elastic{on} 2017

Machine Learning and Statistical Methods for Time Series Analysis

Tom Veasey, Software Engineer, Machine Learning
Steve Dodson, Tech Lead, Machine Learning

Elastic

8th March 2017

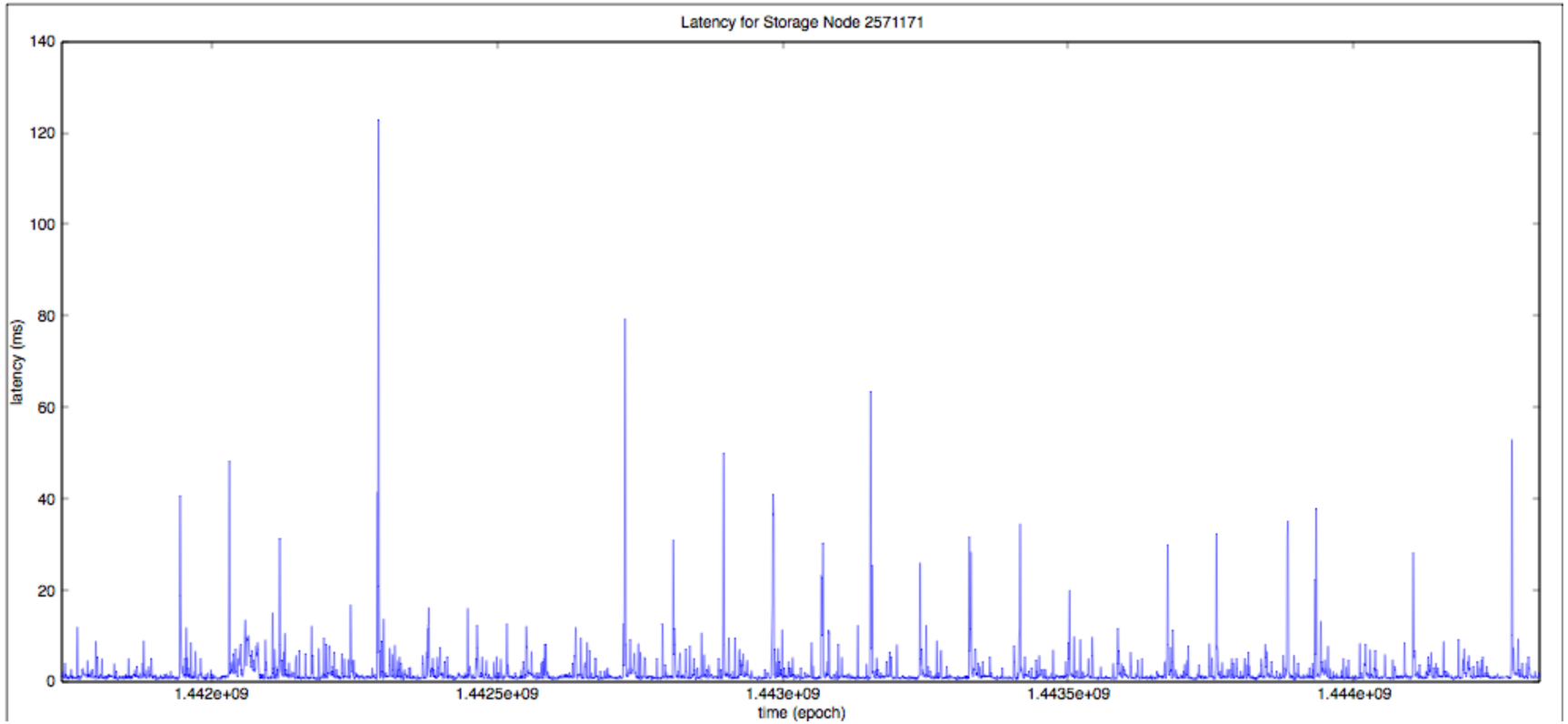
@prelertsteve

Agenda

1. Introduction
2. Machine Learning Overview
3. Time Series Anomaly Detection
4. Demo

Introduction

- Problem Overview and Technology Background
- Problem Examples



Machine Learning Overview

Supervised Machine Learning

Given (Input, Output) pairs, i.e. (x, y)

$y'($  $| \theta) = \text{cat}$

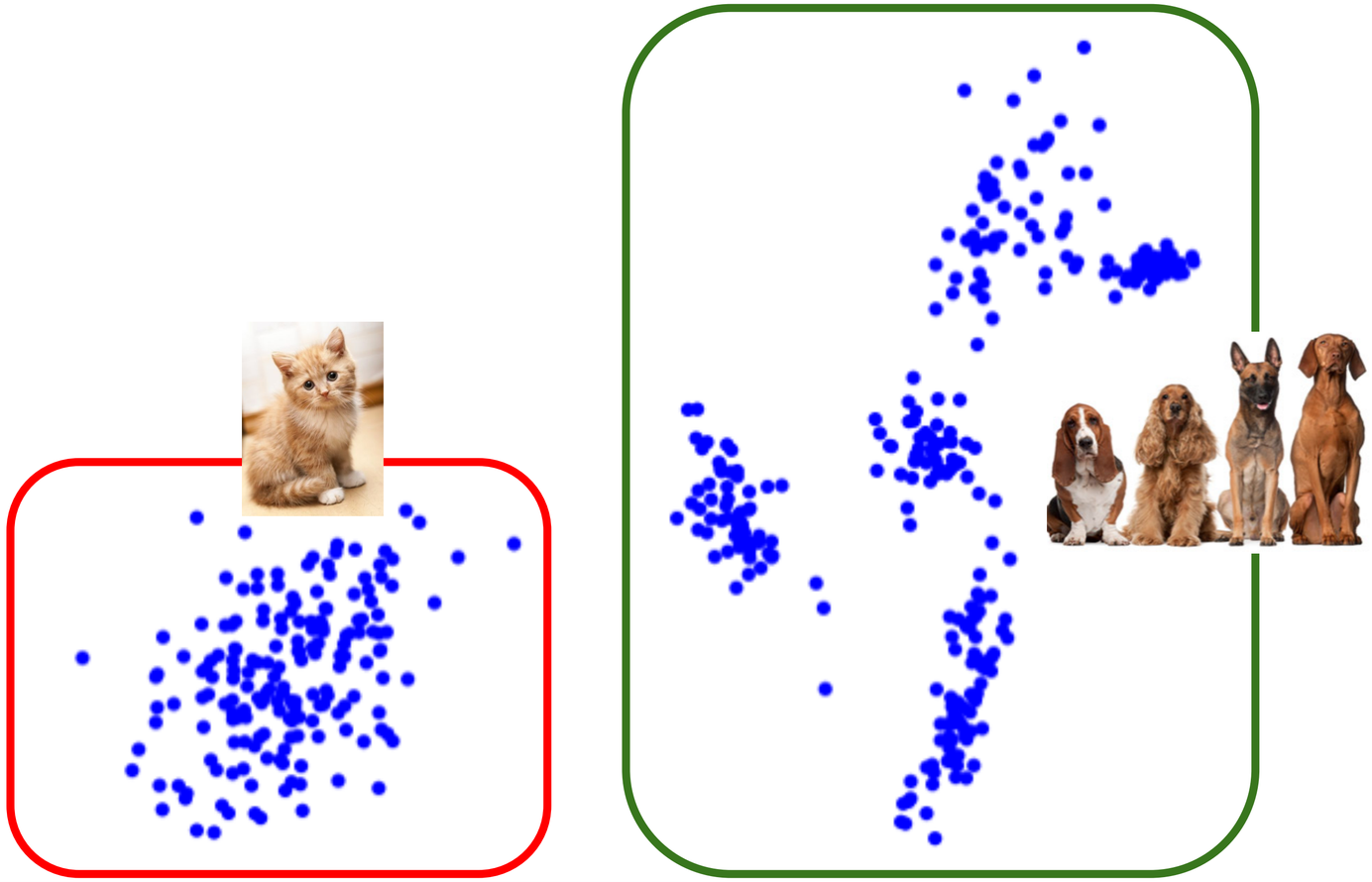
$y'($  $| \theta) = \text{dog}$

Learn θ by minimizing measure of prediction error, i.e. "$|| y - y'(x | \theta) ||$"

Unsupervised Machine Learning

Aren’t told the outputs

What objective does it make sense to target in this case?



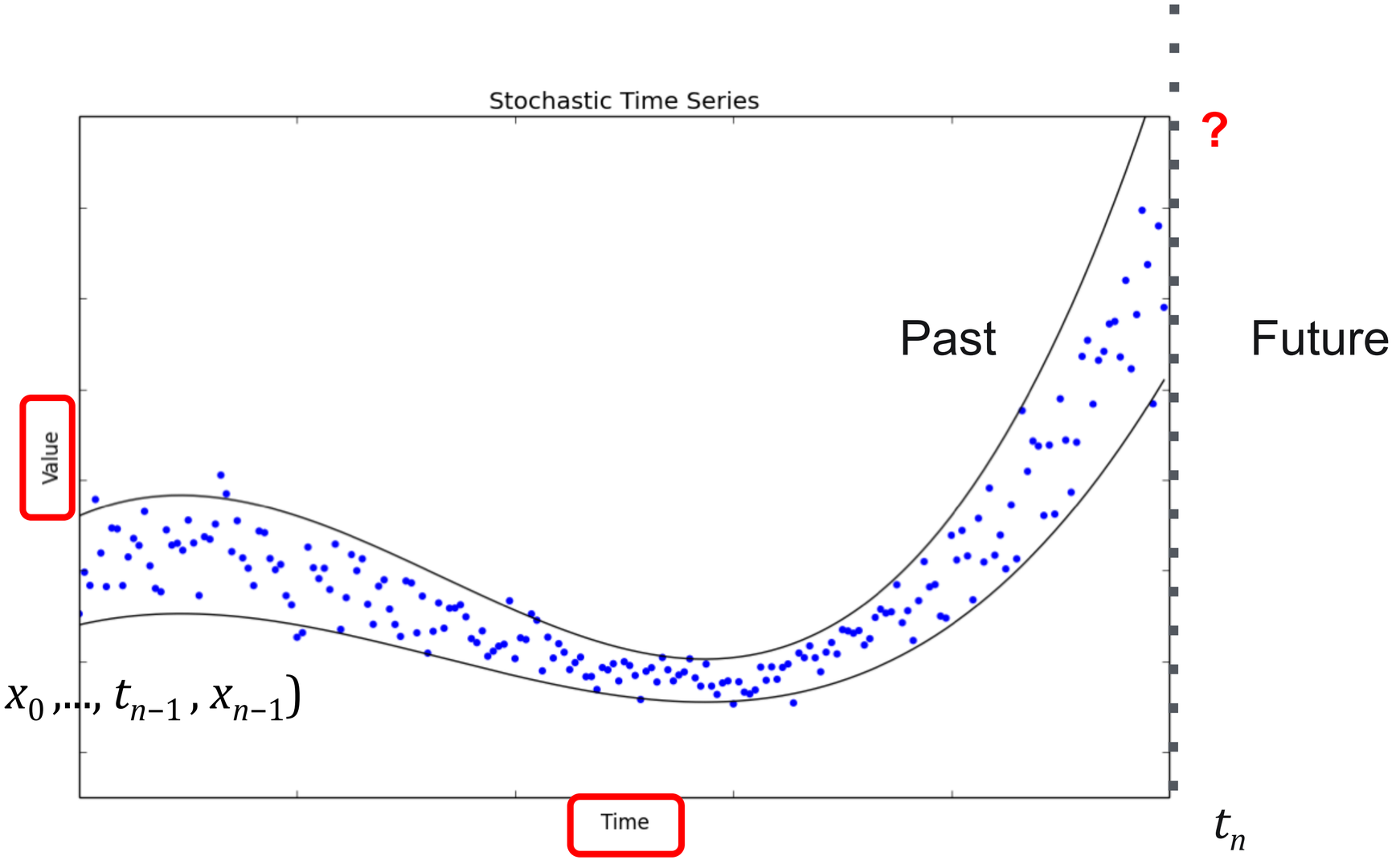
Anomaly Detection

This is useful for anomaly detection

> *“ values which don’t fit the model are the things of interest ”*

Time Series Anomaly Detection

Time Series Anomaly Detection



Estimate $X_n \sim f\,(x_n | t_n\,,\, t_0\,,\, x_0\,,\ldots,\, t_{n-1}\,,\, x_{n-1})$

Time Series Anomaly Detection

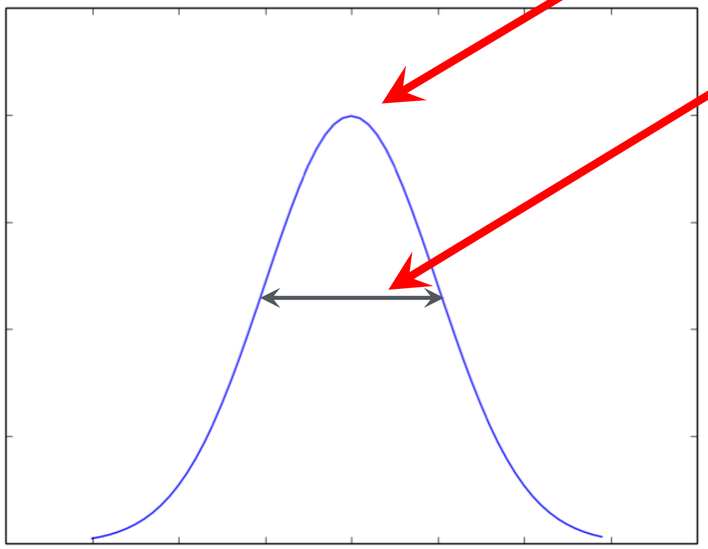
Break problem down

$$f(x_n | t_n, t_0, x_0, \ldots, t_{n-1}, x_{n-1}) = f'(x_n | \theta_0(w_0(t_n) \odot x), \ldots, \theta_m(w_m(t_n) \odot x))$$

+ assumption

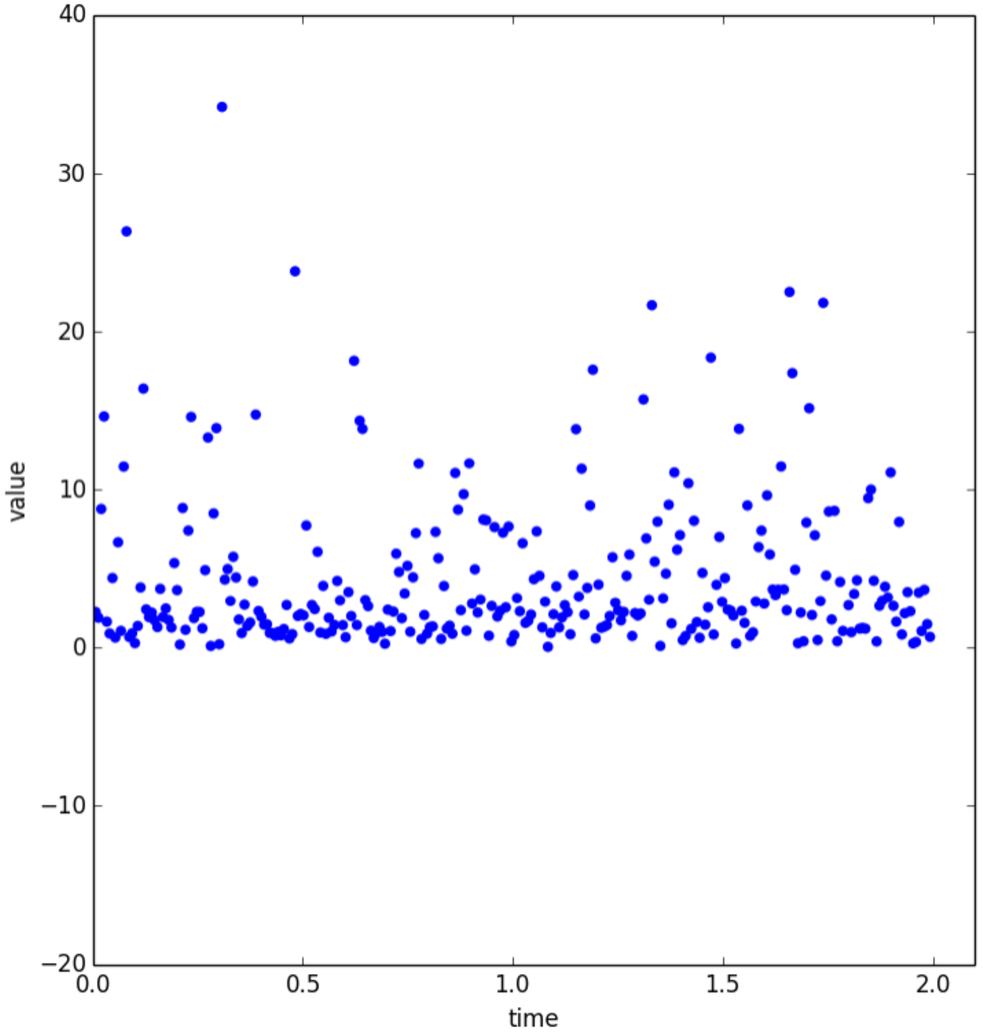
mean

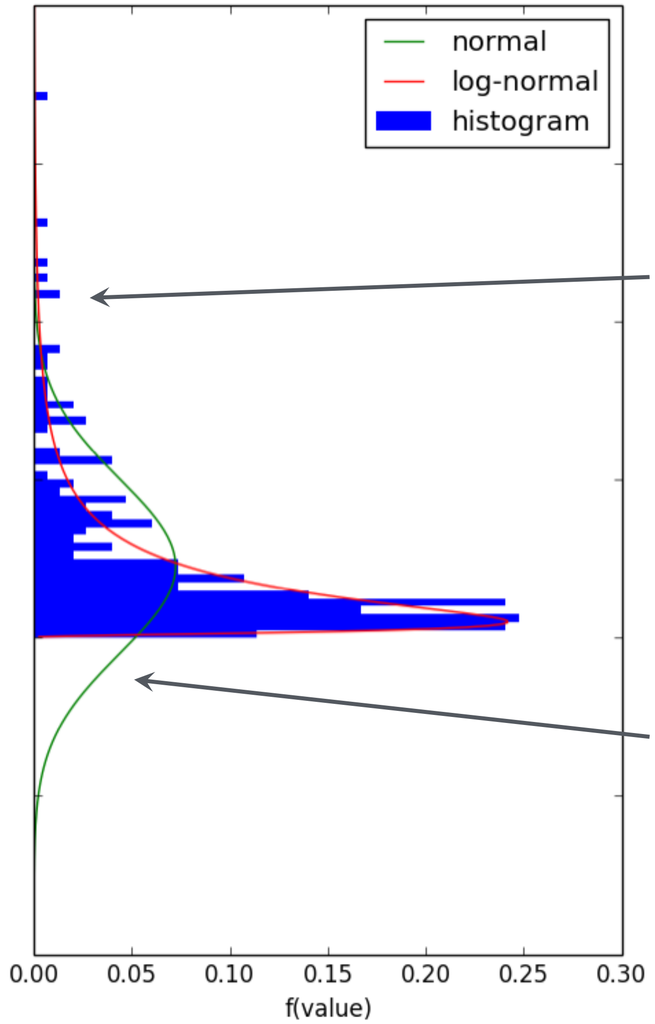
“variance”



$$f(x_n) = 1/\sqrt{2\pi\sigma}\ \exp(-1/2\sigma^2(x_n - m)^2)$$

Stationary Unimodal Example



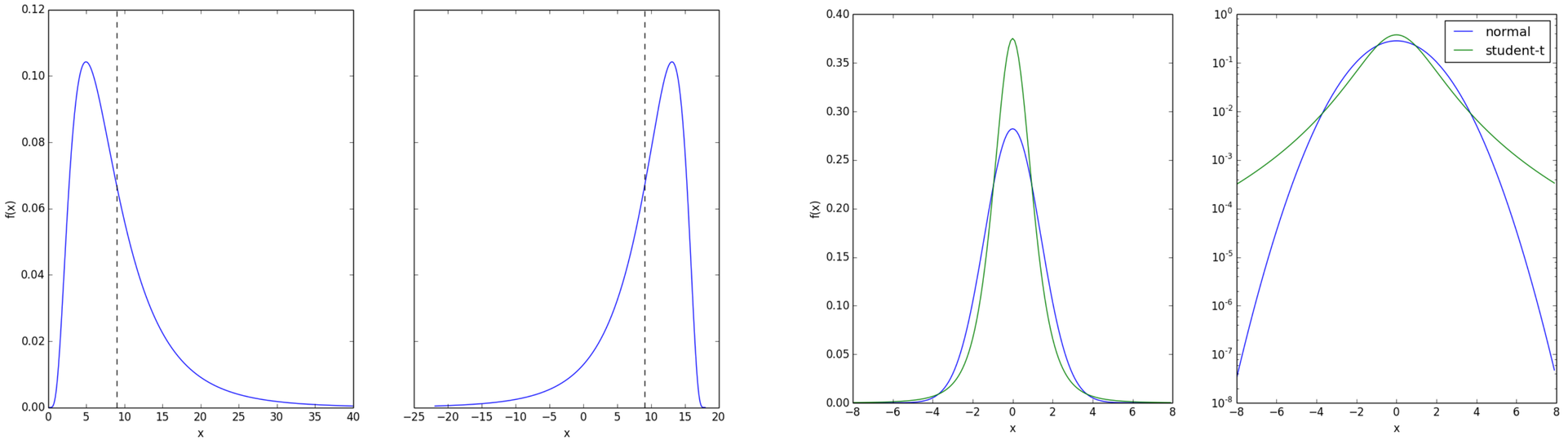


False positives

False negatives

Stationary Unimodal Summary

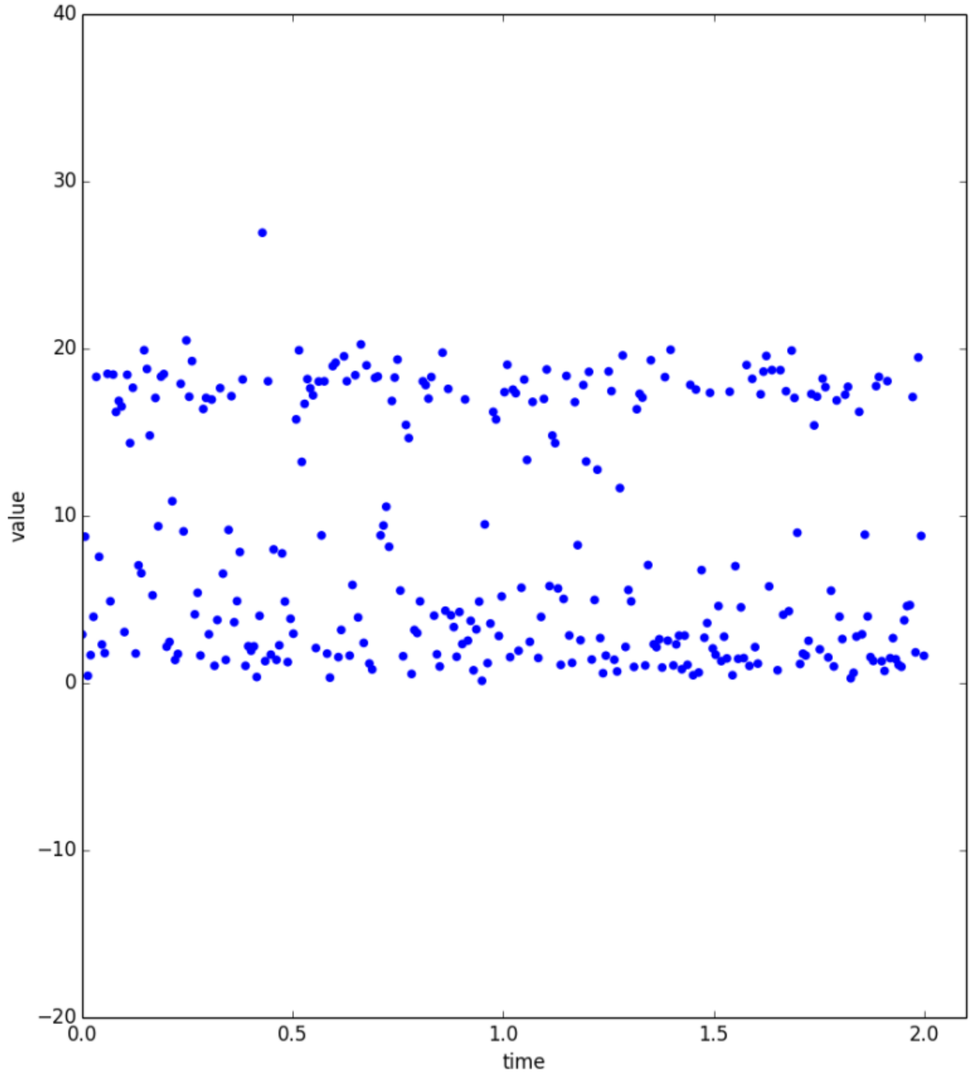
$f'(x_n|\theta_0(w \odot x),...,\theta_m(w \odot x))$ $[w]_i = 1$ for all i.

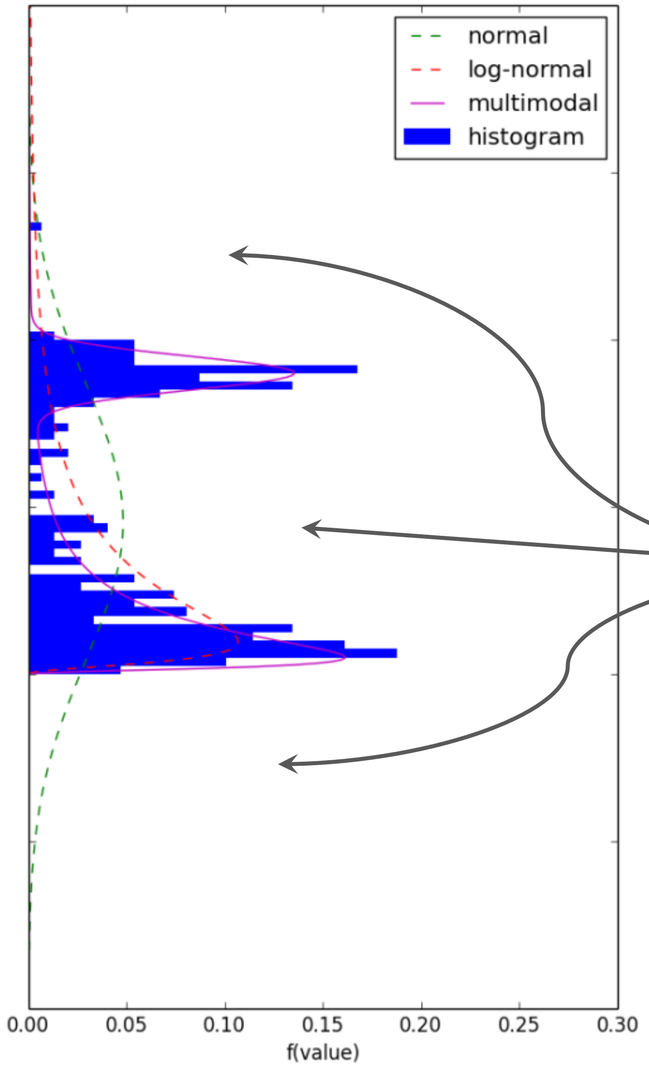


Skewness, heavy tails, etc = “Model Selection”.

Non-parametrics: don’t want to assign “mass” at all data points.

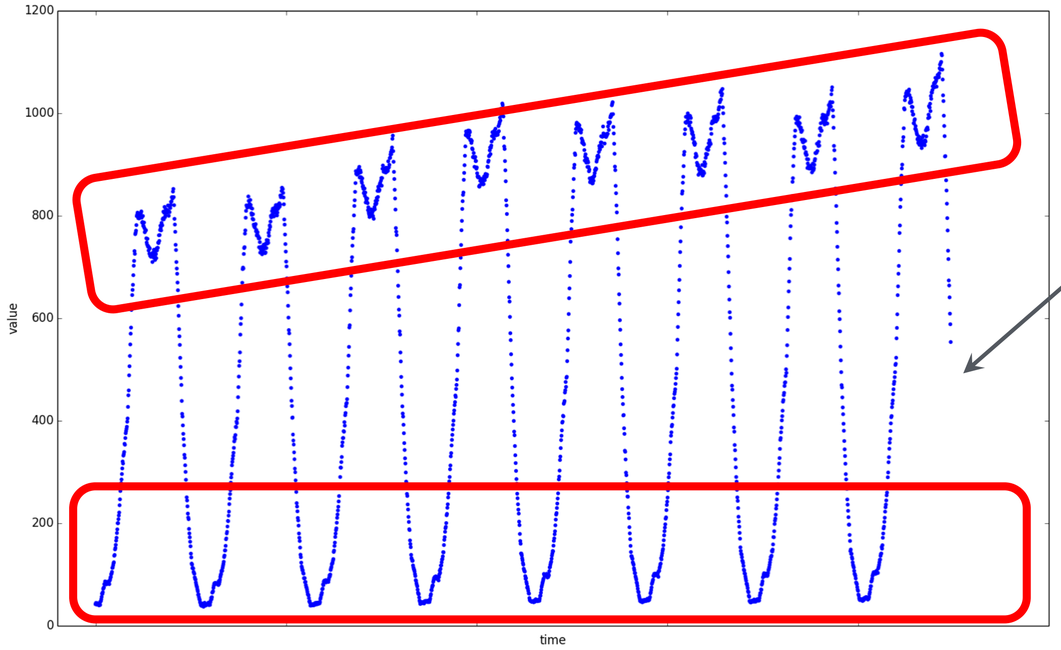
Stationary Multimodal Example



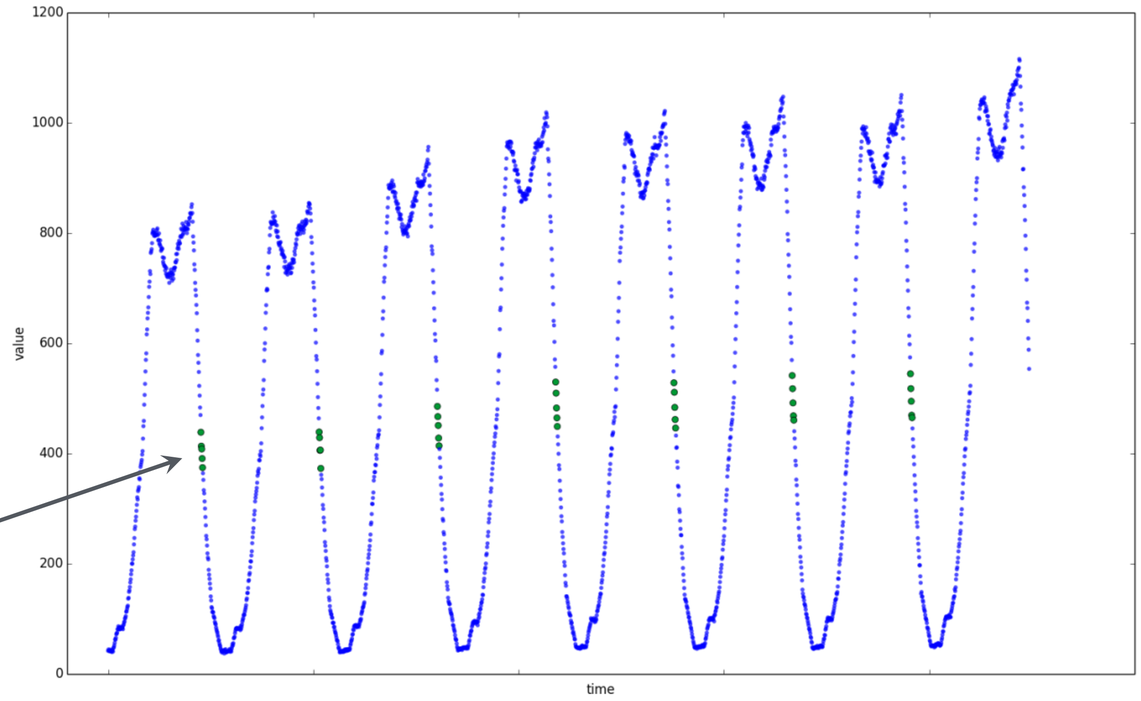


False negatives

Seasonality Example

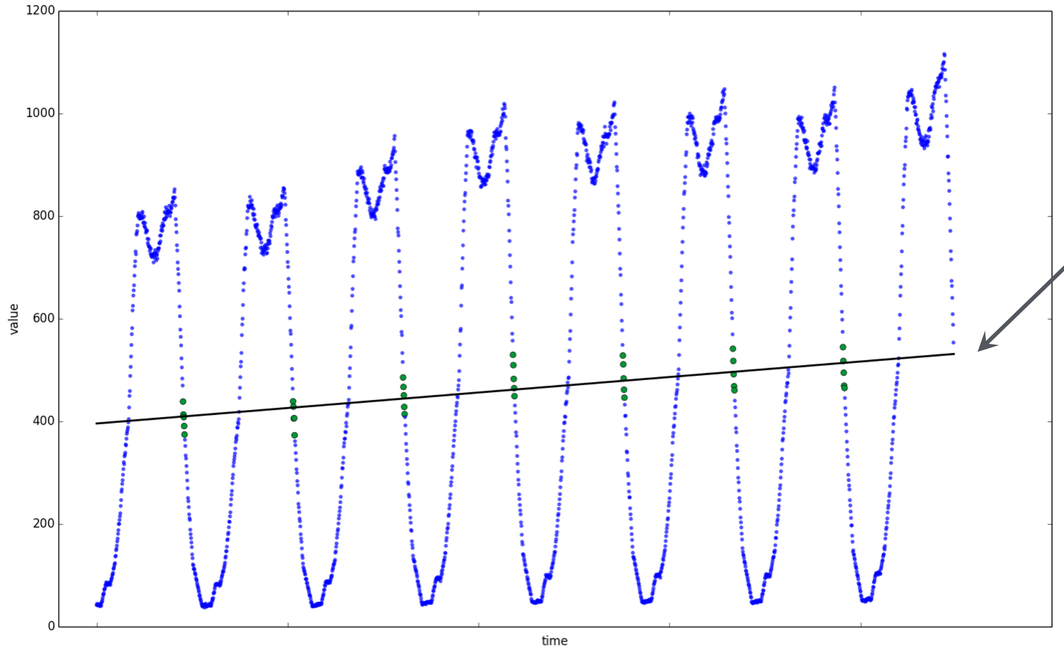


Want to predict the next value

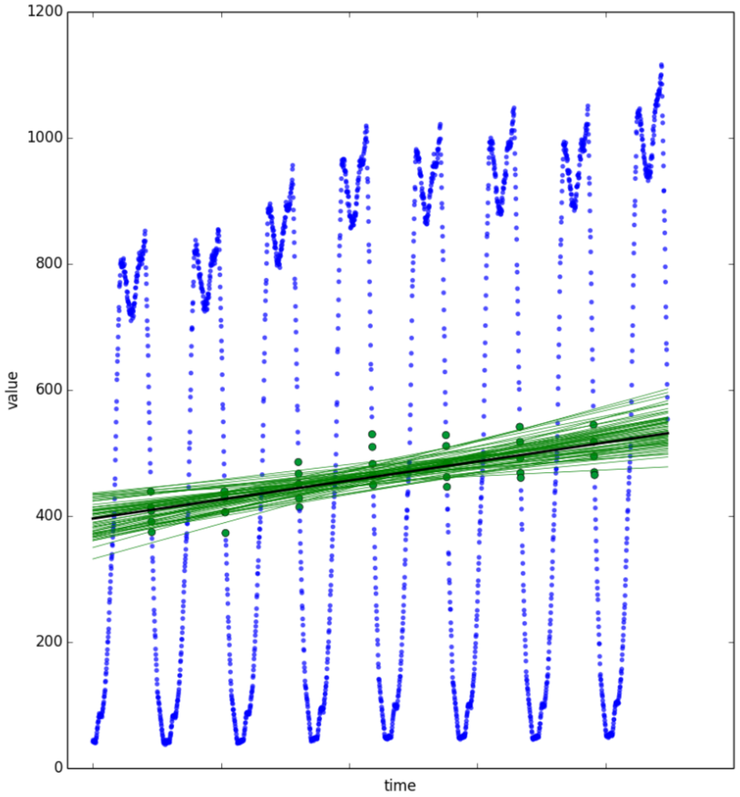


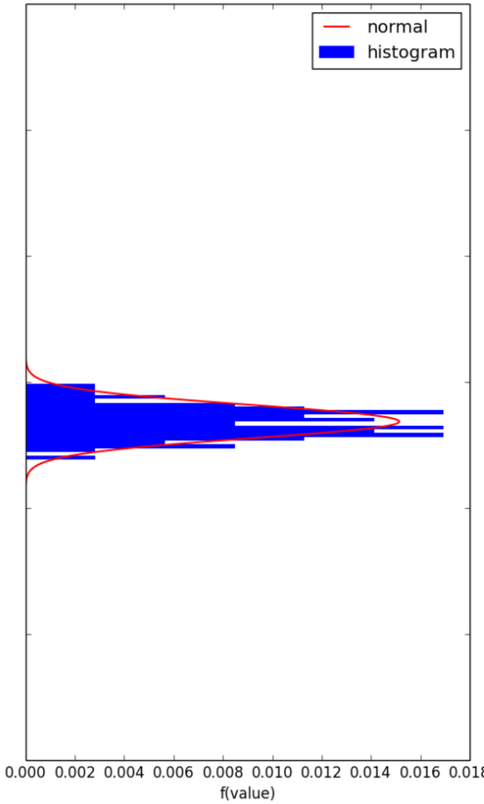
Subset of historical points relevant

Seasonality Example



Best fit OLS





Seasonality Summary

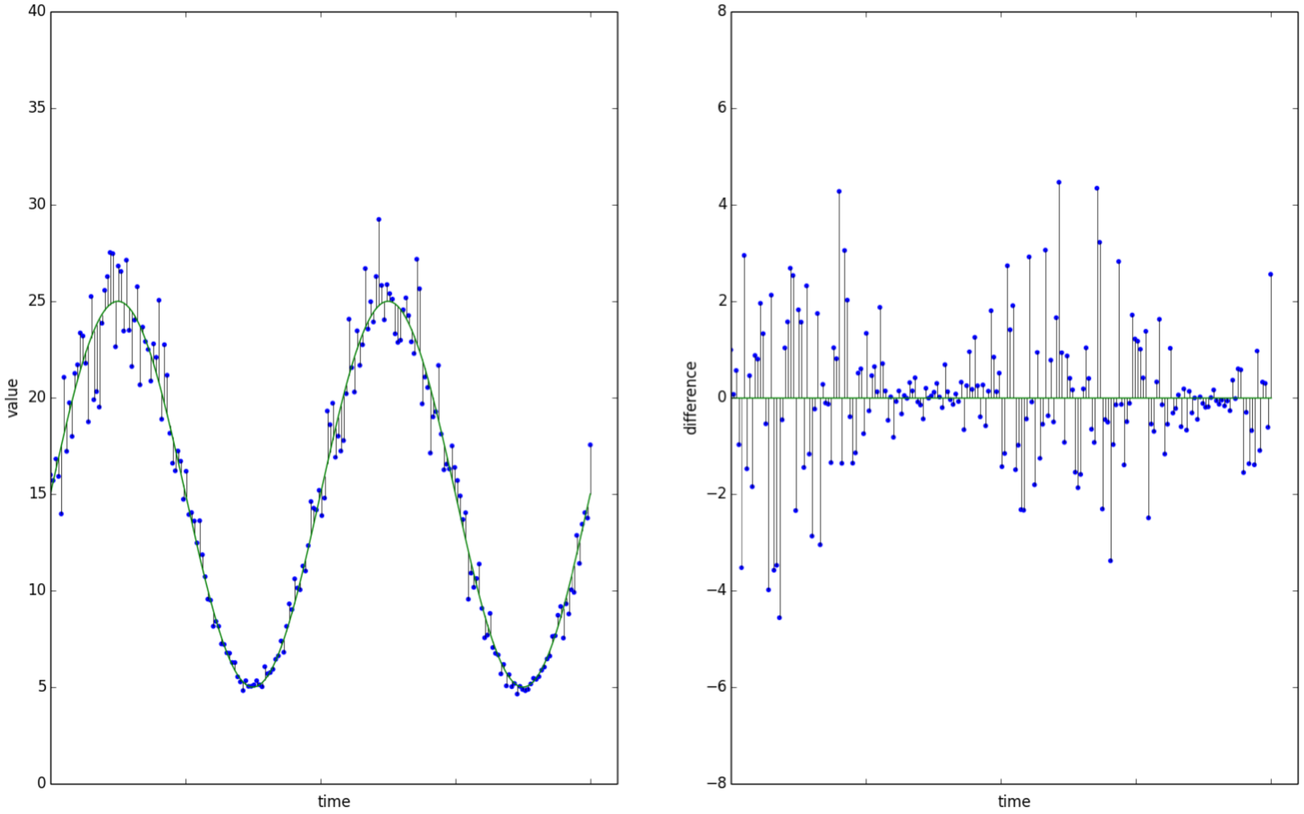
Data pattern repeated *regularly in time*.

$$f'(x_n|\theta_0(w(t_n)\odot x),\ldots,\theta_m(w(t_n)\odot x))$$

$$[w(t)]_k = \sum_i 1\{t_k \in [t-iT-\varepsilon, t-iT+\varepsilon]\}$$

Period is T and interval $\varepsilon \ll T$

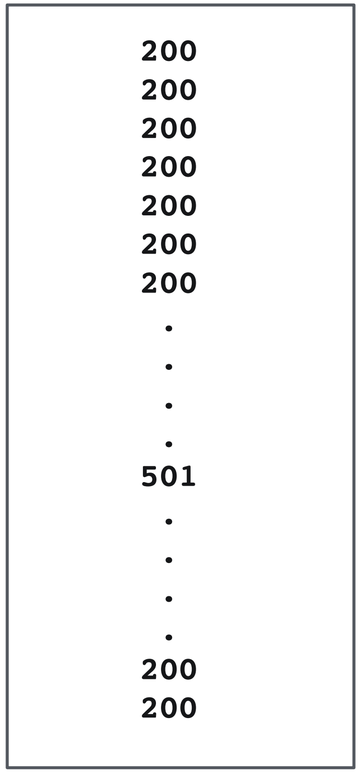
Not just the mean changes...



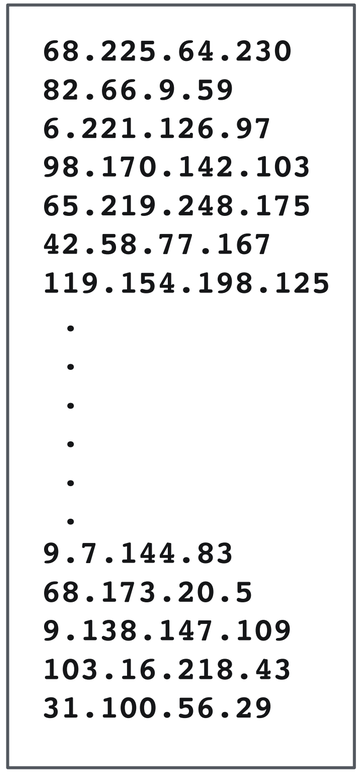
What is an Anomaly?

One way to think about what is an anomaly is “what should I draw to someone’s attention”.

Thought experiment...



status.txt



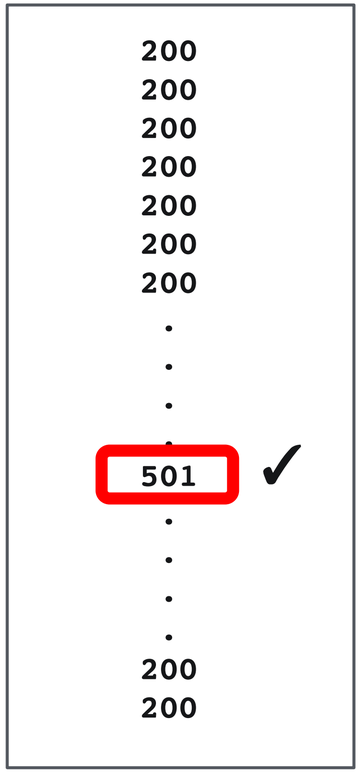
IP.txt

What is an Anomaly?

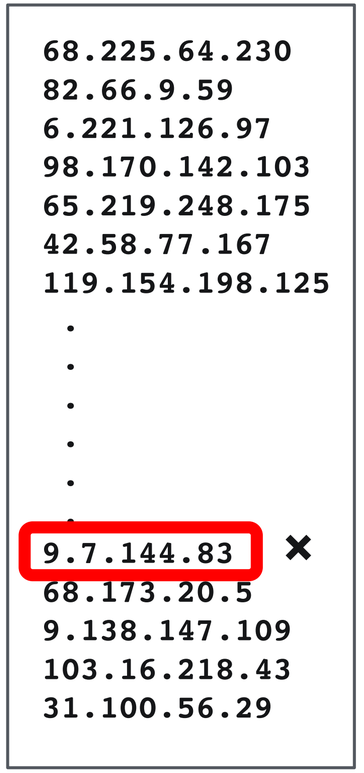
A definition for anomalousness

> “ *The chance of making a draw of a value which is equal or less likely* ”

Generalizes to any density function



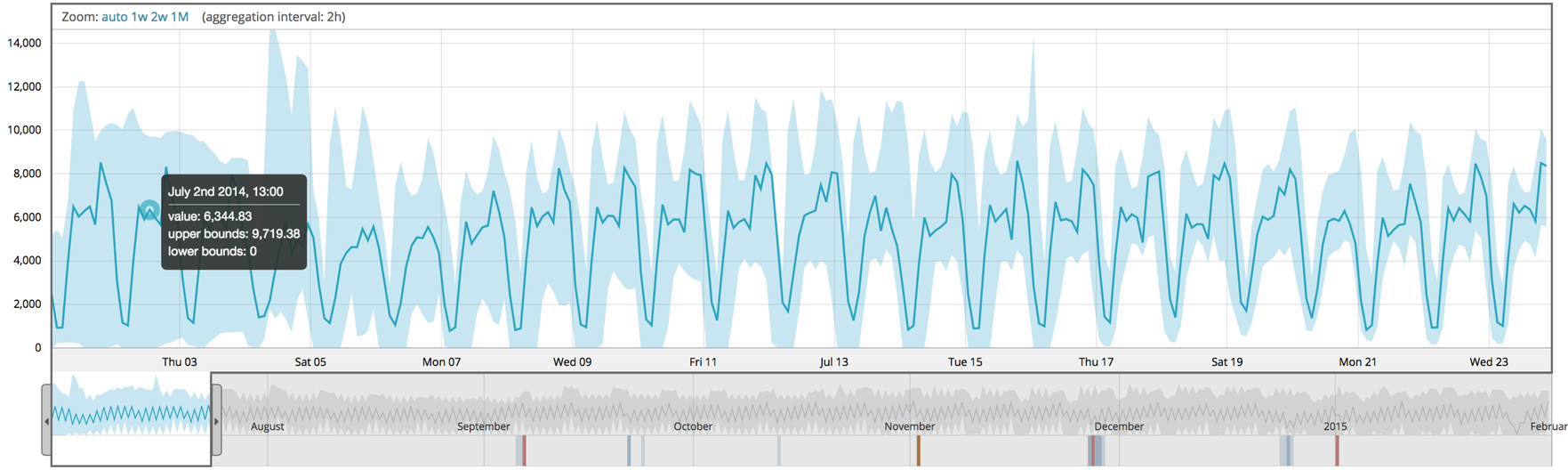
status.txt

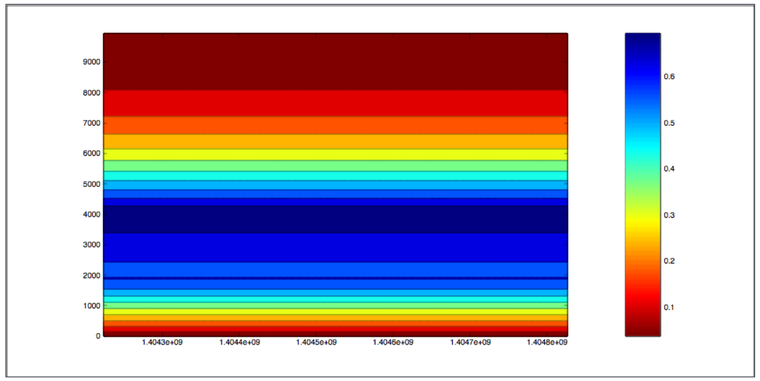


IP.txt

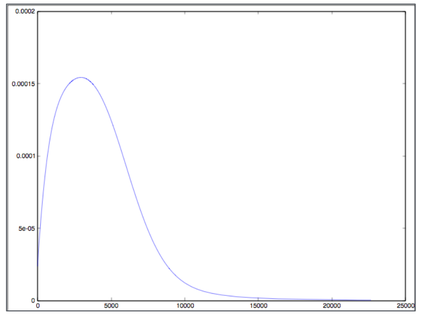
Demo

NYC Yellow Taxi Data

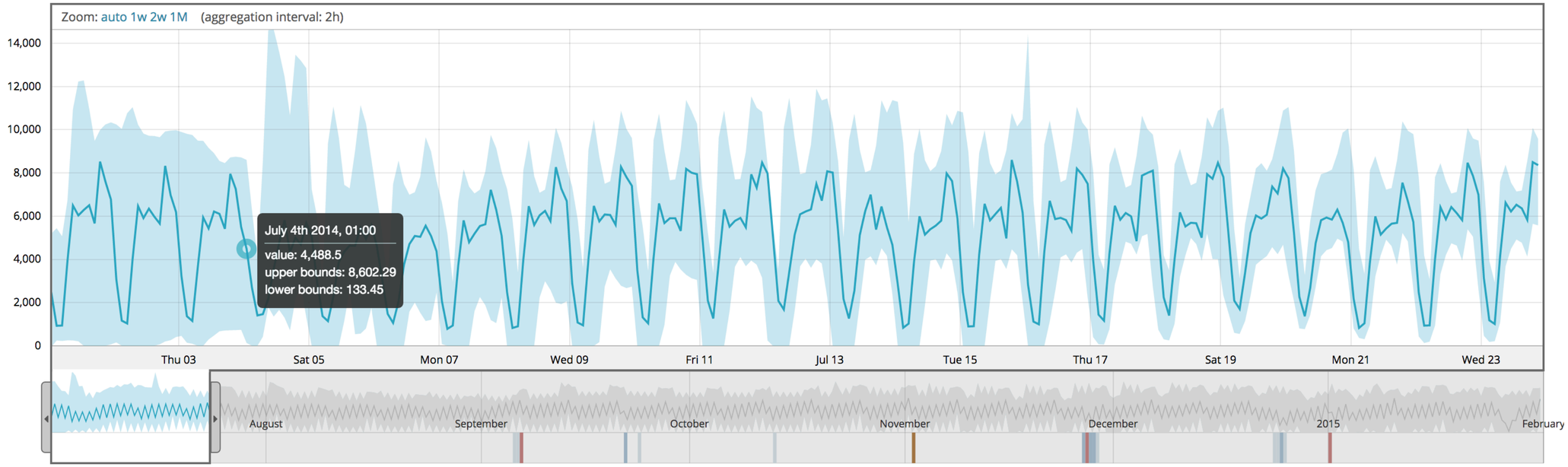


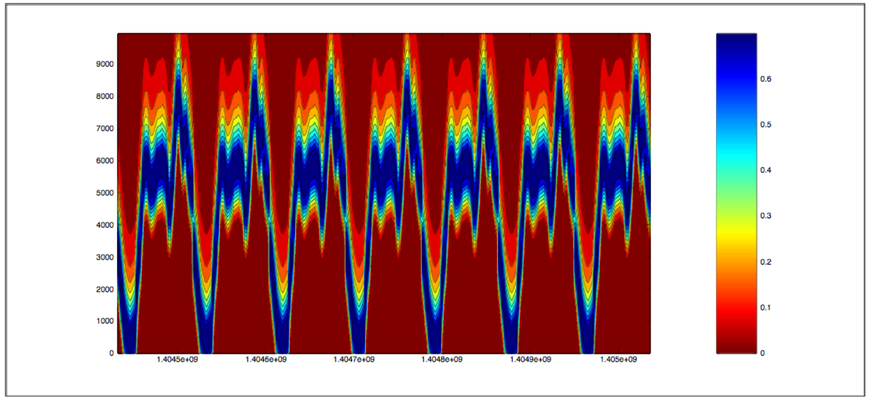


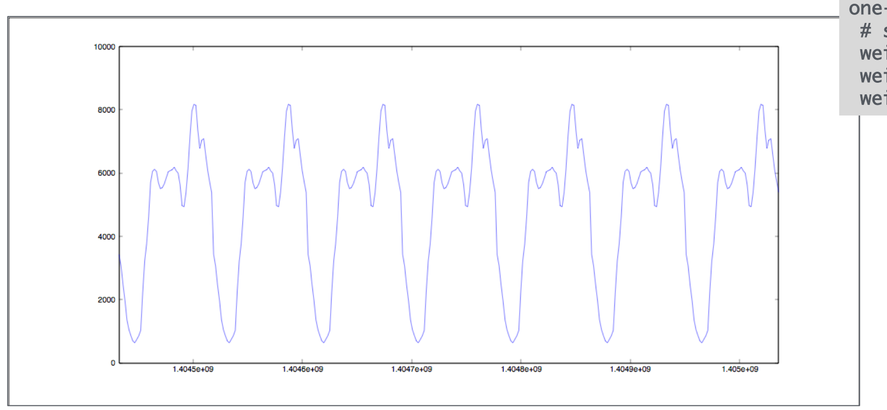

```
one-of-n:
 # samples 27.08807
 weight 0.6127408  gamma mean = 3997.926 sd = 3057.266
 weight 0.1349328  log-normal mean = 4312.742 sd = 4397.683
 weight 1  normal mean = 3998.194 sd = 2459.595
```



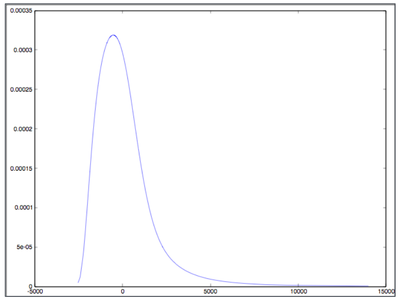
NYC Yellow Taxi Data



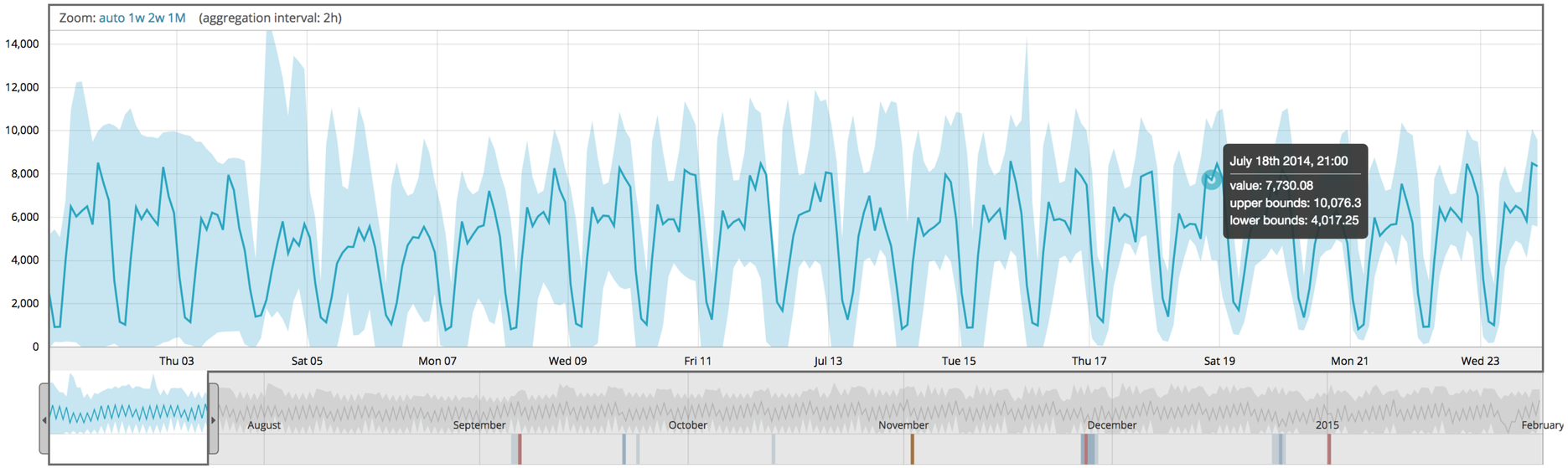


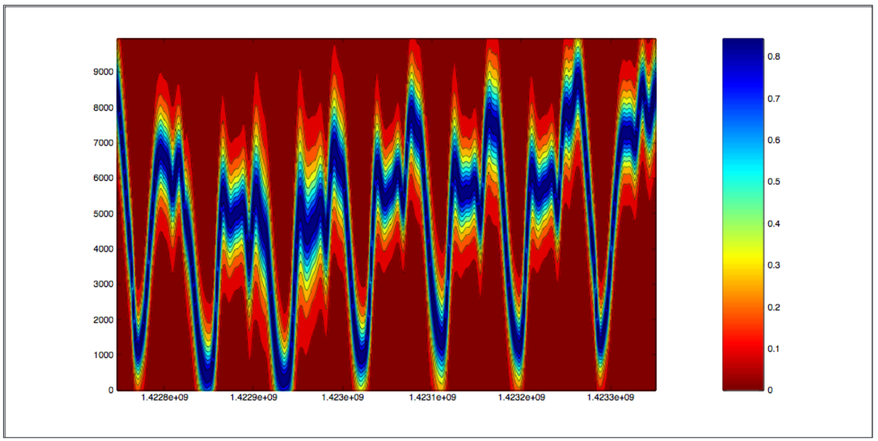


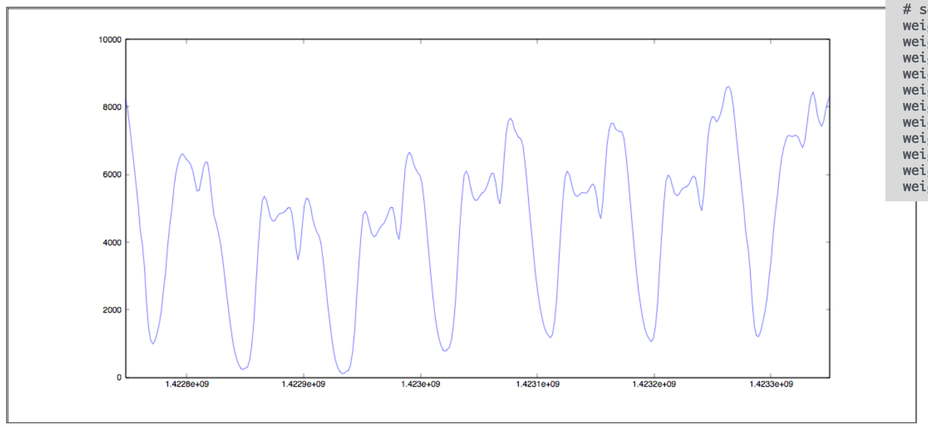

```
one-of-n:
 # samples 10.24265
 weight 0.8721099  gamma mean = -11.0318 sd = 1326.766
 weight 1  log-normal mean = 306.7829 sd = 2202.448
 weight 0.7571213  normal mean = -11.5318 sd = 983.4198
```



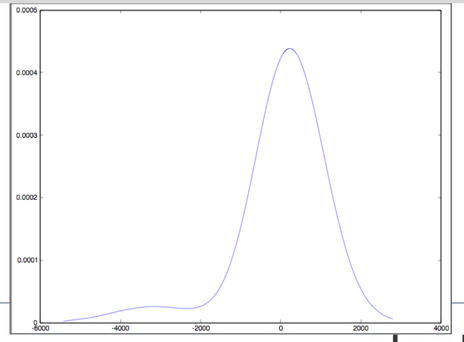
NYC Yellow Taxi Data








```
one-of-n:
 # samples 2419.166
 weight 1  multimodal:
 weight 1   weight 0.06826681  one-of-n:
 weight 1   weight 0.06826681   # samples 165.1488
 weight 1   weight 0.06826681   weight 0.2297682  gamma mean = -3130.213 sd = 1068.919
 weight 1   weight 0.06826681   weight 0.08569148  log-normal mean = -3128.735 sd = 1083.164
 weight 1   weight 0.06826681   weight 1  normal mean = -3130.713 sd = 1054.263
 weight 1   weight 0.9317332  one-of-n:
 weight 1   weight 0.9317332   # samples 2254.017
 weight 1   weight 0.9317332   weight 0.382987  gamma mean = 251.7118 sd = 849.5769
 weight 1   weight 0.9317332   weight 0.2192974  log-normal mean = 251.858 sd = 852.8664
 weight 1   weight 0.9317332   weight 1  normal mean = 251.2118 sd = 846.9203
```



More Questions?

Visit us at the AMA

www.elastic.co

Please attribute Elastic with a link to elastic.co