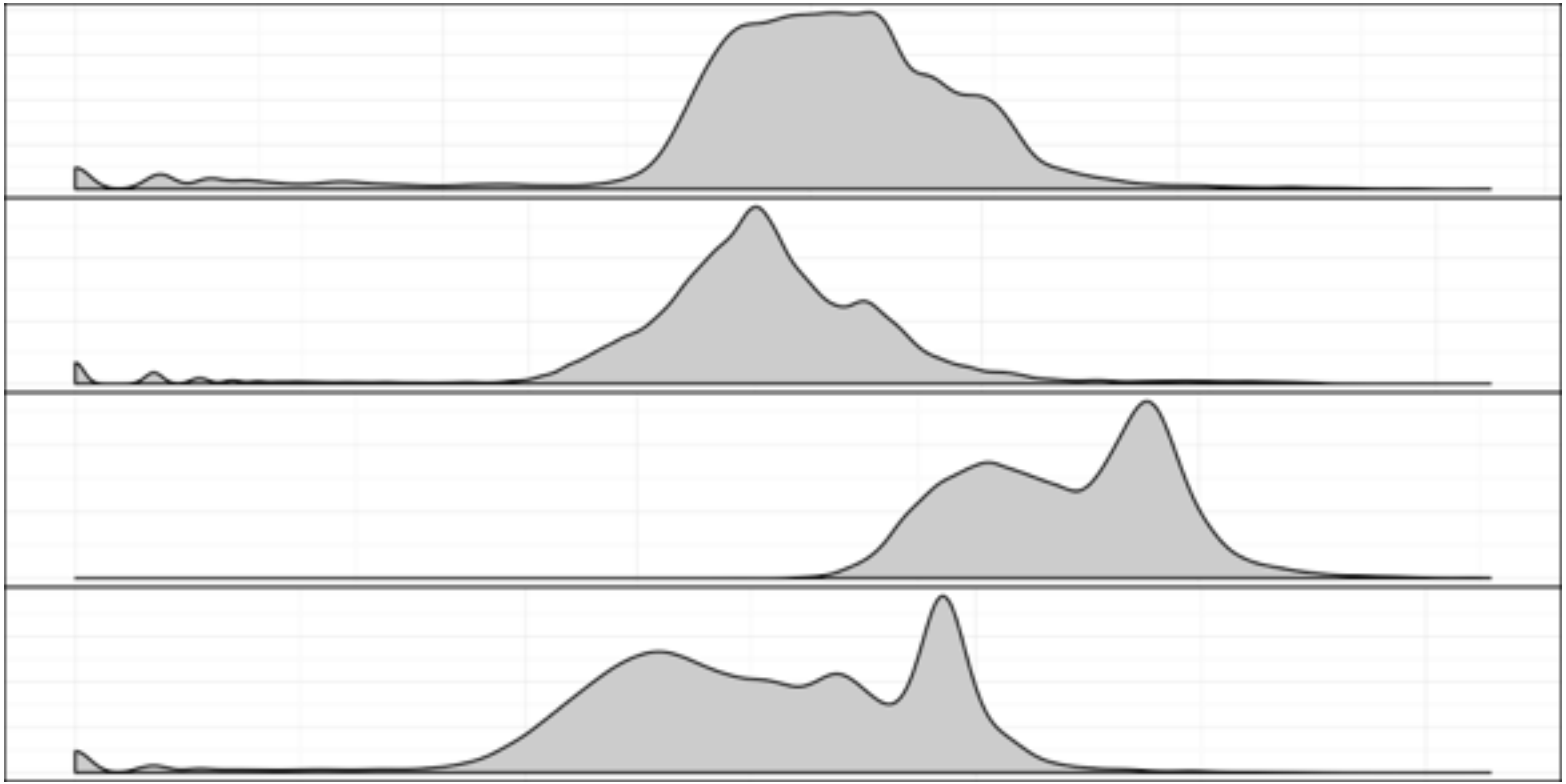
Response Times in Log Space



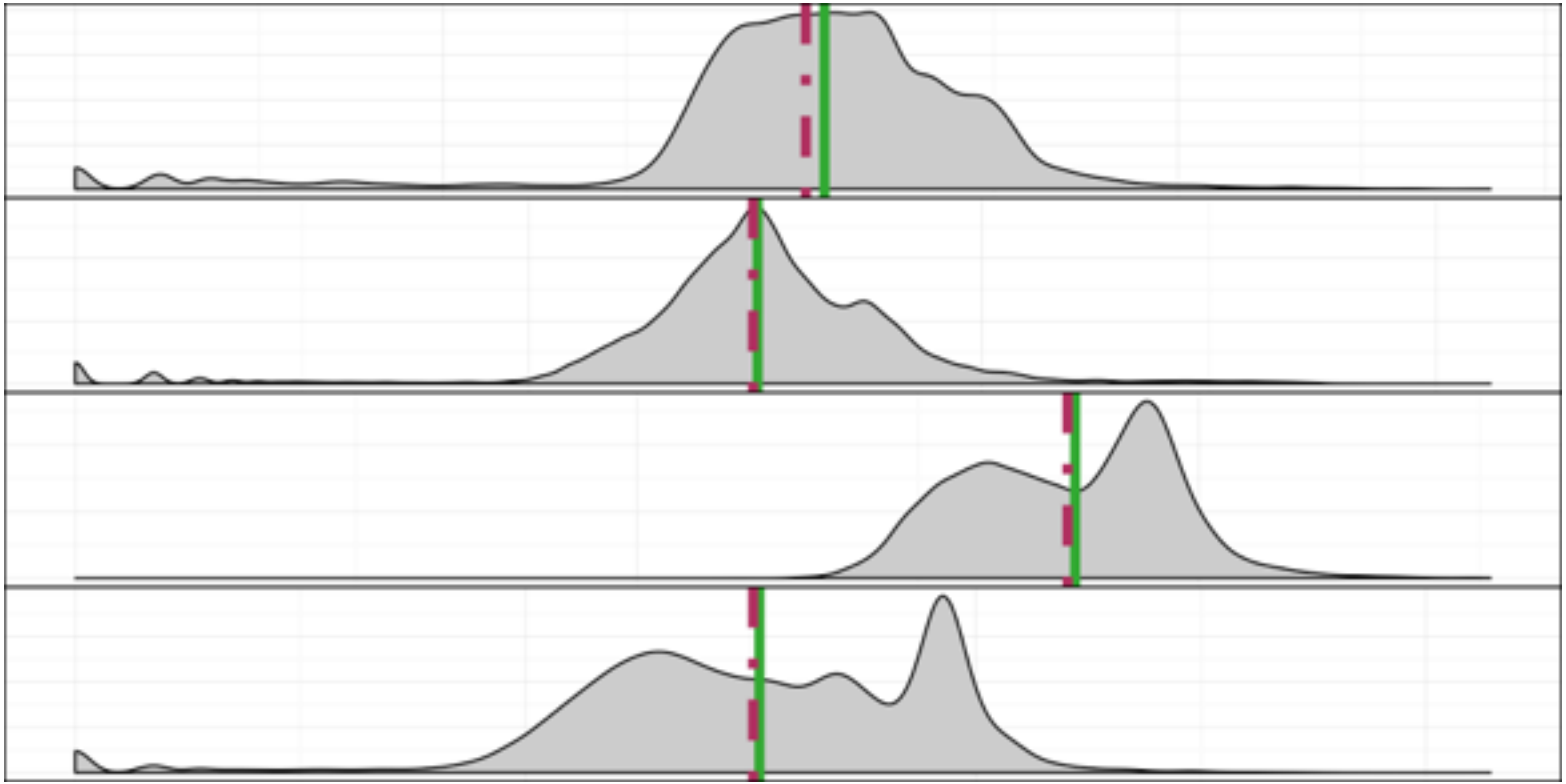
Response Time Frequency Charts



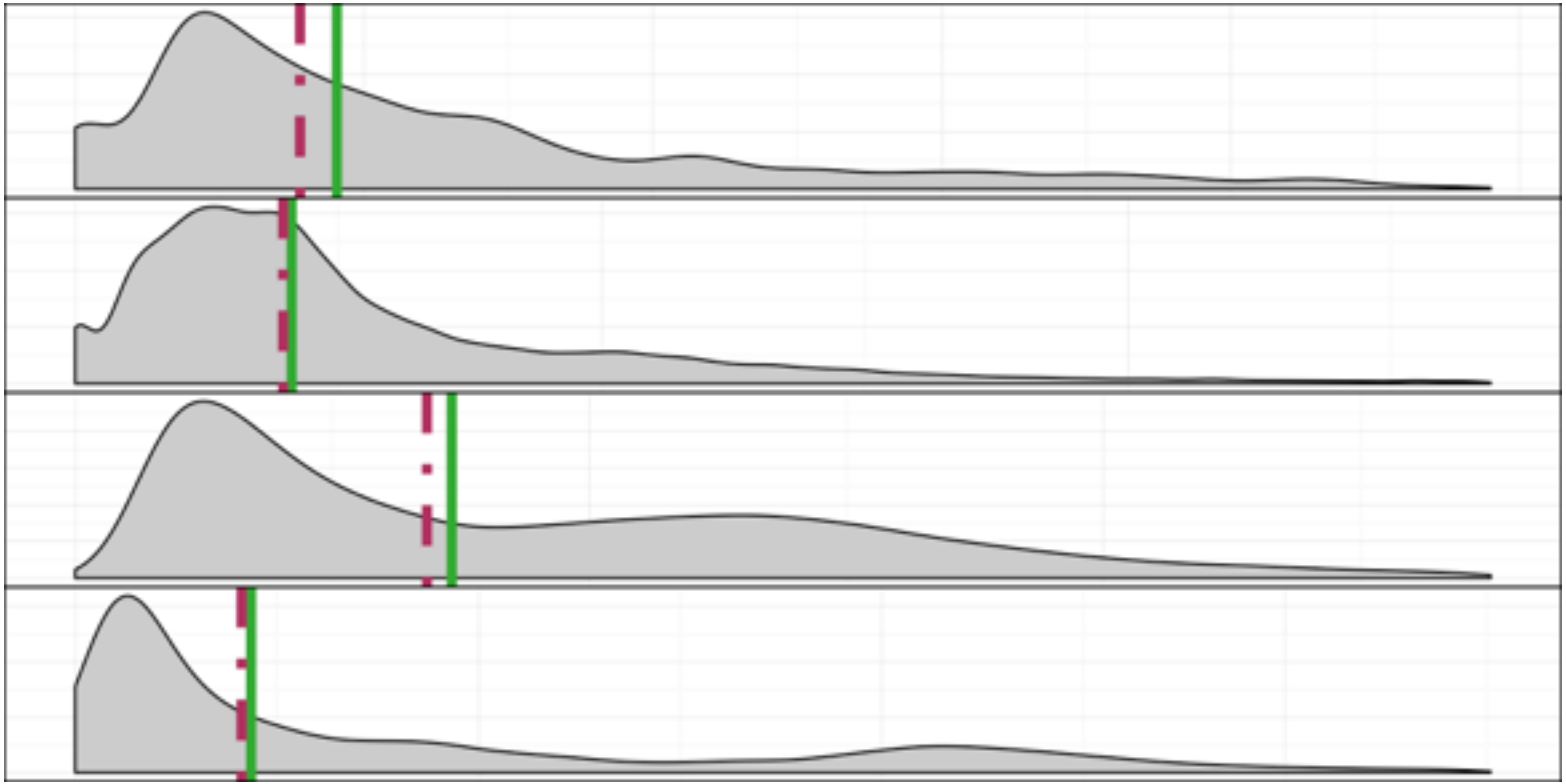
Frequency Chart of the *log()* of Response Times



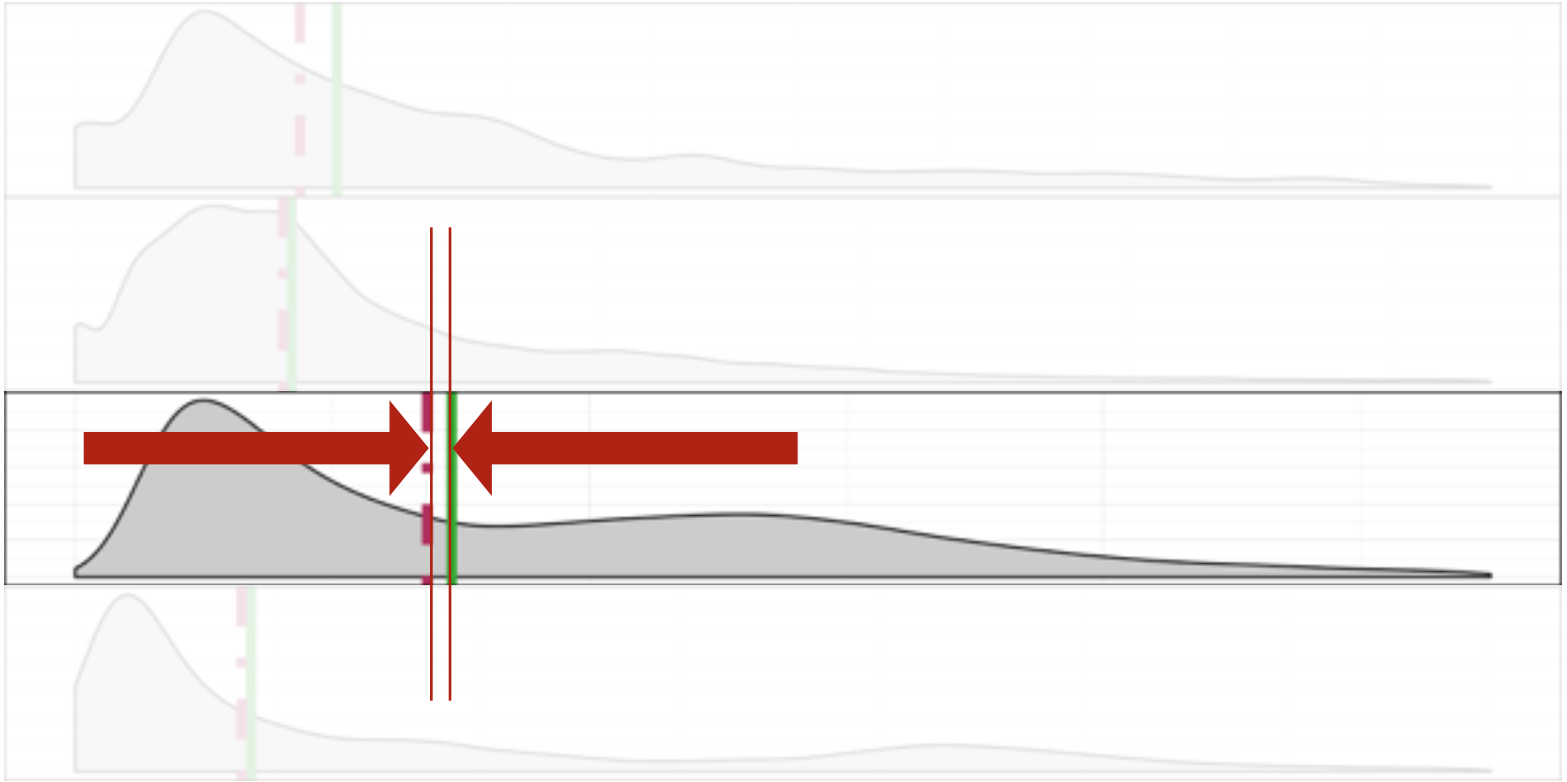
Mean (solid green) and Median (dashed red) Values in Log Space



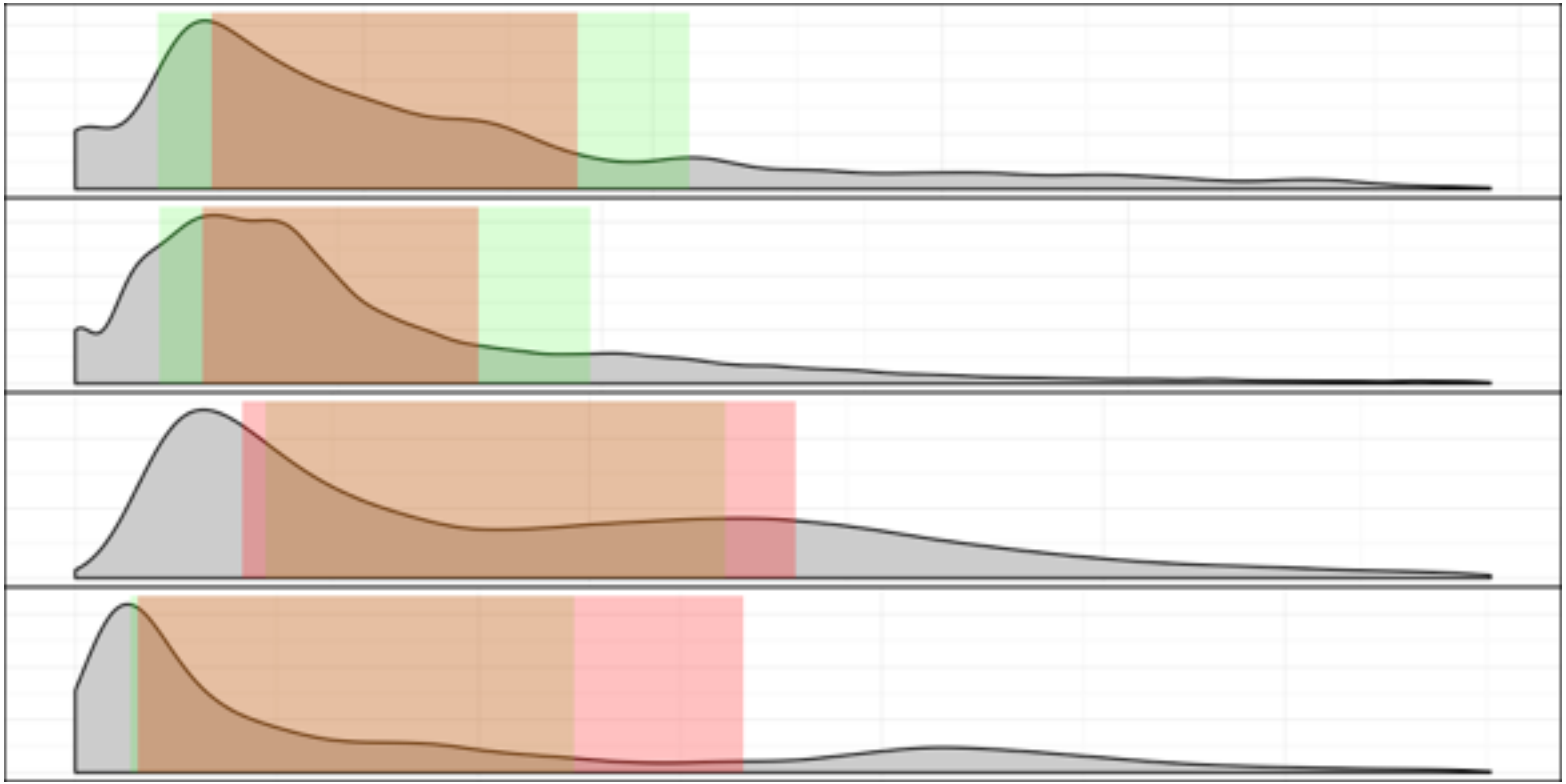
Take *exp(mean)* and plot back in untransformed histograms (dashed red) and compare with the median (solid green).



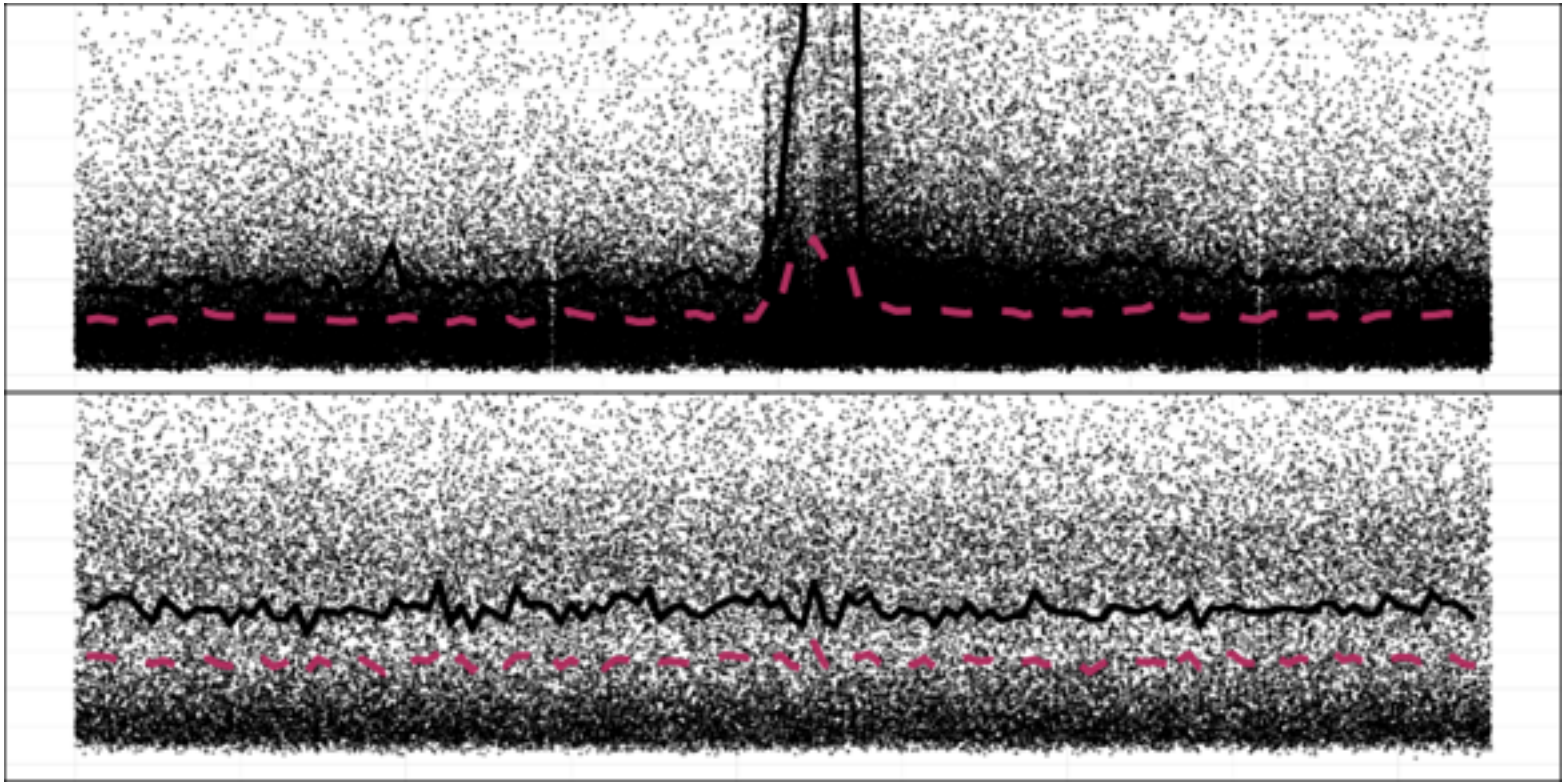
The mean in log space makes an excellent approximation of the median!



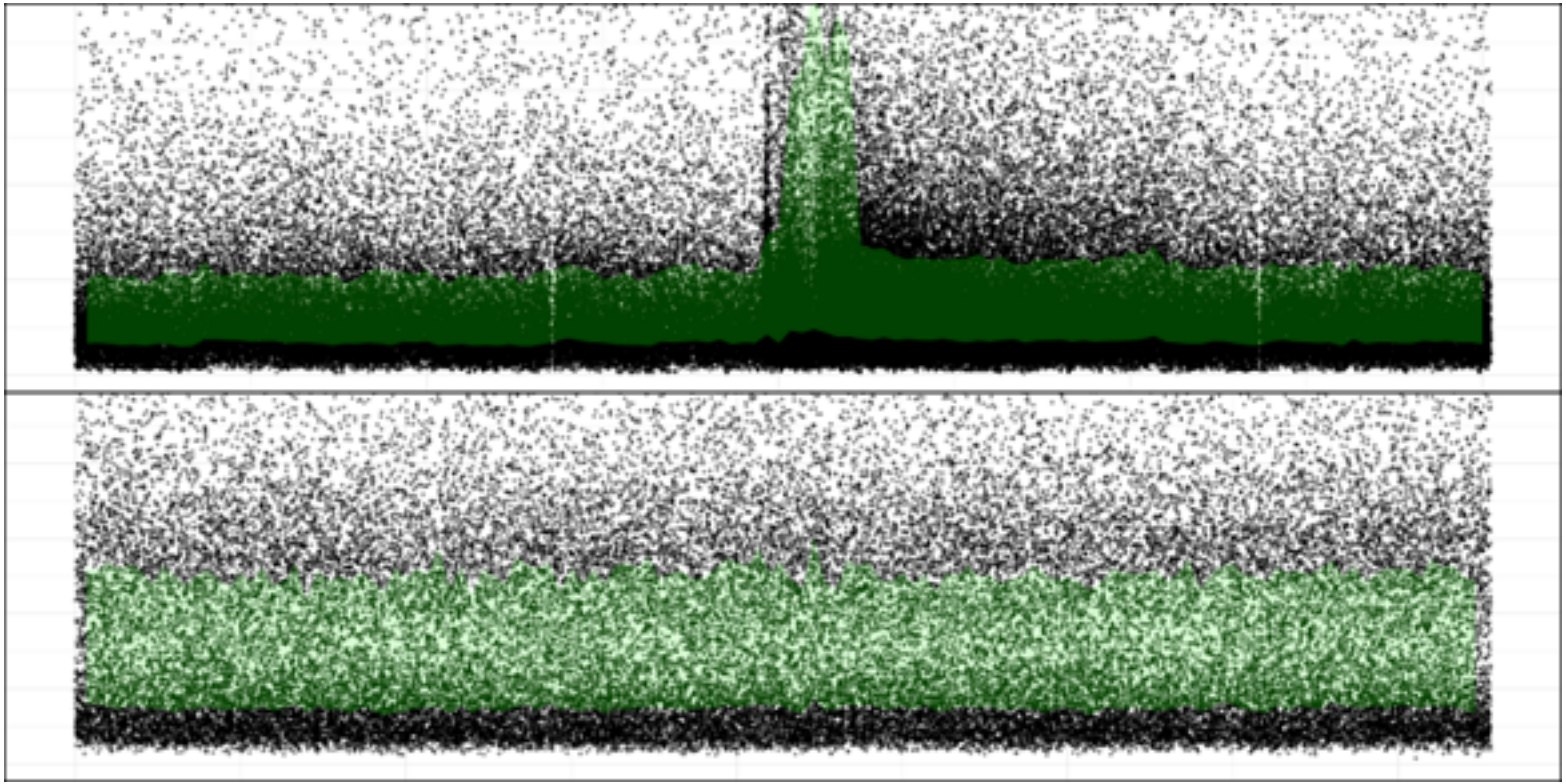
Estimating inner quartiles is also pretty close.



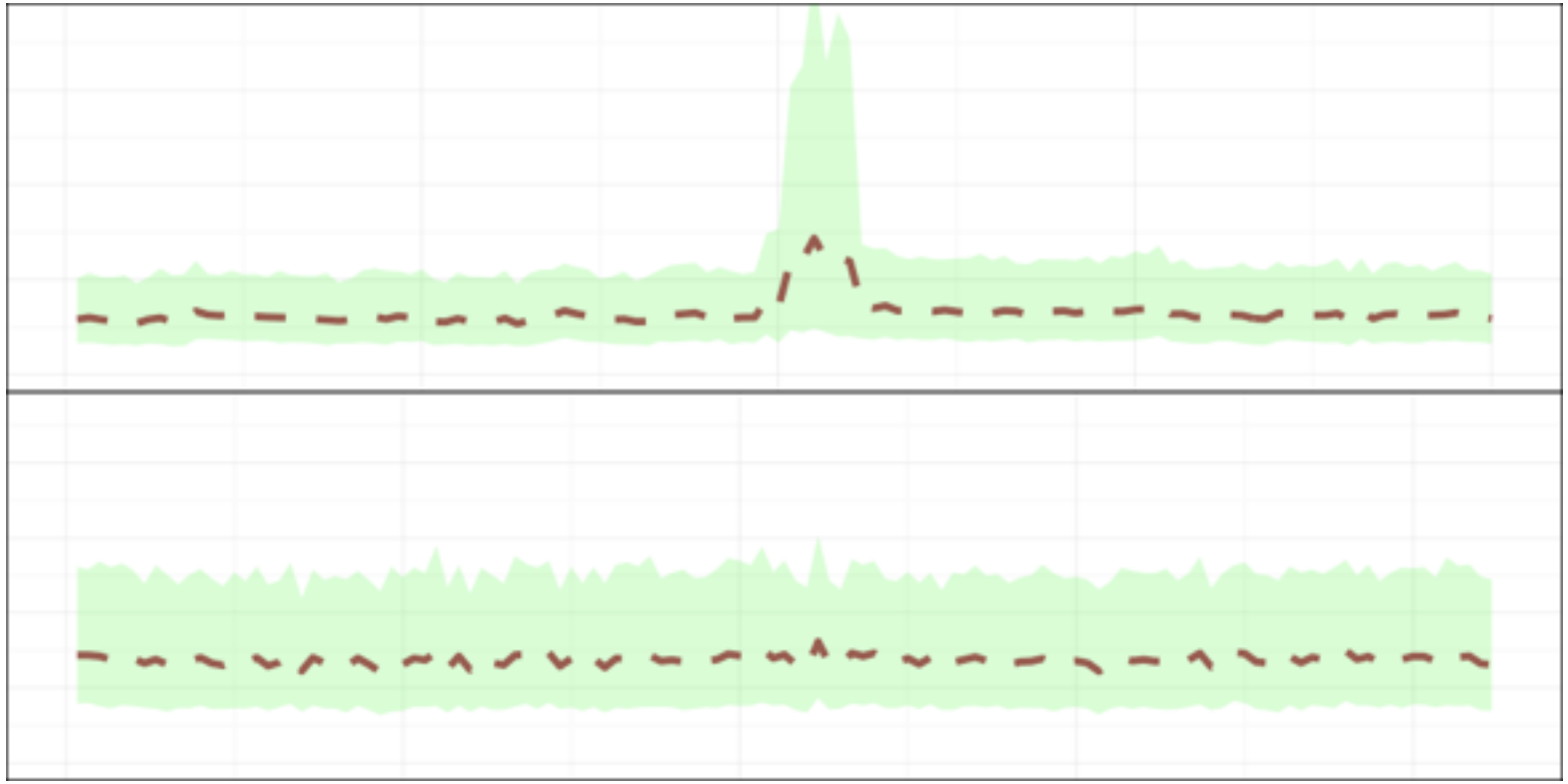
Arithmetic Mean (solid black) vs Geometric Mean (dashed red)



Geometric Quartiles



Geometric Quartiles



Calculating the Geometric Mean and Standard Deviation

$$GM = exp\left(\frac{1}{n}\sum_{i=1}^{n} ln(t_i)\right)$$

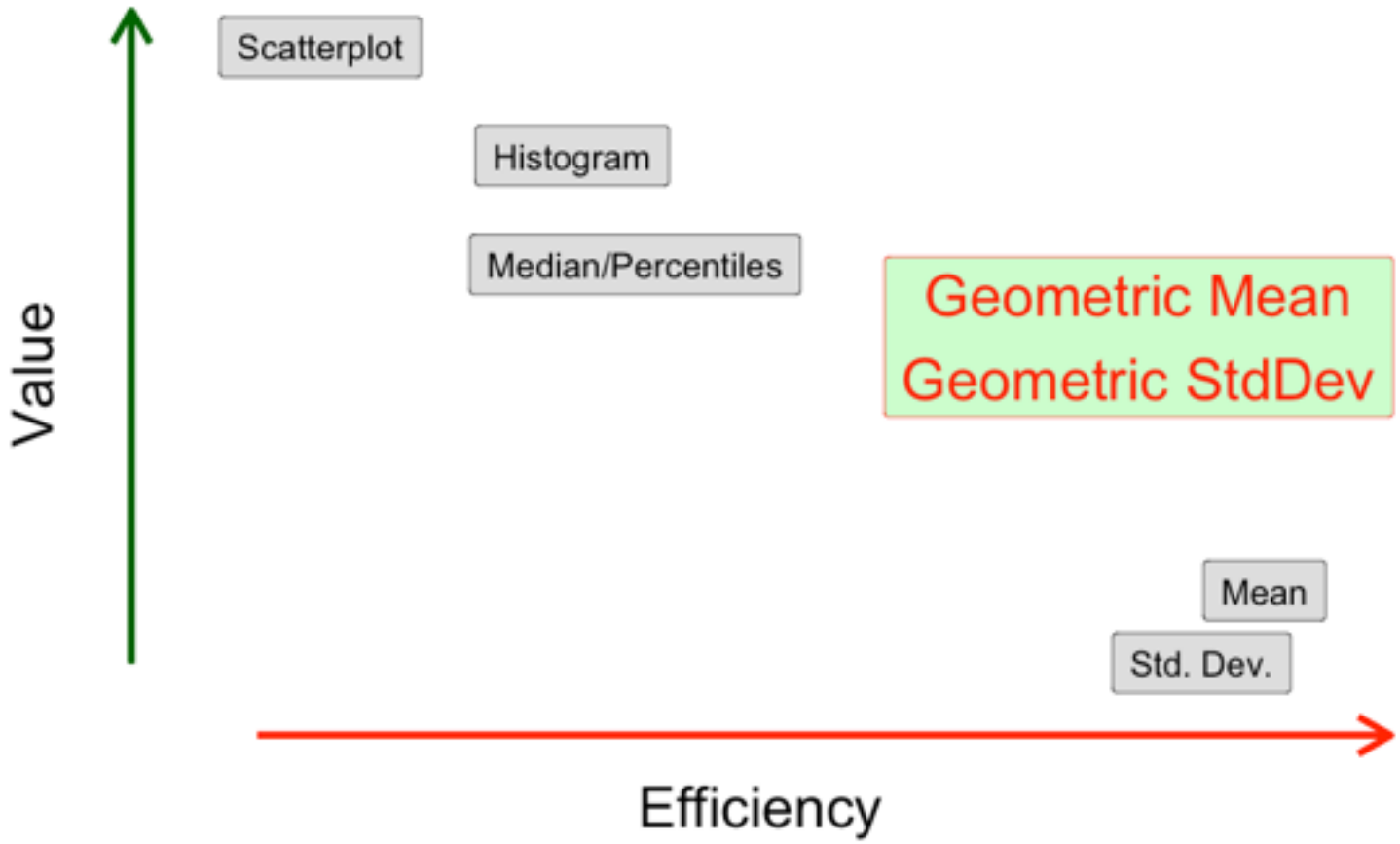
Instead of collecting the sum of the response times, you collect the sum of the log of the response times.

$$GSD = exp\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n} ln(t_i)^2 - \left(\frac{1}{n}\sum_{i=1}^{n} ln(t_i)\right)^2}\right)$$

Instead of collecting the sum of squares to calculate the standard deviations, you collect the sum of the squared log of response times.

$$\left[\frac{GM}{GSD^Z} \cdots GM \times GSD^Z\right]$$

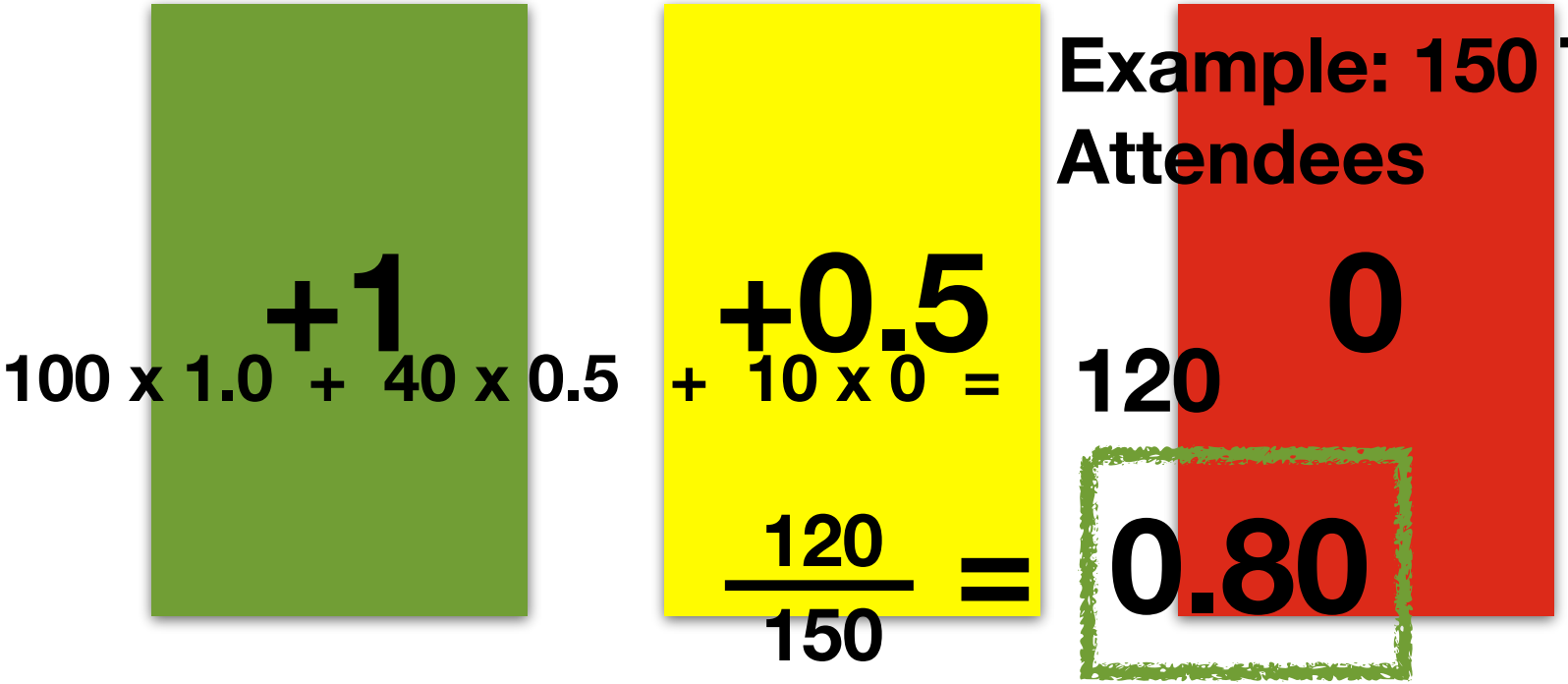
To calculate the inner quartile range you use a formula based on the GSD.


Scatterplot
Histogram
Median/Percentiles
Geometric Mean
Geometric StdDev
Mean
Std. Dev.
Value
Efficiency

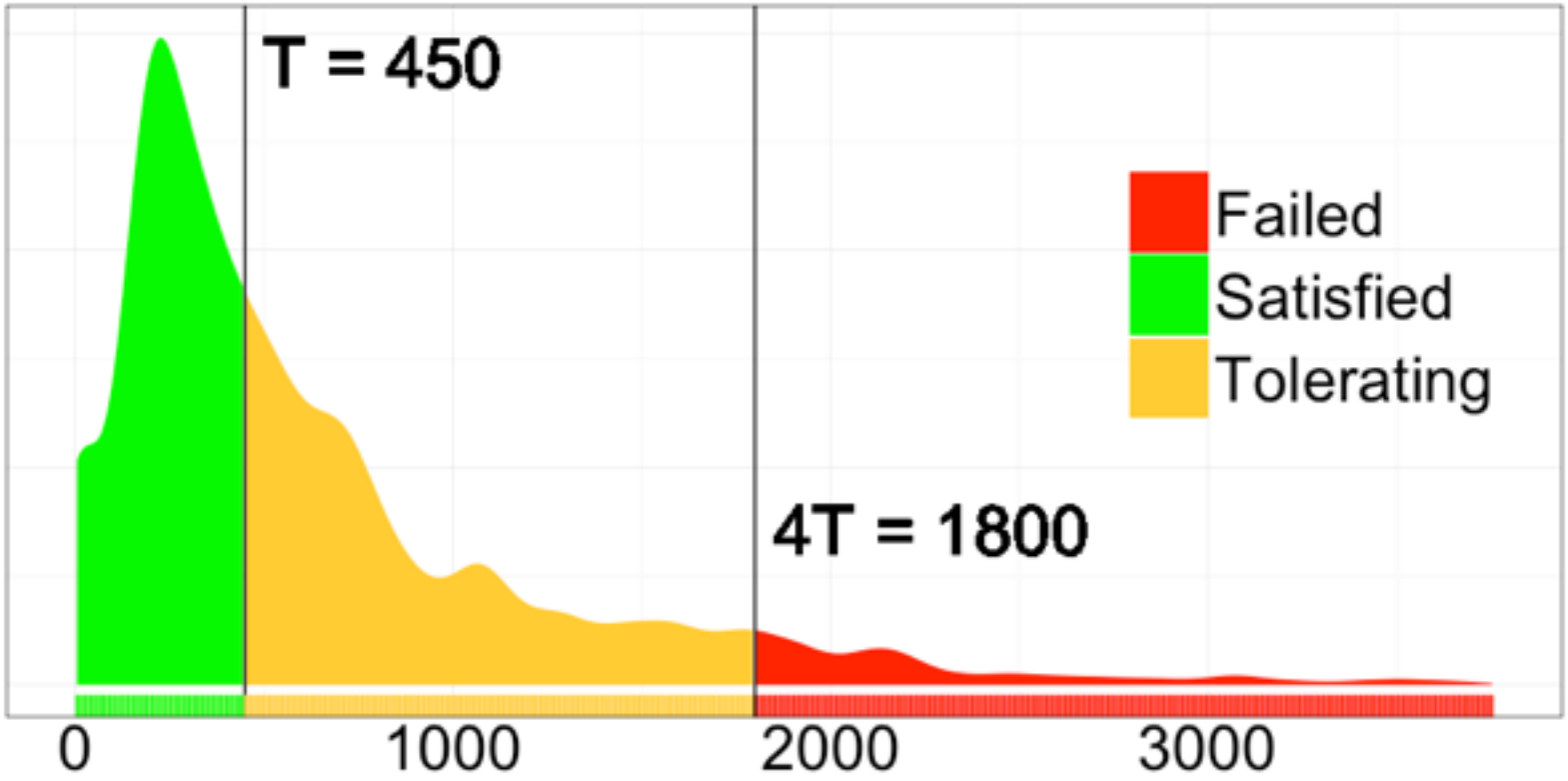
Exploring Apdex

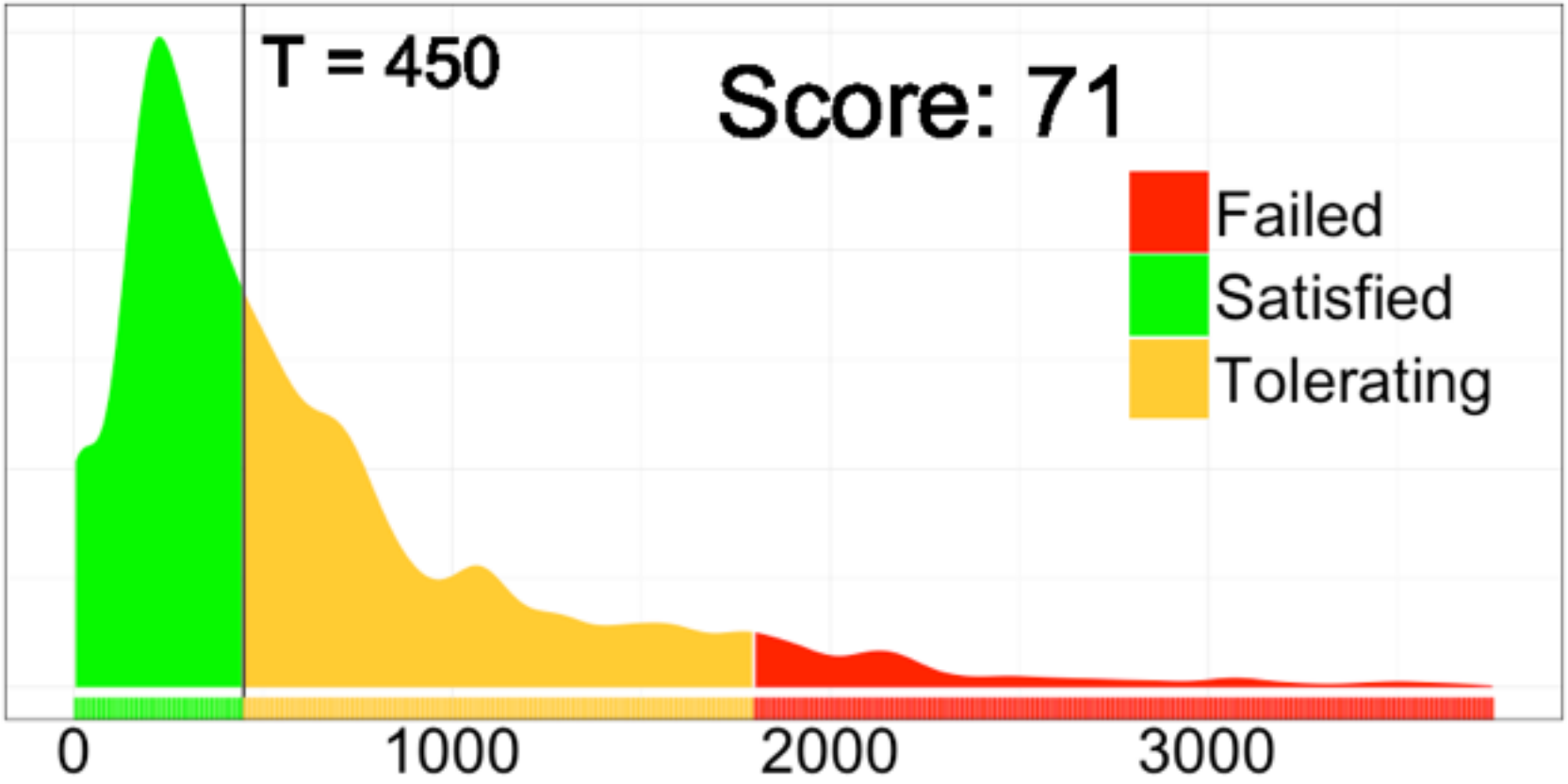
What is Apdex?

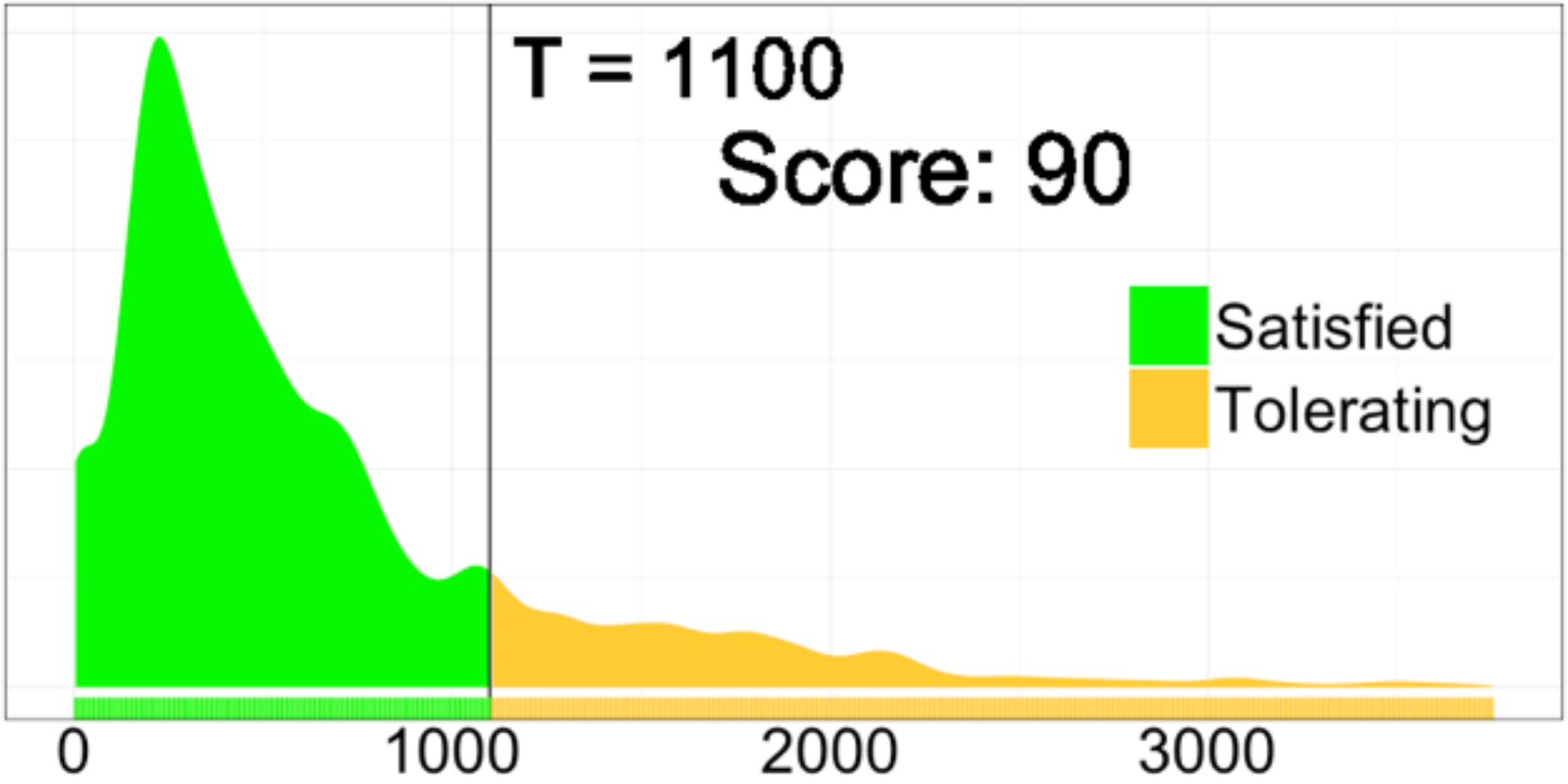
- An Ap… standa… indica… ur site is me… ions.
- Score… (total f… very re…).
- Score… ponse… rding … e thres…
  - **Sa…**
  - **To…**
  - **Fa…**
- Timeouts and errors automatically placed in **Failed** bin.



Apdex is a Three Bucket Histogram

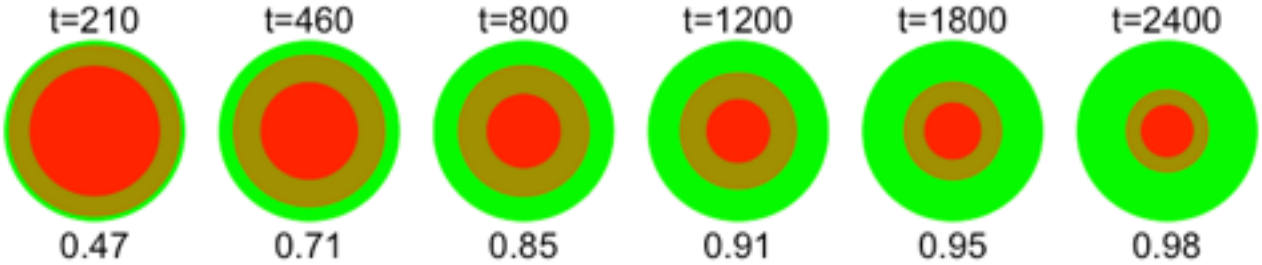






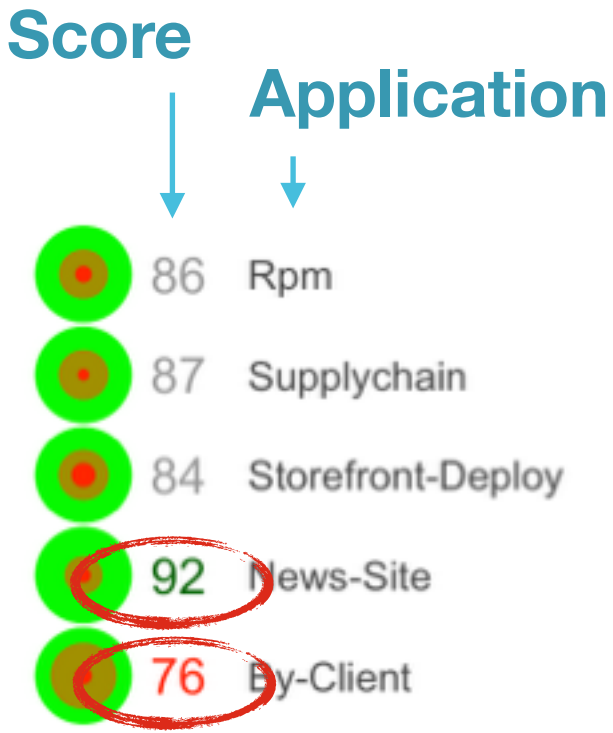
Apdex Buttons

- Apdex Scores for a group of apps can be enumerated with guages, but a richer visual is the "Apdex Button”.

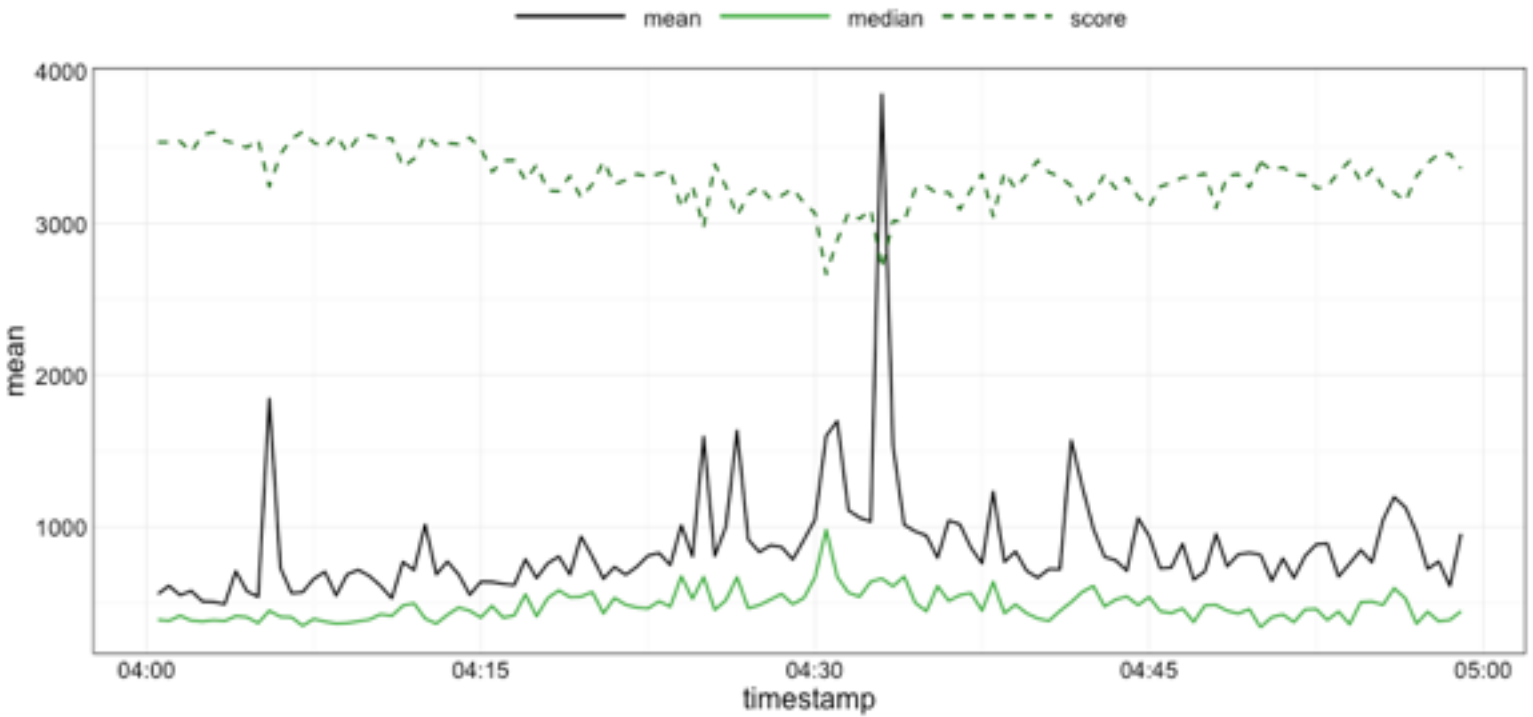


- Each application's status is represented with its own "Traffic Light"
  - **Green** → Good
  - **Red** → Bad

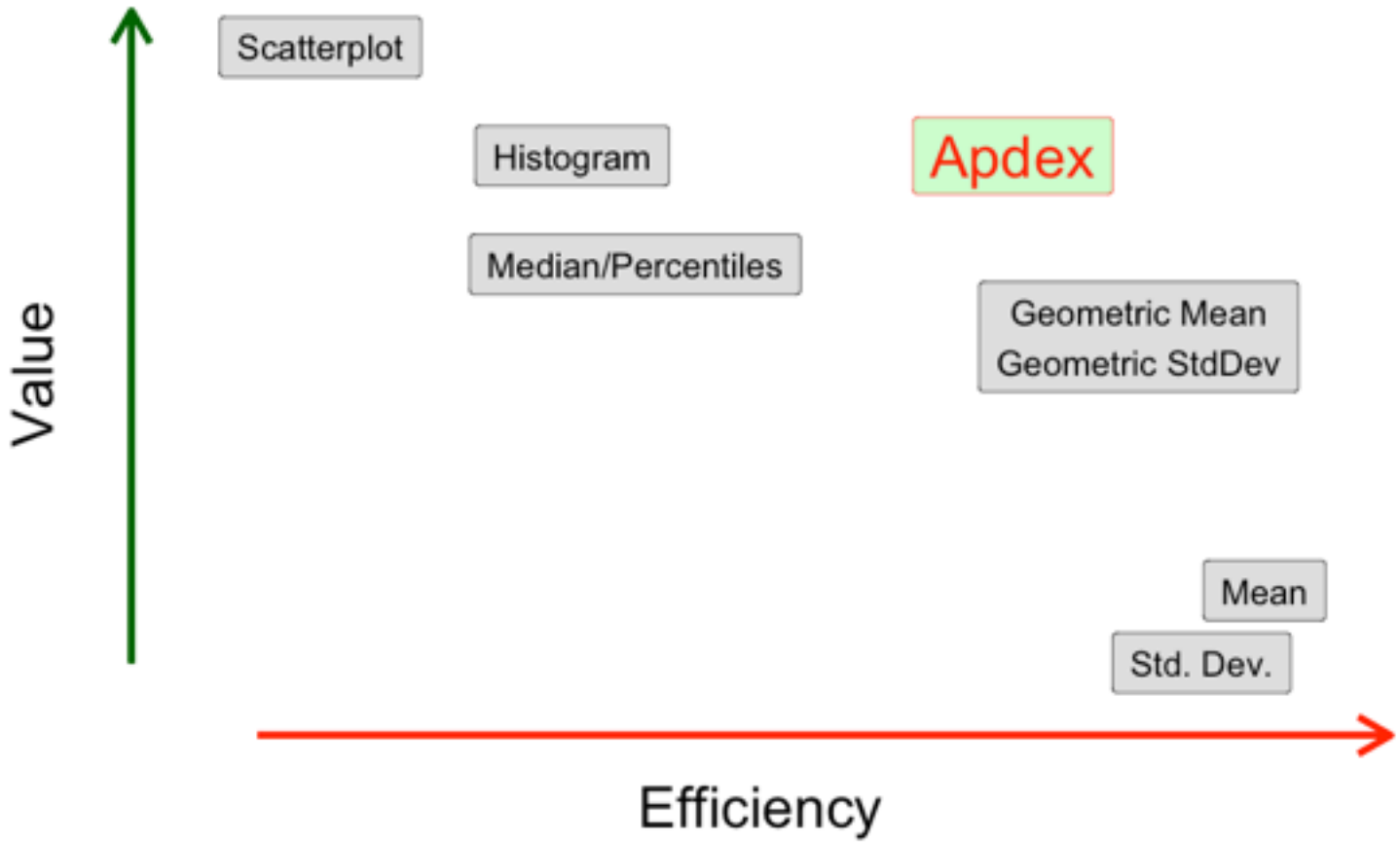
Apdex Buttons in Dashboards



Apdex vs. Mean



New Relic.


Scatterplot
Histogram
Apdex
Median/Percentiles
Geometric Mean
Geometric StdDev
Mean
Std. Dev.
Value
Efficiency

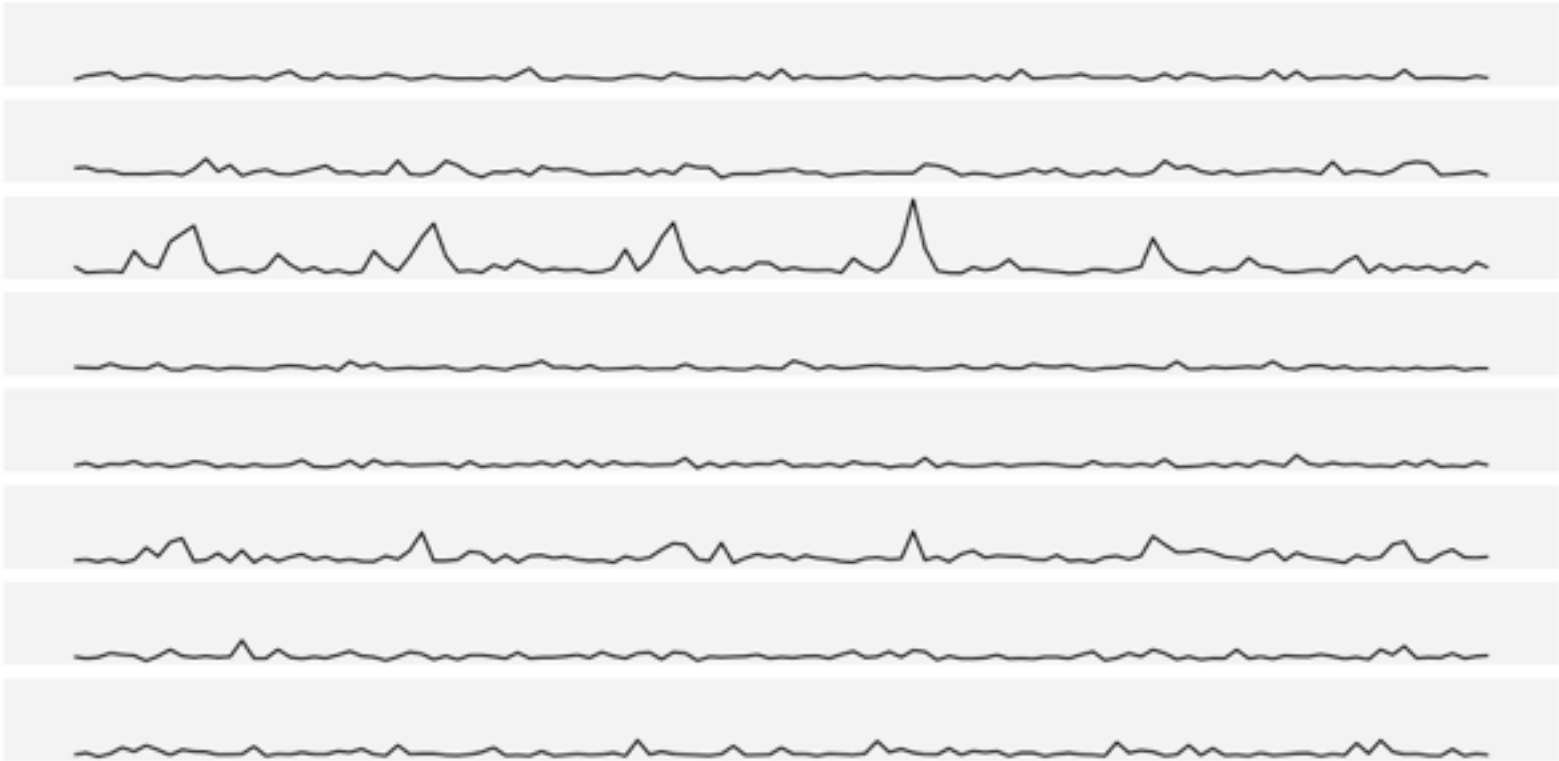
Other Visualizations

Faceted Views

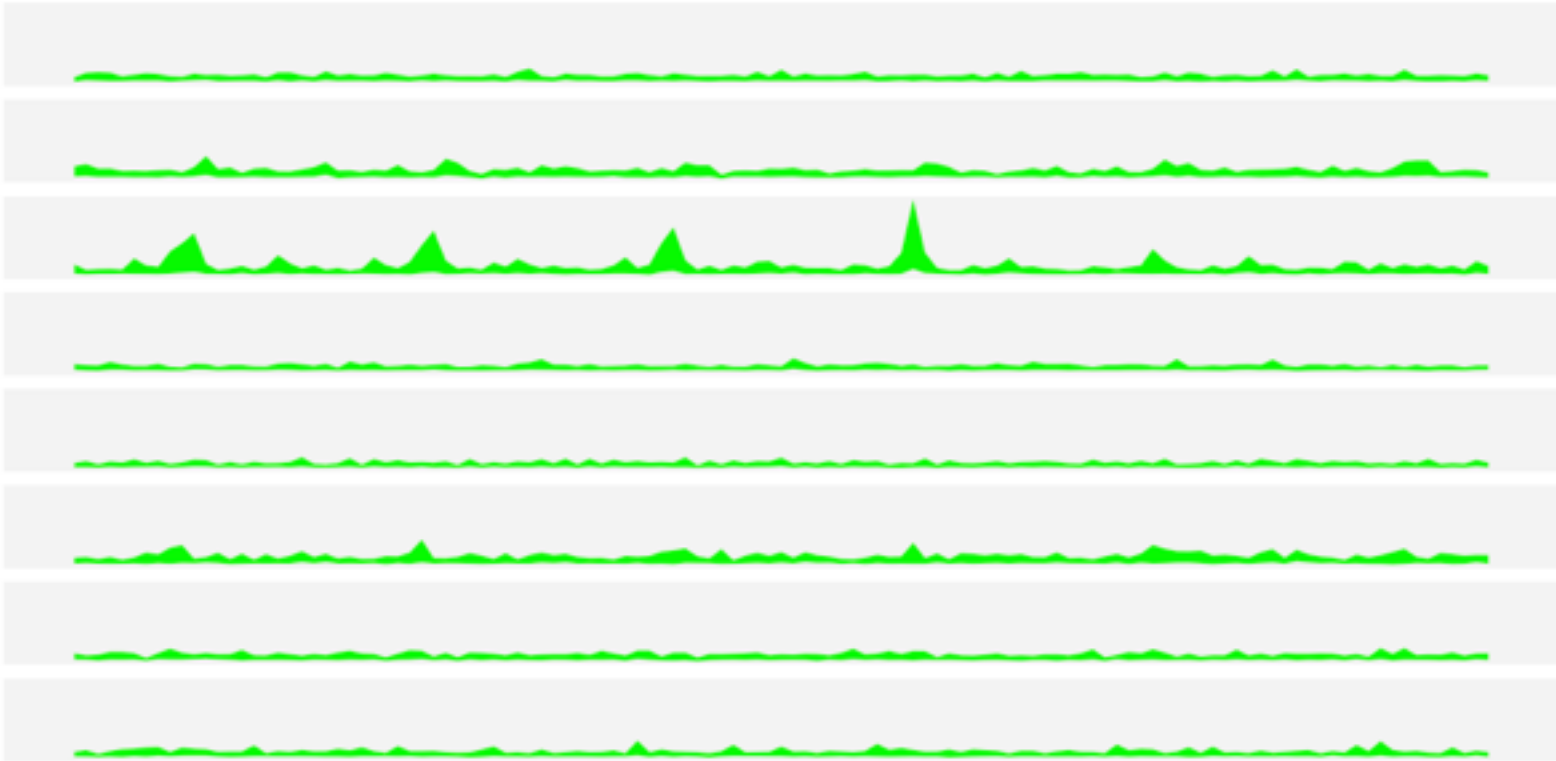
Looking at collection of related things:

- Cluster of servers
- Application server nodes
- Pages in an application
- Response times broken down by browser

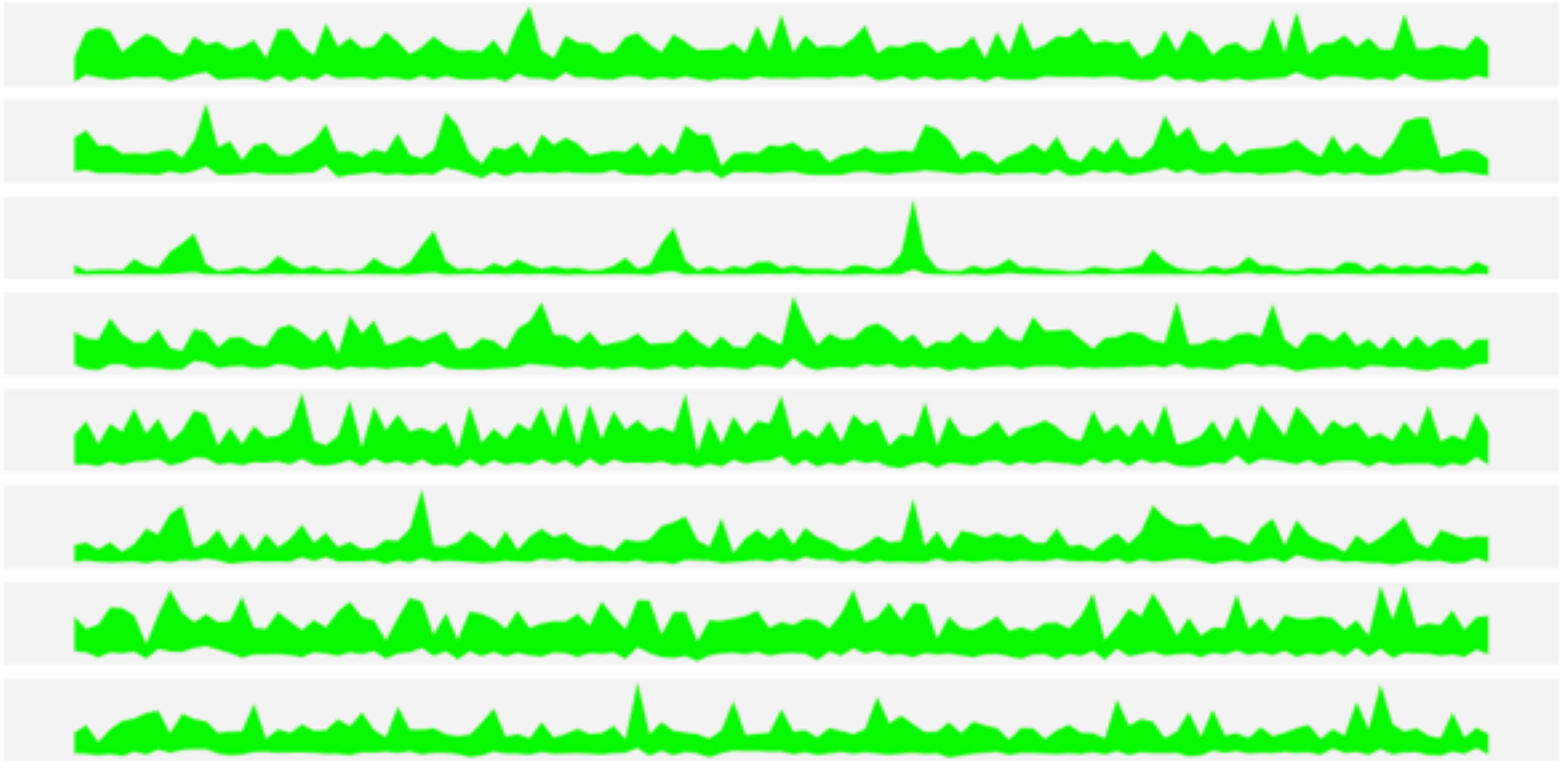
Sparklines - Mean



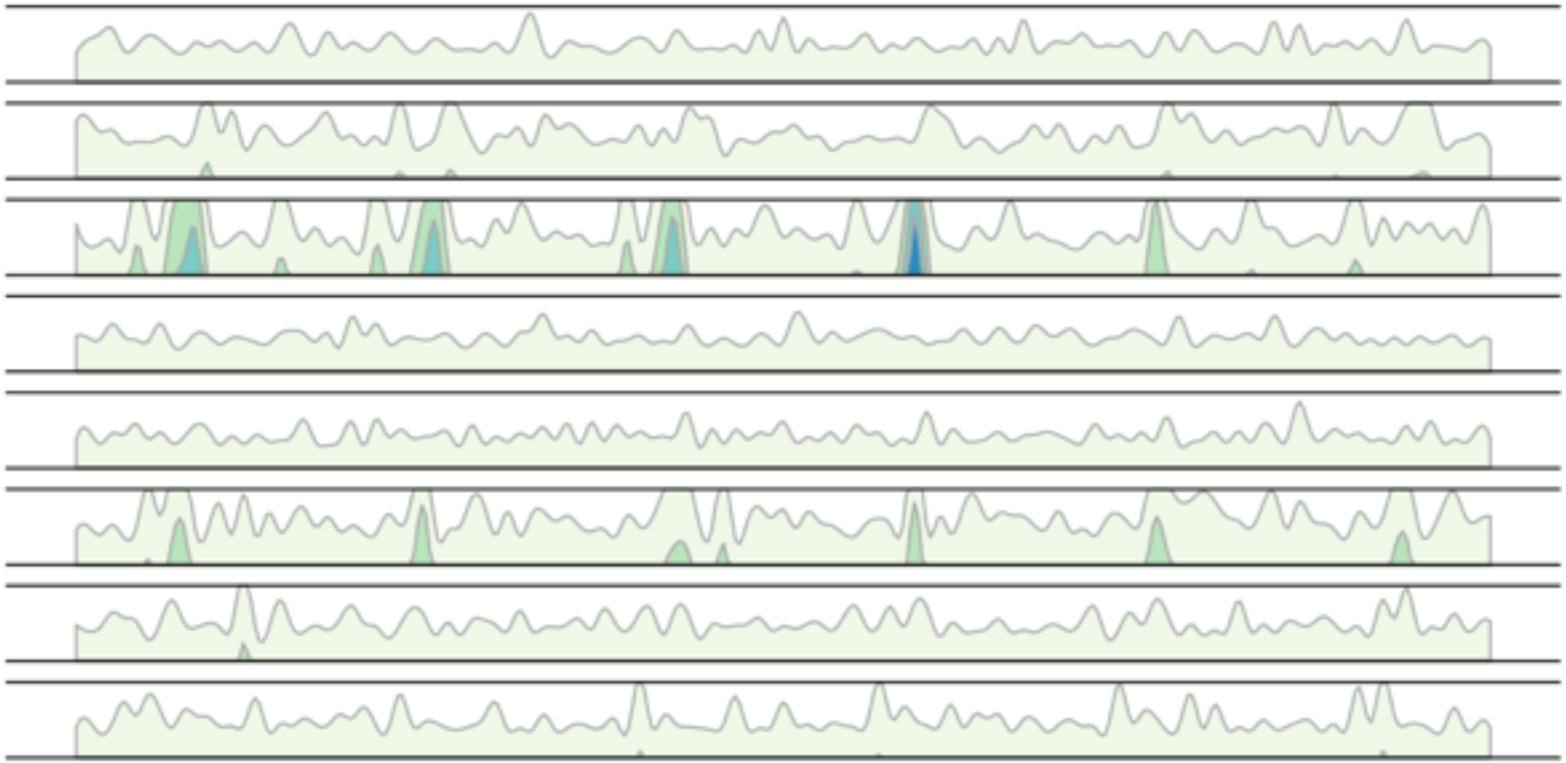
Sparklines - Inner Quartiles



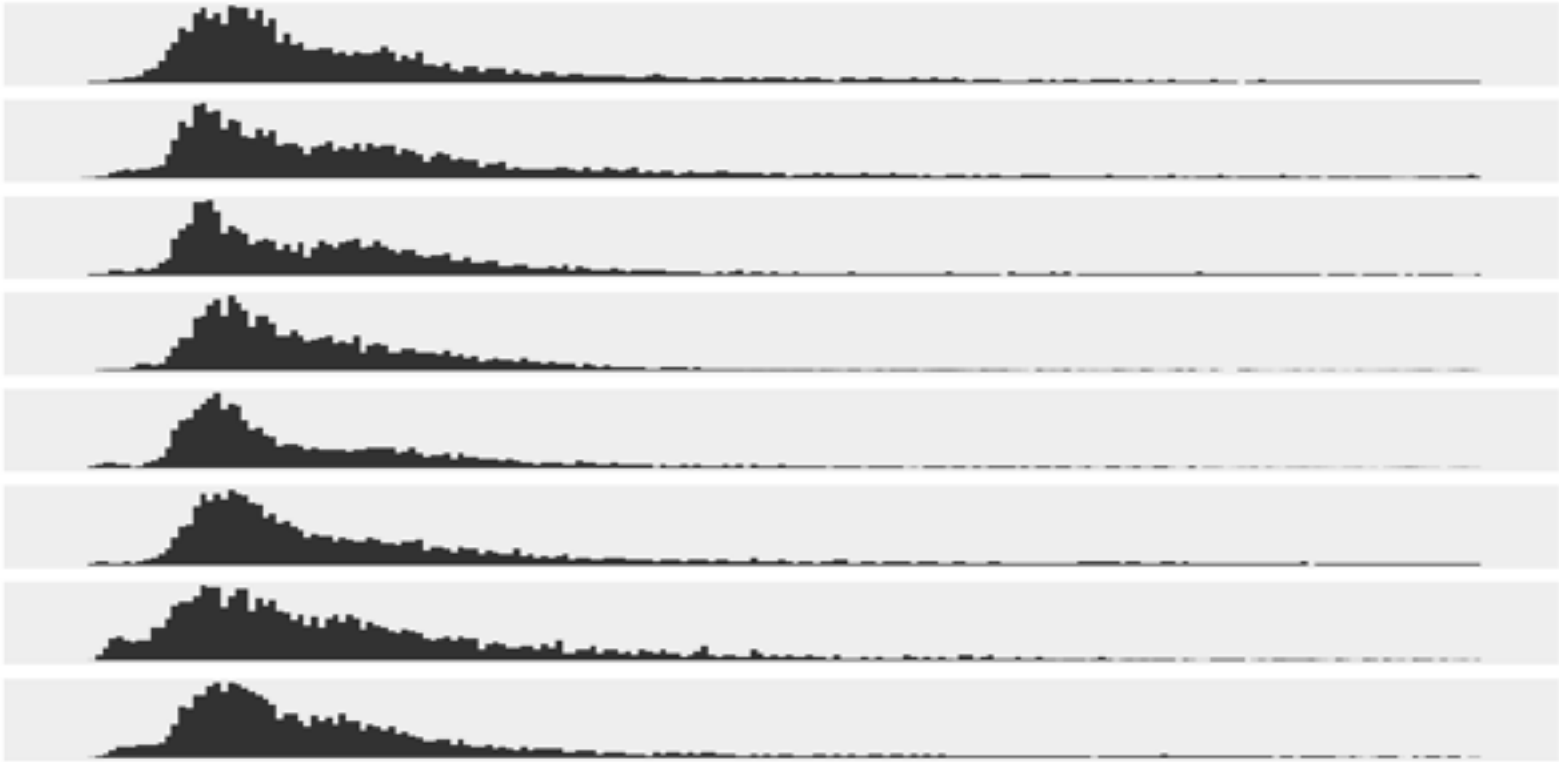
Sparklines - Inner Quartiles, Free Y Axis



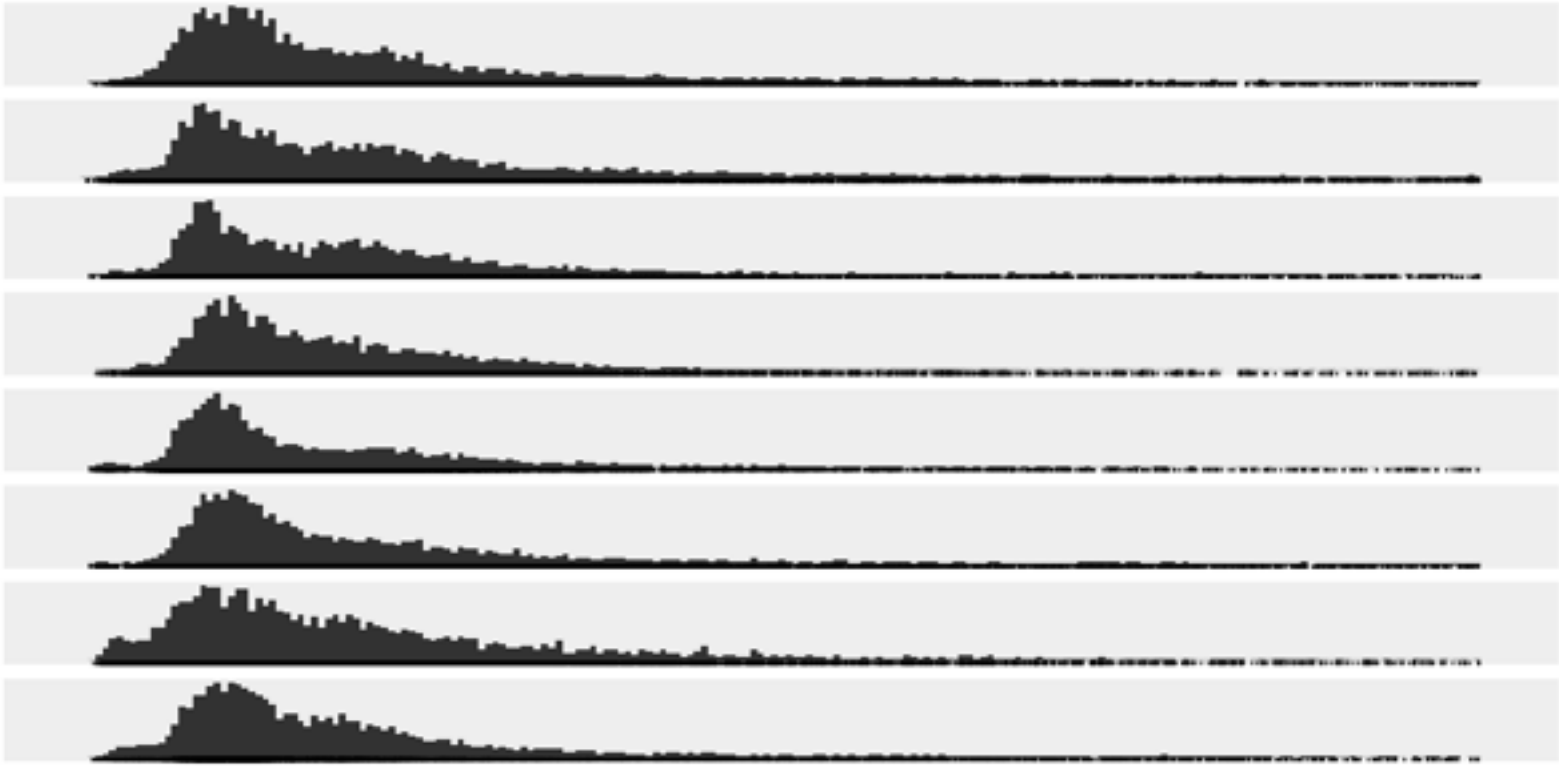
Sparklines - Horizon Plots



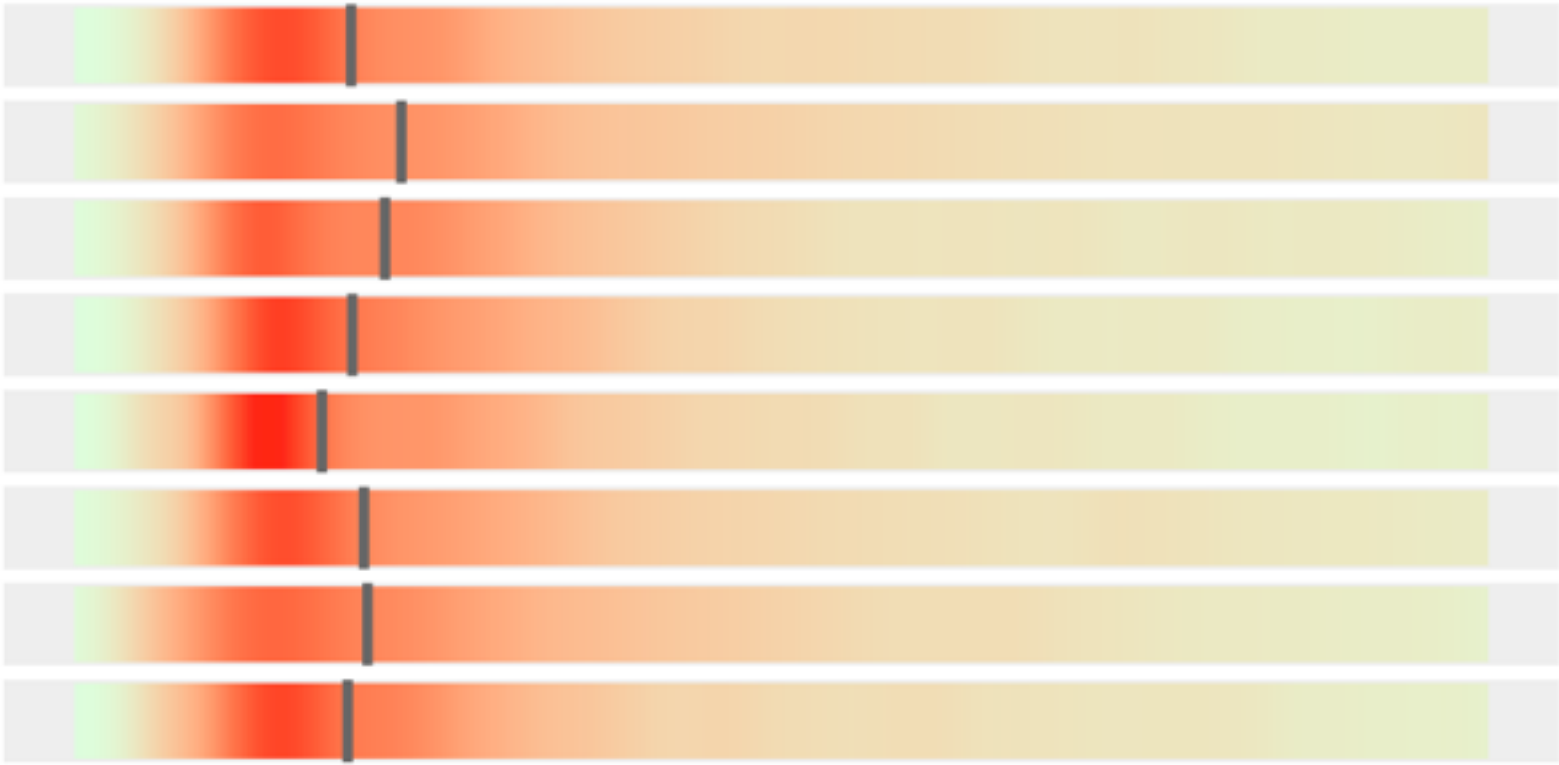
Sparklines - Frequency Plots



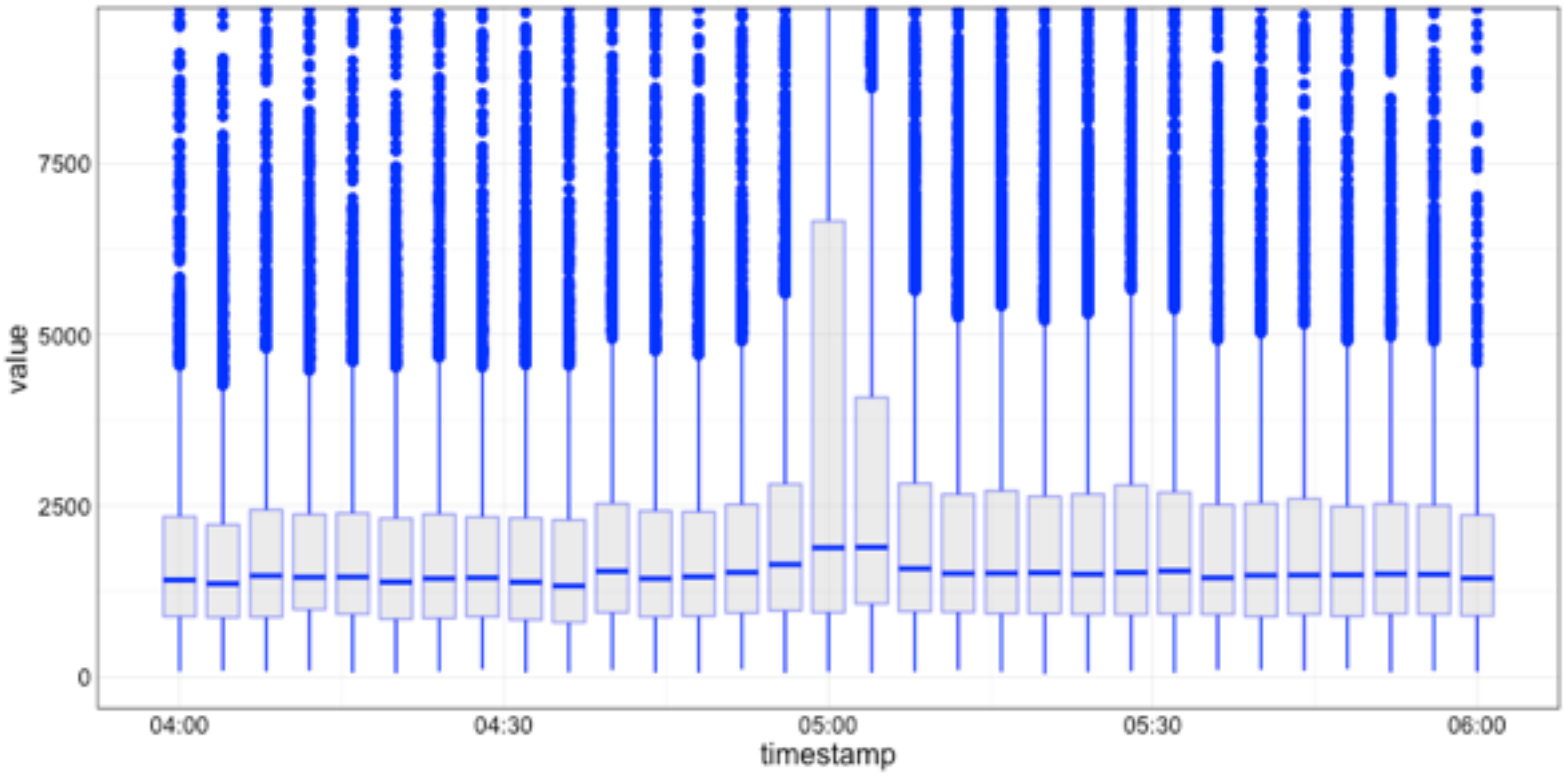
Sparklines - Frequency + Rug



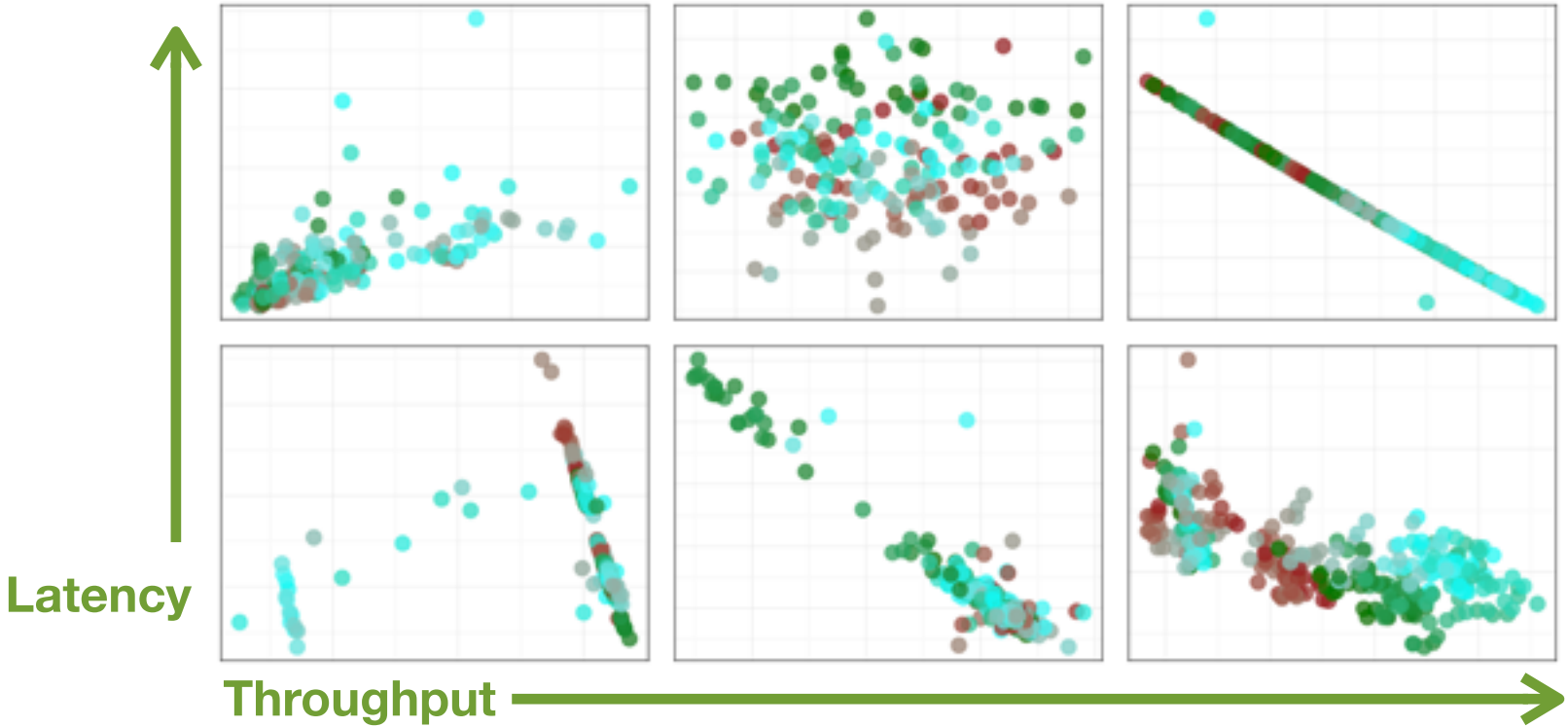
Sparklines - Density Filament



Box Plots

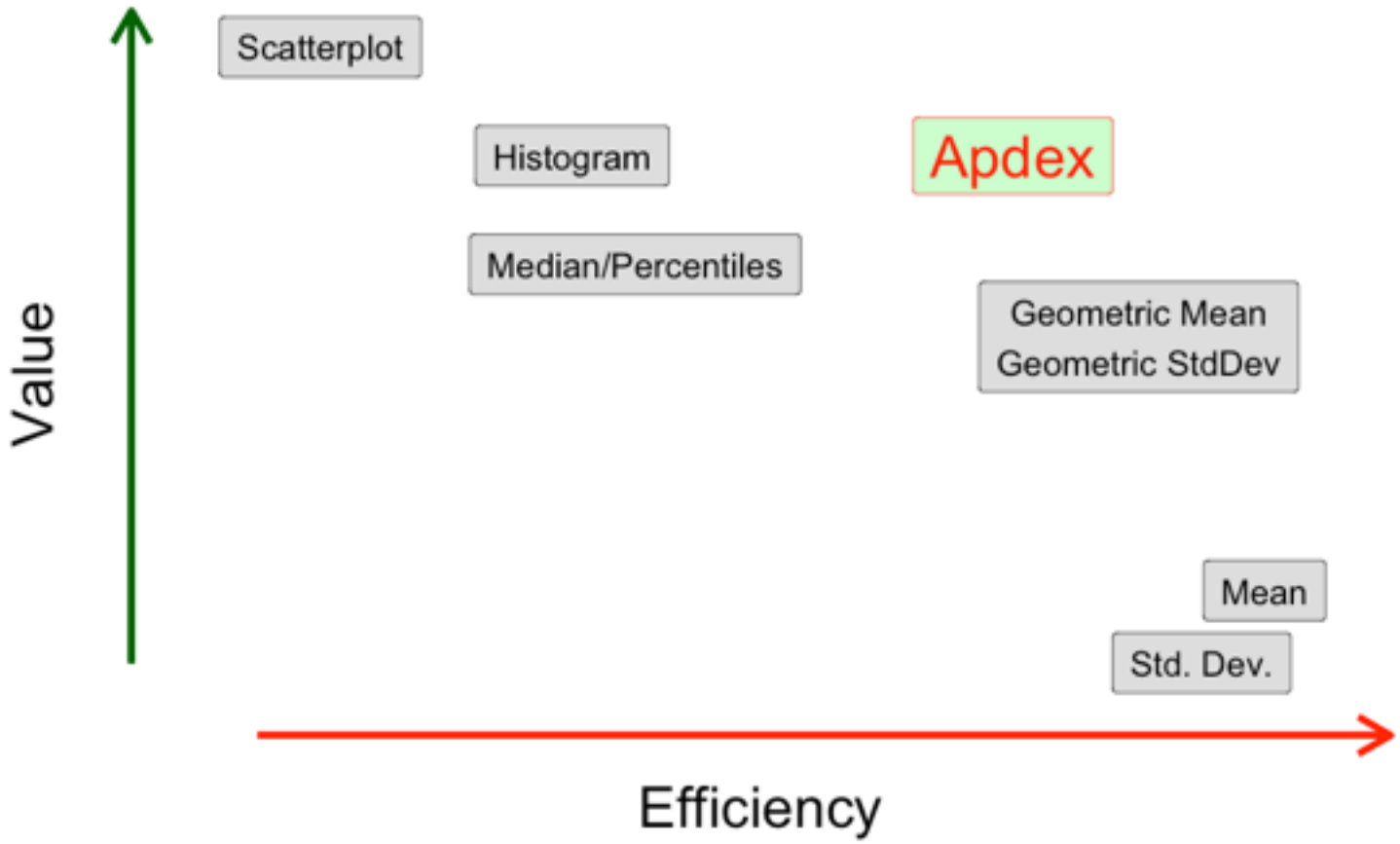


Throughput vs. Response Time



Traffic Animation Demo

New Relic.

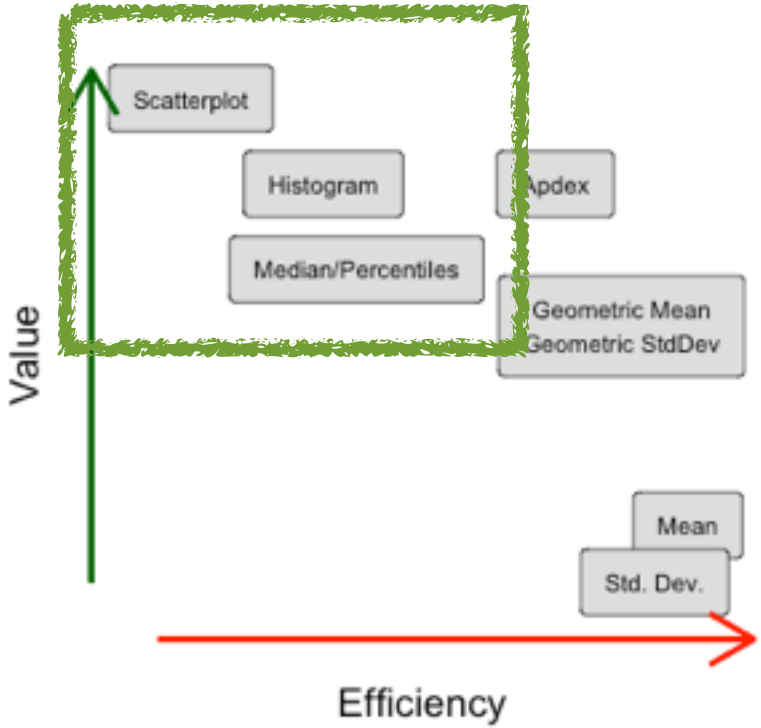

Scatterplot
Histogram
Apdex
Median/Percentiles
Geometric Mean
Geometric StdDev
Mean
Std. Dev.
Value
Efficiency

Summary

Monitoring Latency - When You Can Store a Lot of Data

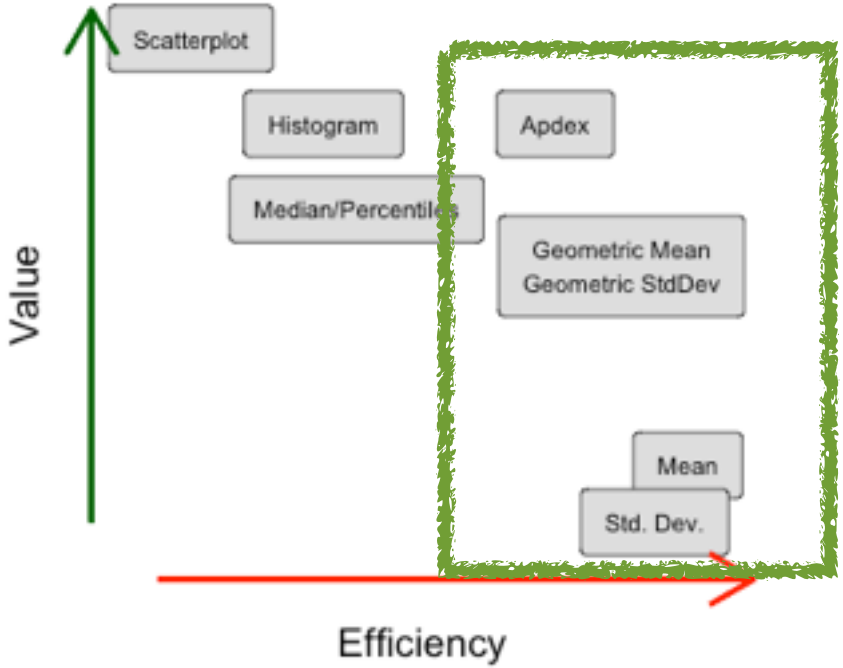
When you can collect a lot of data:

- Use Scatterplots for deep visibility
- Use Median, Inner Quartiles for plots and sparkcharts
- Use histograms to identify multi-modal behavior
- Collect multiple dimensions for faceted views when possible



Monitoring Latency - When Space is a Premium

- Be careful using the mean
- Use Geometric Mean to approximate the median
- Use Geometric Standard Deviation to approximate inner quartile regions
- Use Apdex for monitoring the quality of the customer experience



Monitoring Latency - Consider Alternative Visualizations

- Throughput vs Response Time Scatterplots
- Density Filaments
- Horizon Plots
- Animation

Thanks!

- Bill Kayser (@bravoking) - www.newrelic.com
- This talk: **bkayser.github.io/apmviz**
- Source for this talk: **github.com/bkayser/apmviz**
- D3/Javascript version live demo: **http://marlowe.datanerd.us**
- D3/Javascript source: **github.com/newrelic/marlowe**
- R Library for getting New Relic Data: github.com/bkayser/NewRelicR