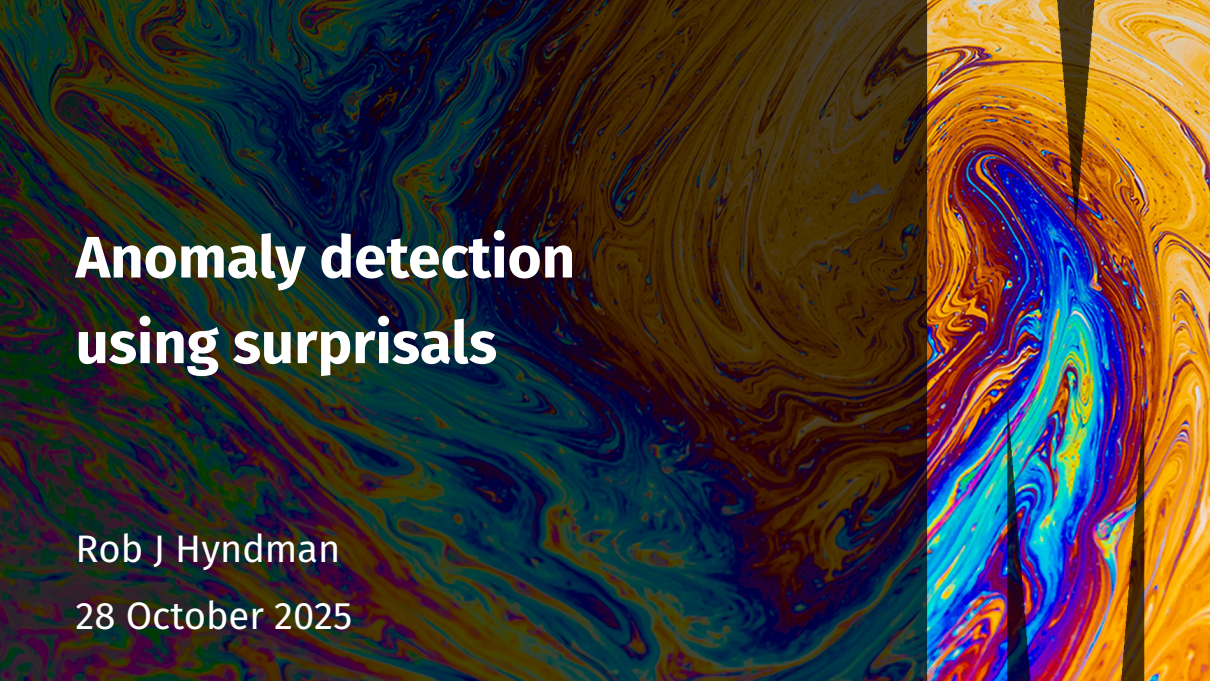


Anomaly detection using surprisals

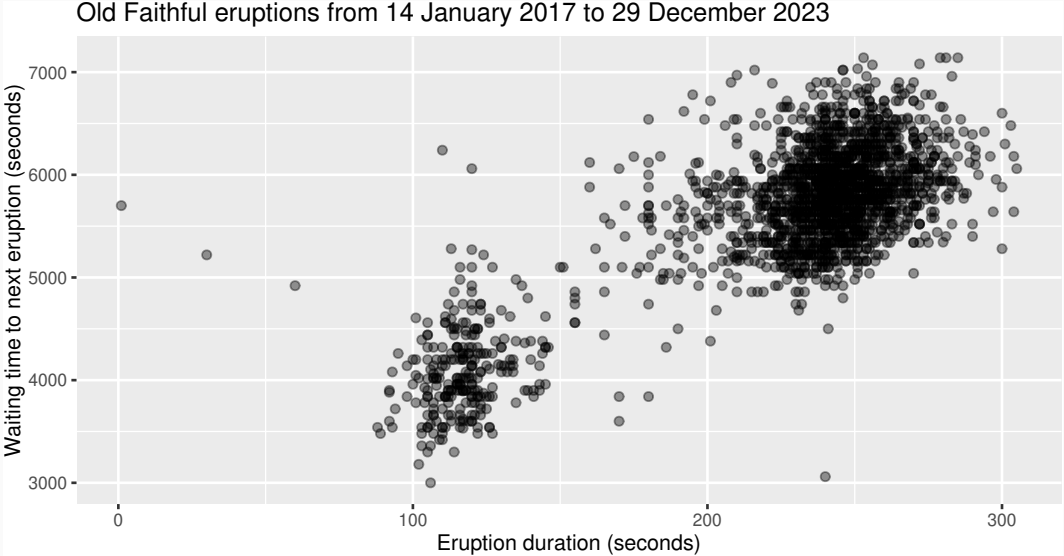
Rob J Hyndman

28 October 2025

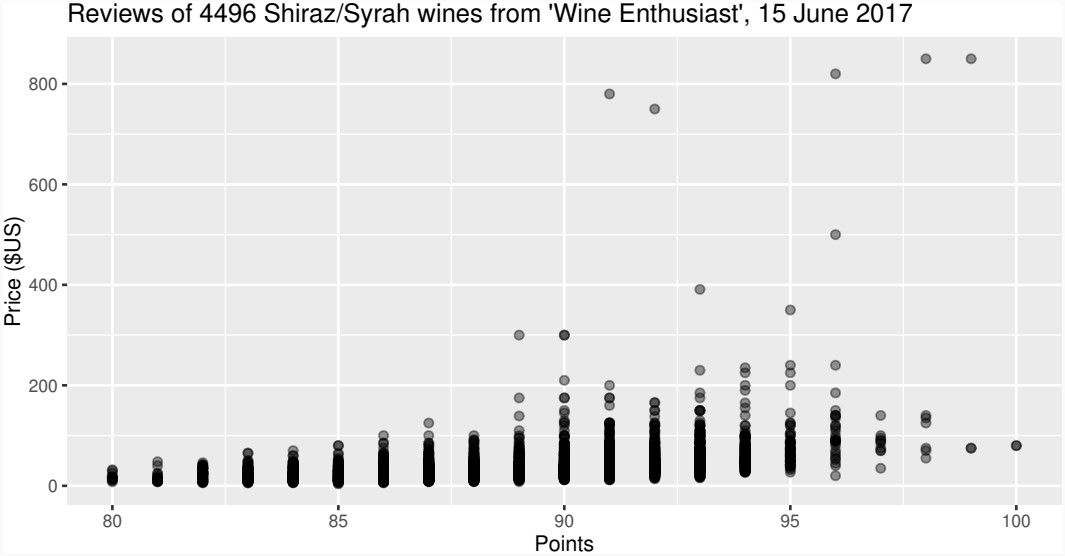
Outline

Outline

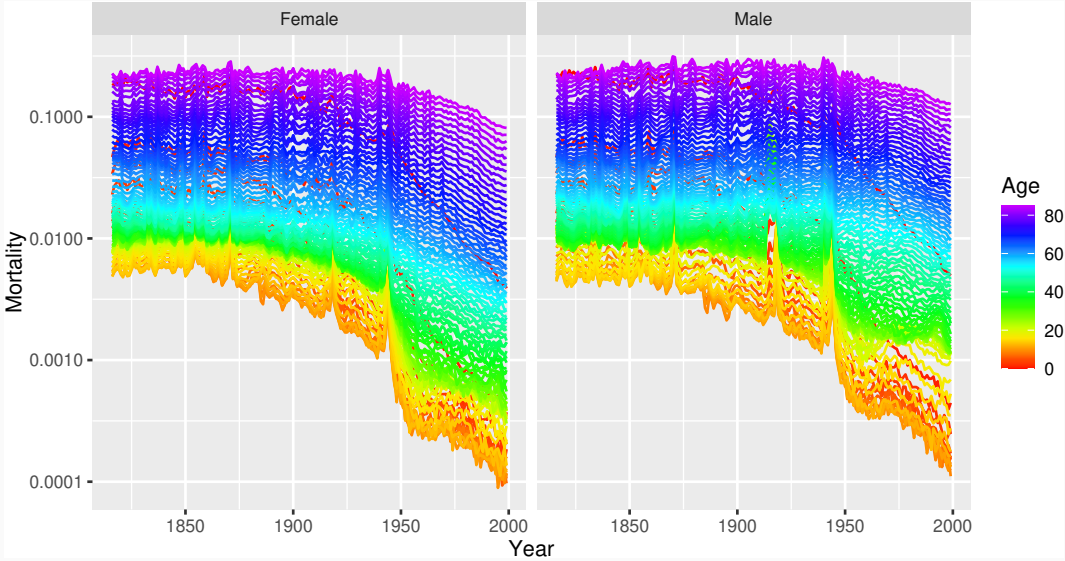
Old faithful eruptions



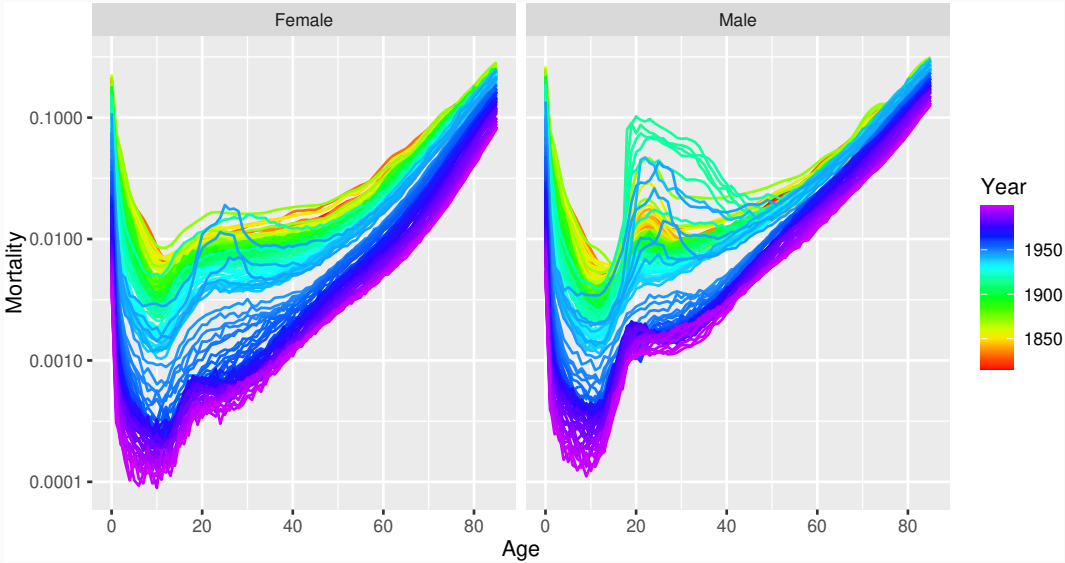
Wine quality and prices



French mortality



French mortality



Definitions of anomalies

an observation (or a subset of observations) which appears to be inconsistent with the remainder of that set of data.

(Barnett & Lewis, 1978)

Definitions of anomalies

an observation (or a subset of observations) which appears to be inconsistent with the remainder of that set of data.

(Barnett & Lewis, 1978)

an observation which deviates so much from other observations as to arouse suspicion it was generated by a different mechanism. (Hawkins, 1980)

Definitions of anomalies

Definition: Anomaly

Given a set of observations $\{y_1, \ldots, y_n\}$ drawn from probability distribution F, y_i is an **anomaly** if

$$\Pr(f(Y) < f(y_i)) < \alpha$$

where $Y \sim F$, f is the generalized density of F, and $\alpha > 0$ is a chosen threshold.

Definitions of anomalies

Definition: Anomaly

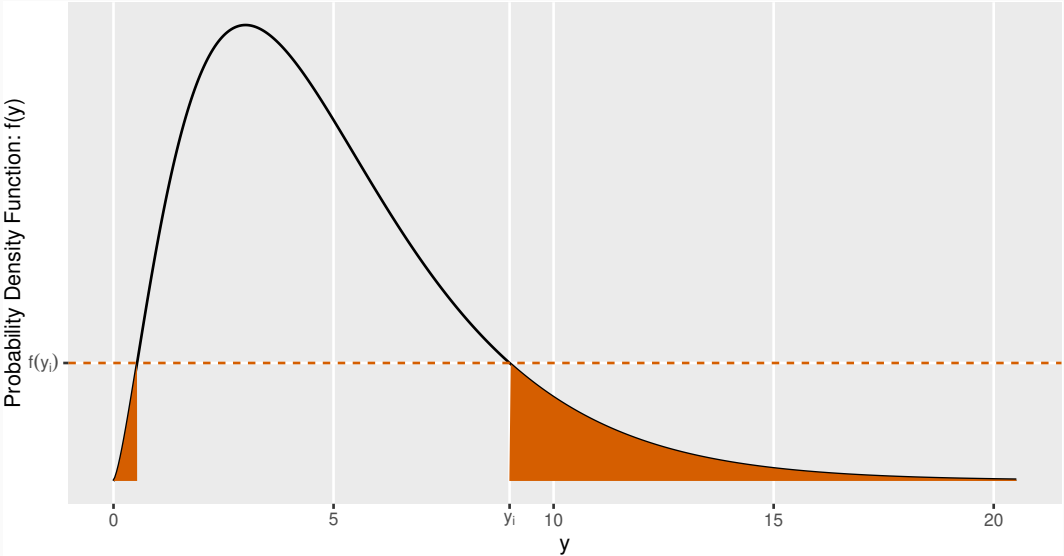
Given a set of observations $\{y_1, \ldots, y_n\}$ drawn from probability distribution F, y_i is an **anomaly** if

$$\Pr(f(Y) < f(y_i)) < \alpha$$

where $Y \sim F$, f is the generalized density of F, and $\alpha > 0$ is a chosen threshold.

- y_i can be a scalar, vector or a more complex object
- f can be a conditional density, and can be known or estimated

Definitions of anomalies



Anomaly detection: Normal distribution

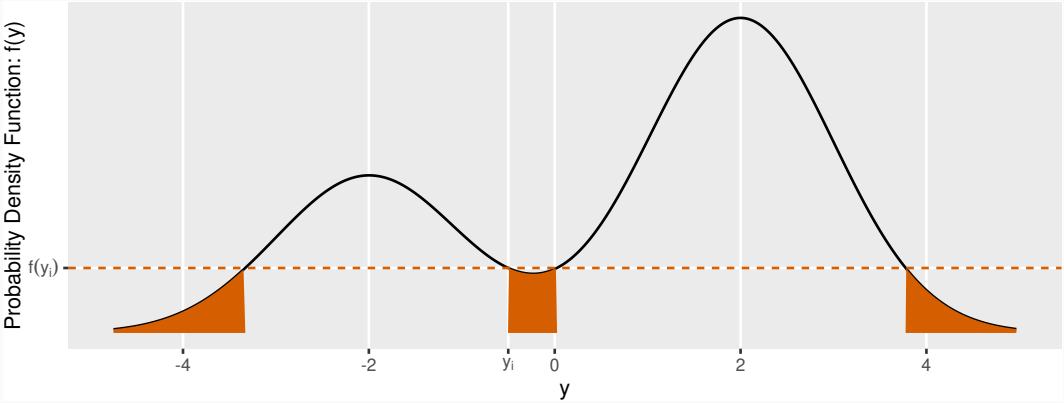
If $F \sim N(\mu, \sigma^2)$, then $p_i = 2\left[1 - \Phi(|y_i - \mu|/\sigma)\right]$

Equivalent to a two-sided p-value from a z-score test.

Anomaly detection: Highest density regions

HDR with probability $1-\alpha$ is $R_\alpha = \{y : f(y) \geq c_\alpha\}$ where c_α is largest constant s.t. $\Pr(Y \in R_\alpha) \geq 1-\alpha$.

An observation is an anomaly if $y_i \notin R_\alpha$.



Surprisals

Definition: Surprisal

The **surprisal** of an observation y_i drawn from probability distribution F with generalized density f is defined as

$$s_i = -\log f(y_i)$$

- Better known as “log scores” in statistics.
- “Surprisal” coined by Tribus (1961).
- Average surprisal = entropy of random variable
- Sum of surprisals = negative log likelihood

Anomaly detection using surprisals

Let $G(s) = P(S \leq s)$ be the **surprisal distribution** where $S = -\log f(Y)$ and $Y \sim F$.

$$G(s) = P(-\log f(Y) \leq s) = P(f(Y) \geq e^{-s})$$

The **surprisal score** is

$$p_i = 1 - G(s_i)$$

and an observation is an **anomaly** if $p_i < \alpha$.

Outline

Fisher-Tippett-Gnedenko theorem

Consider n iid rvs $S_1, \ldots, S_n$ with cdf G and $M_n = \max\{S_1, \ldots, S_n\}$. If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$P\left\{(M_n - b_n)/a_n \leq z\right\} \to H(z) \quad \text{as} \quad n \to \infty,$$

for a non-degenerate cdf H, then

$$H(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

- $\xi > 0$: Fréchet distribution (G heavy-tailed)
- $\xi \to 0$: Gumbel distribution (G light-tailed)
- $\xi < 0$: Weibull distribution (G bounded upper tail)

Pickands-Balkema-De Haan theorem

If G satisfies the FTG theorem, then the upper tail of G can be approximated by the Generalized Pareto Distribution (GPD):

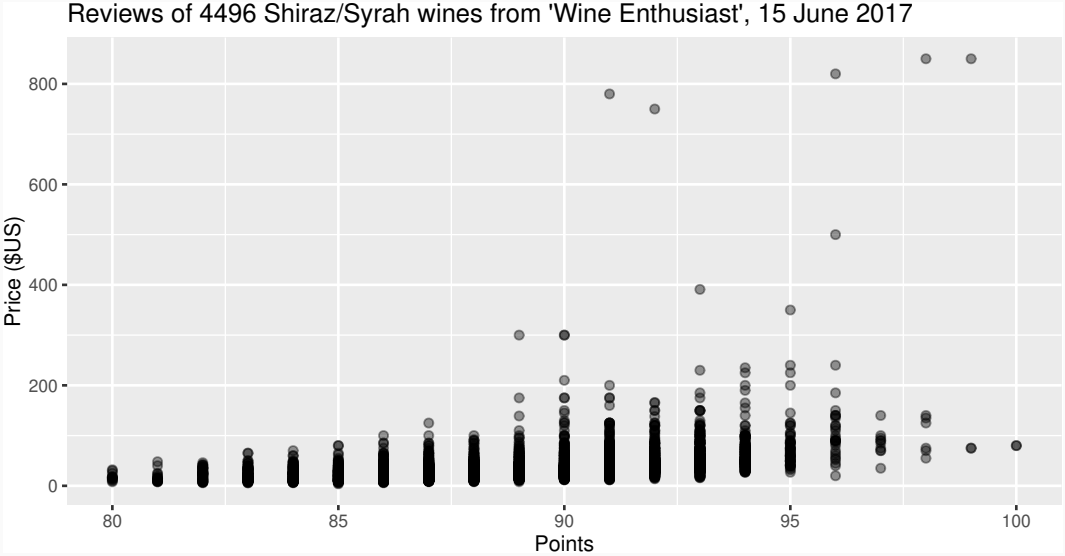
$$K(x) = \Pr(S \leq u + s \mid S > u) = 1 - \left(1 + \frac{\xi s}{\sigma_u}\right)^{-1/\xi}$$

for large enough u, where $\sigma_u = \sigma + \xi(u - \mu)$.

Surprisals and EVT

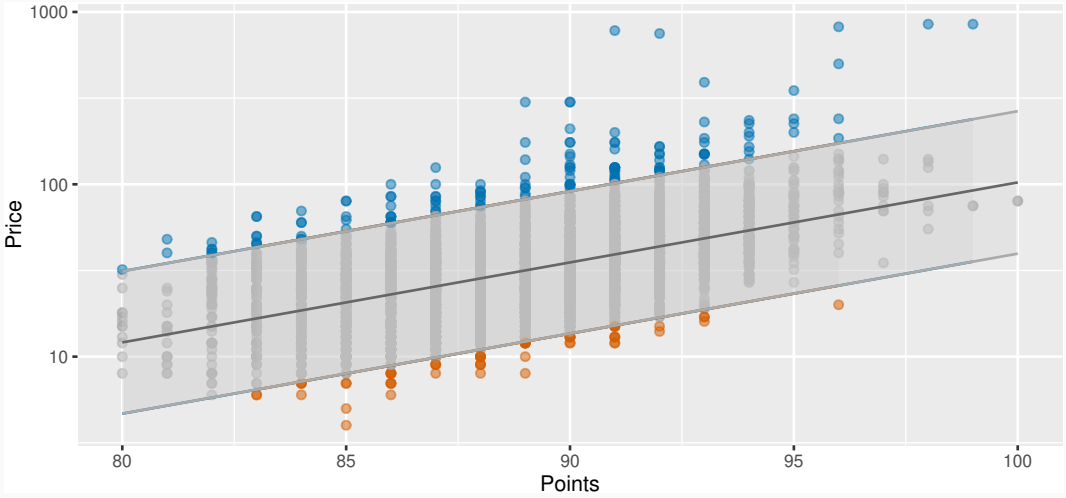
- Suppose we have n iid observations $Y_1, \ldots, Y_n$ from distribution F with density f.
- Let $S_i = -\log f(Y_i)$ be the surprisal of Y_i.
- Then $S_1, \ldots, S_n$ are iid from the surprisal distribution $G(s) = P(S \leq s)$.
- If G satisfies the FTG theorem, then we can approximate the upper tail of G by a GPD.

Application to wine quality and prices



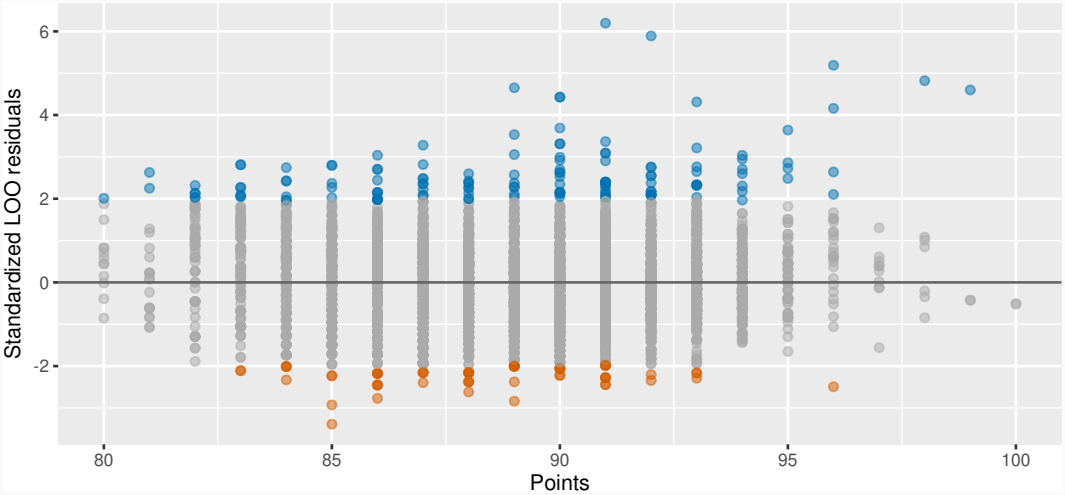
Application to wine quality and prices

Proposed model: $\log \text{Price} \mid \text{Points} \sim N(a + b\text{Points}, \sigma^2)$.



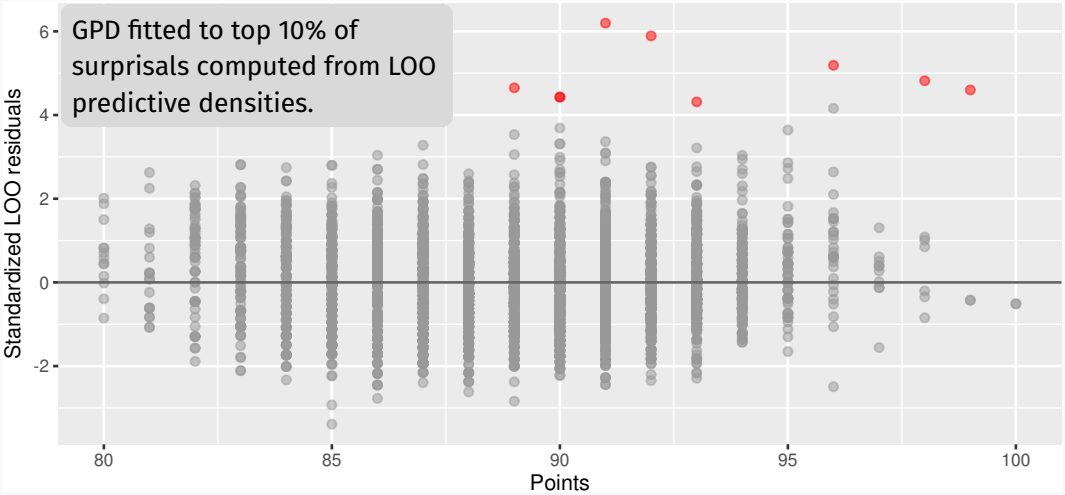
Application to wine quality and prices

Proposed model: $\log \text{Price} \mid \text{Points} \sim N(a + b\text{Points}, \sigma^2)$.

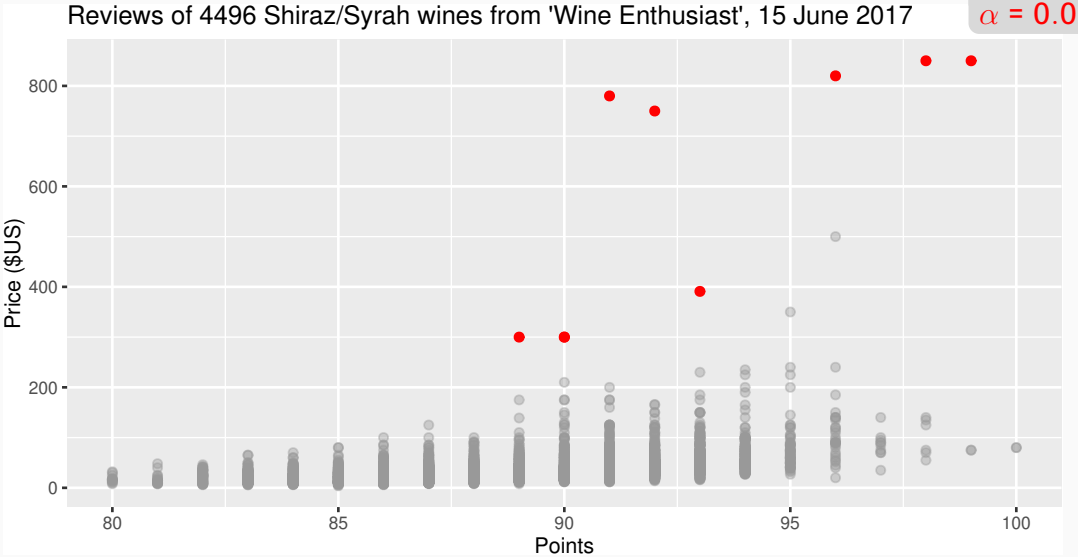


Application to wine quality and prices

Proposed model: $\log \text{Price} \mid \text{Points} \sim N(a + b\text{Points}, \sigma^2)$. $\alpha = 0.001$



Application to wine quality and prices



Application to French mortality

Outline

KDE

Bandwidth selection

Persistent homology

Application to Wine quality and prices

Outline