

# Anomaly detection using surprisals

Rob J Hyndman

28 October 2025

# Outline

- 1 Anomalies and surprisals
- 2 Extreme value theory and surprisals
- 3 Lookout algorithm
- 4 Conclusions

# Outline

1 Anomalies and surprisals

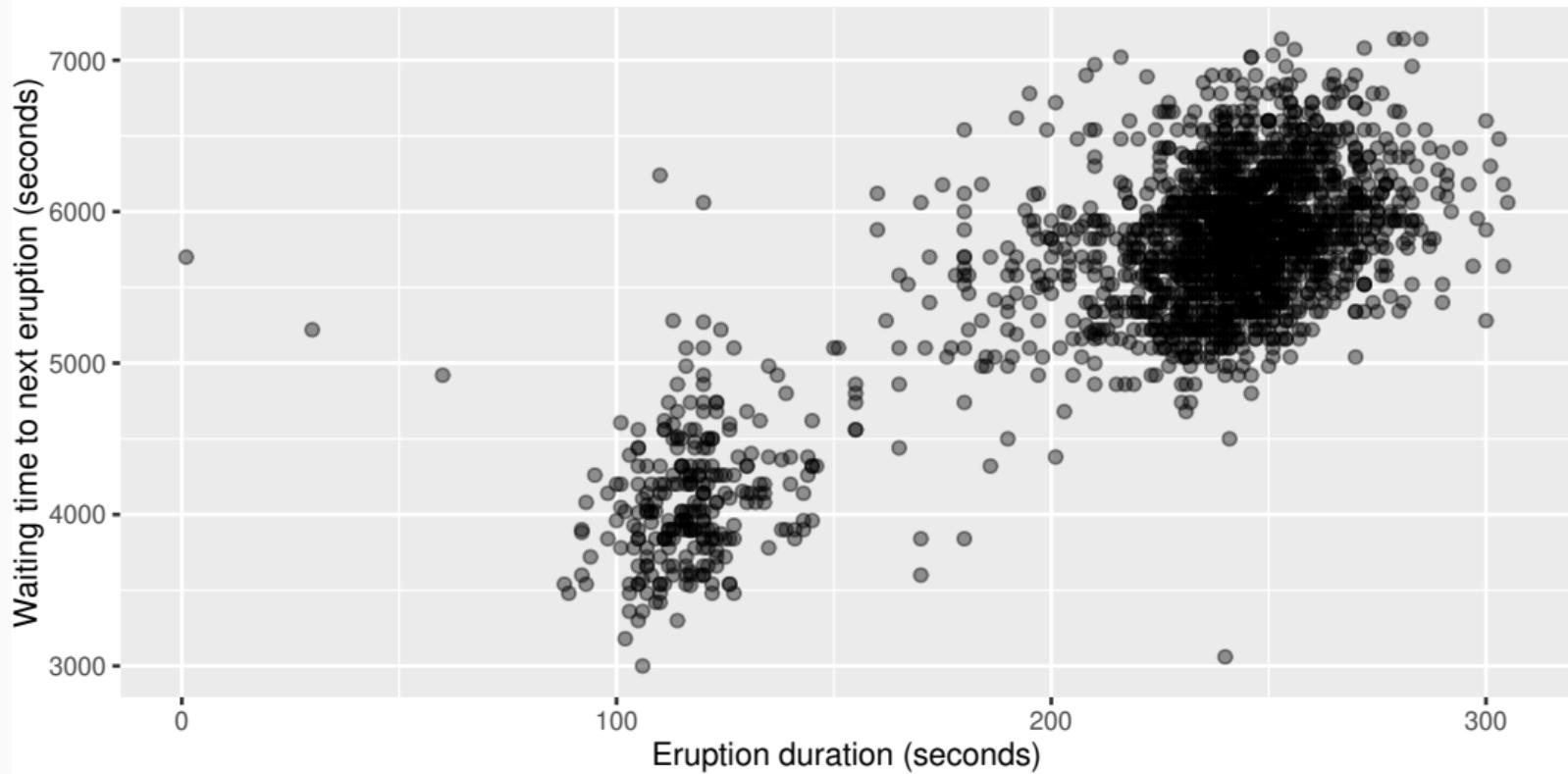
2 Extreme value theory and surprisals

3 Lookout algorithm

4 Conclusions

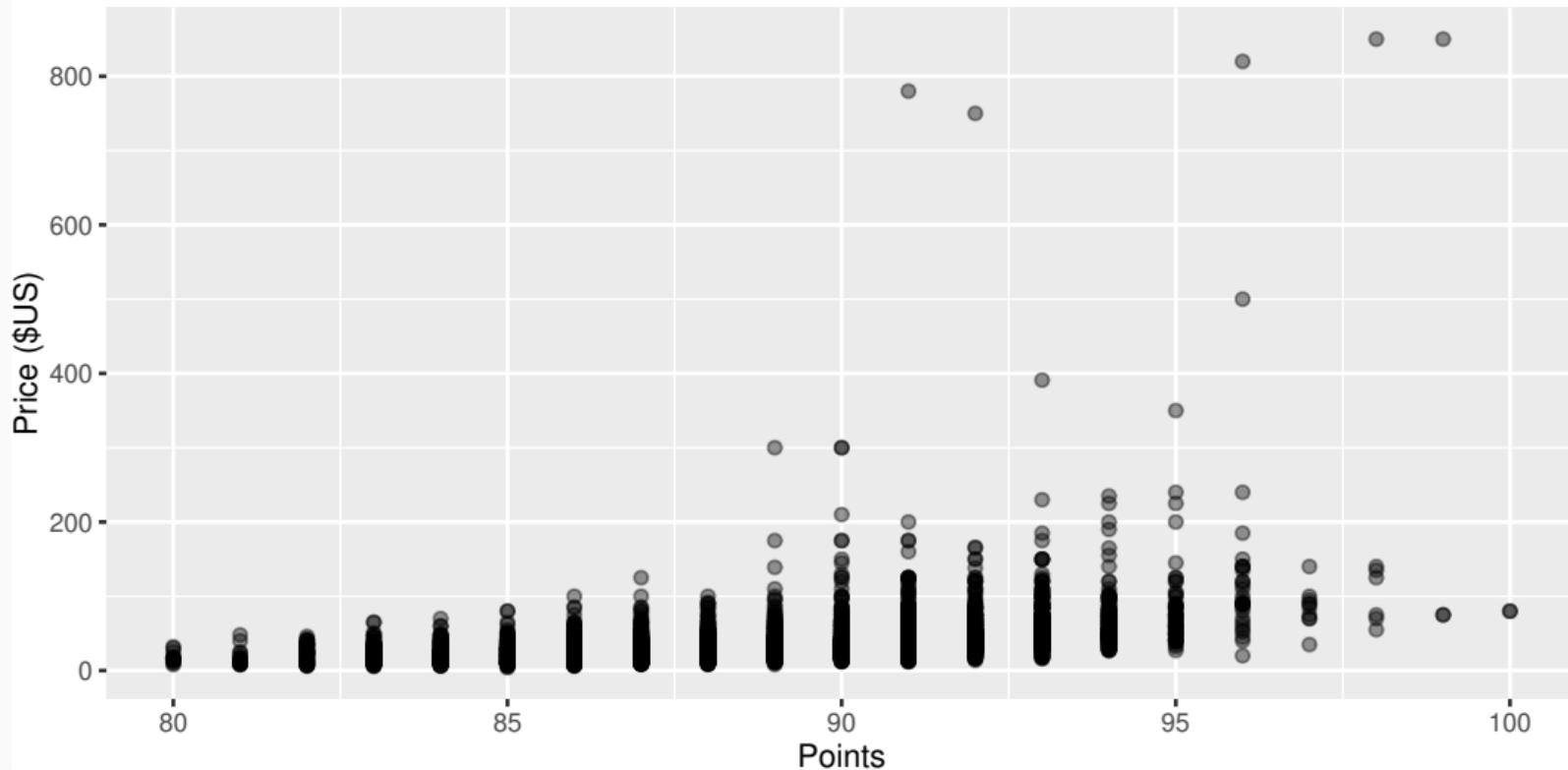
# Old faithful eruptions

Old Faithful eruptions from 14 January 2017 to 29 December 2023

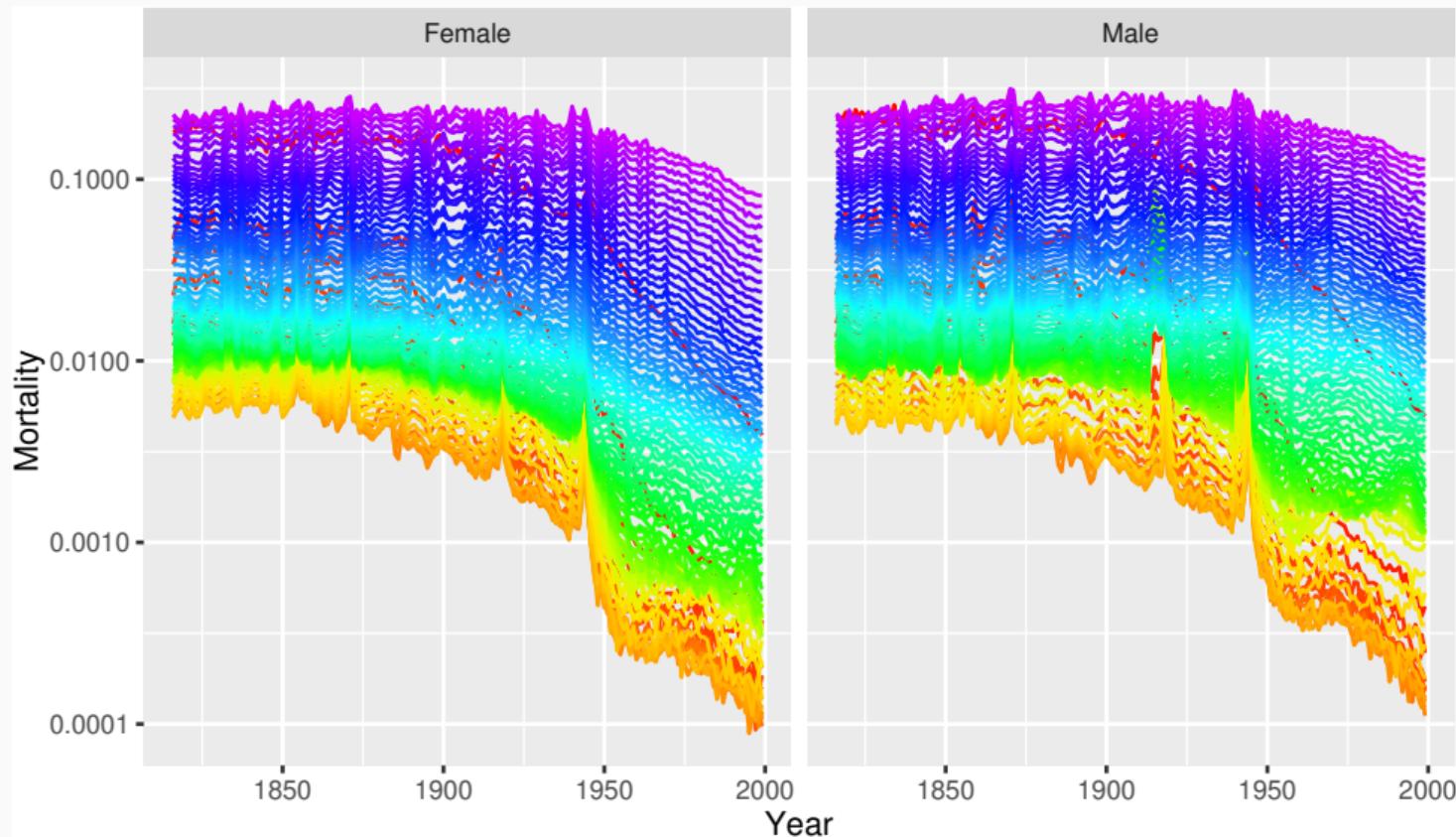


# Wine quality and prices

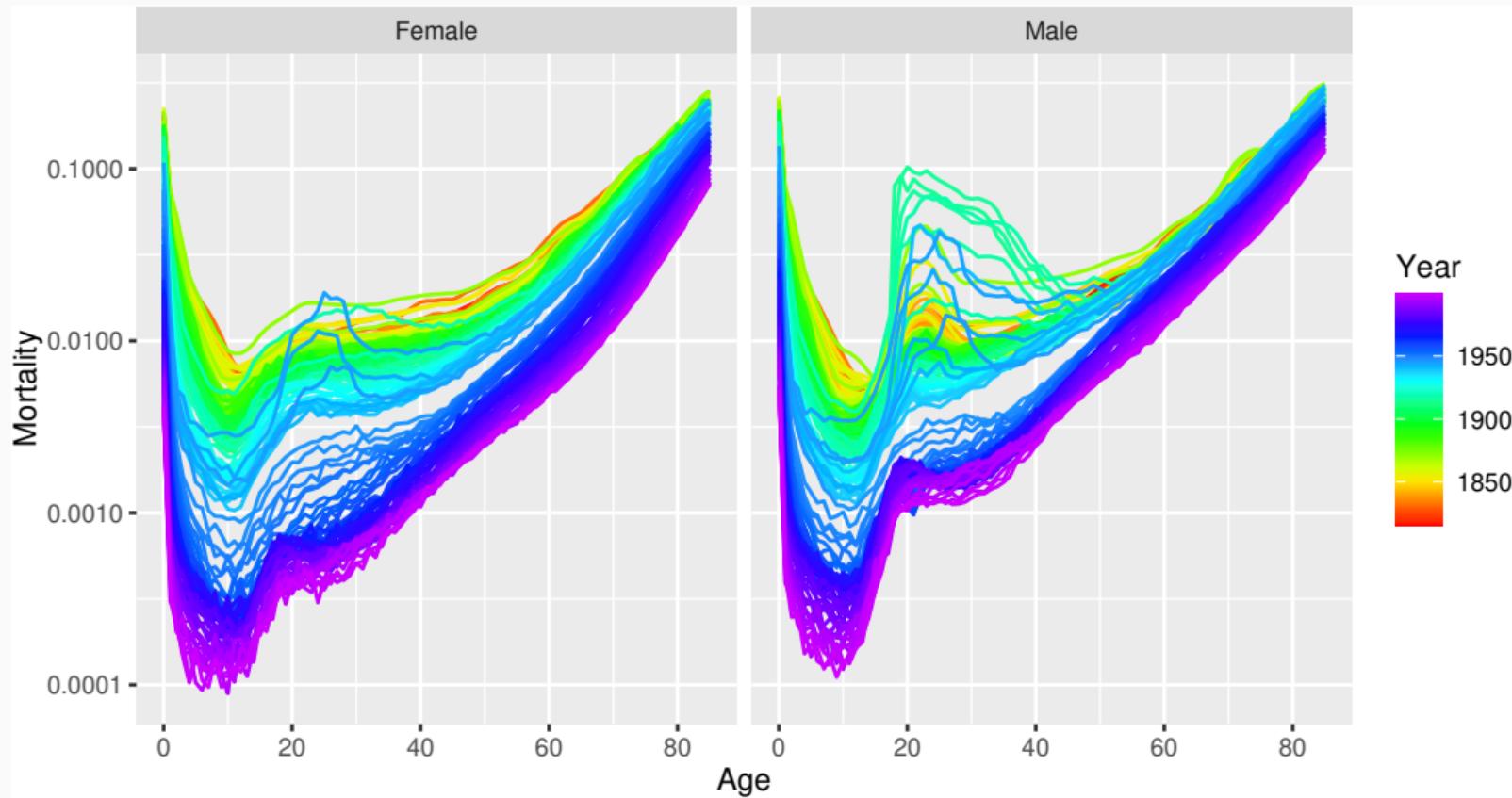
Reviews of 4496 Shiraz/Syrah wines from 'Wine Enthusiast', 15 June 2017



# 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  $\{y_1, \dots, y_n\}$  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

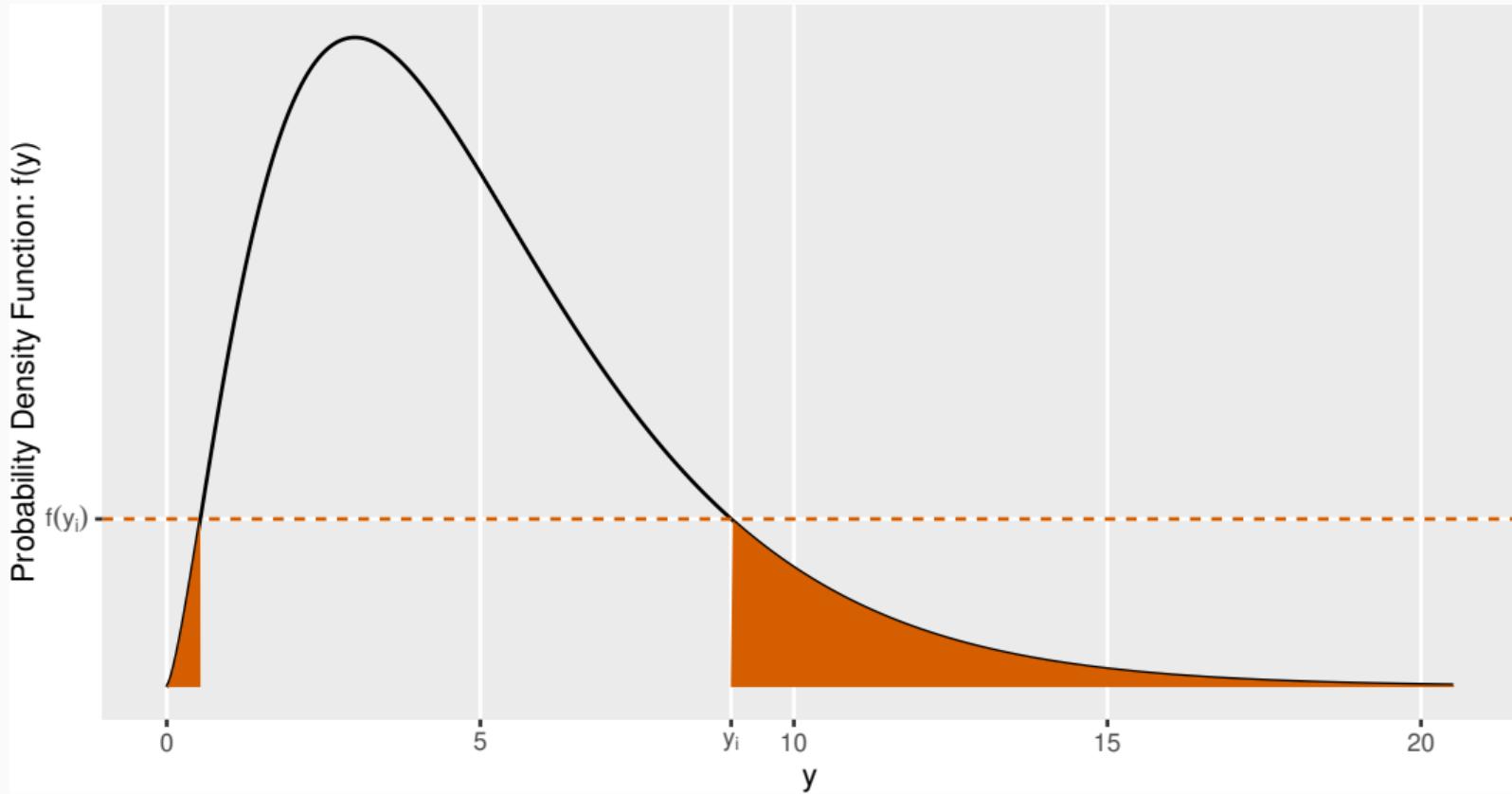
Given a set of observations  $\{y_1, \dots, y_n\}$  drawn from probability distribution  $F$ ,  $y_i$  is an **anomaly** if

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

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

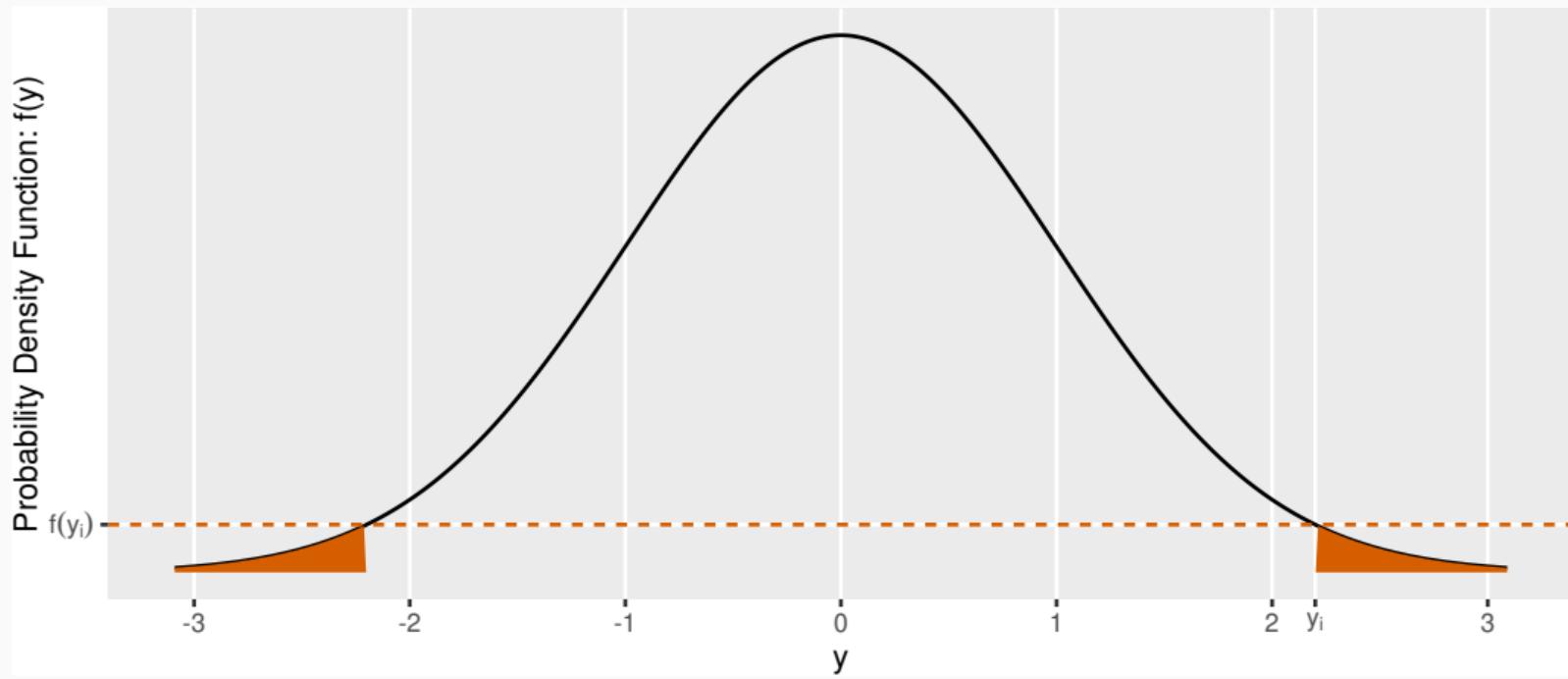
# Definitions of anomalies



# Anomaly detection: Normal distribution

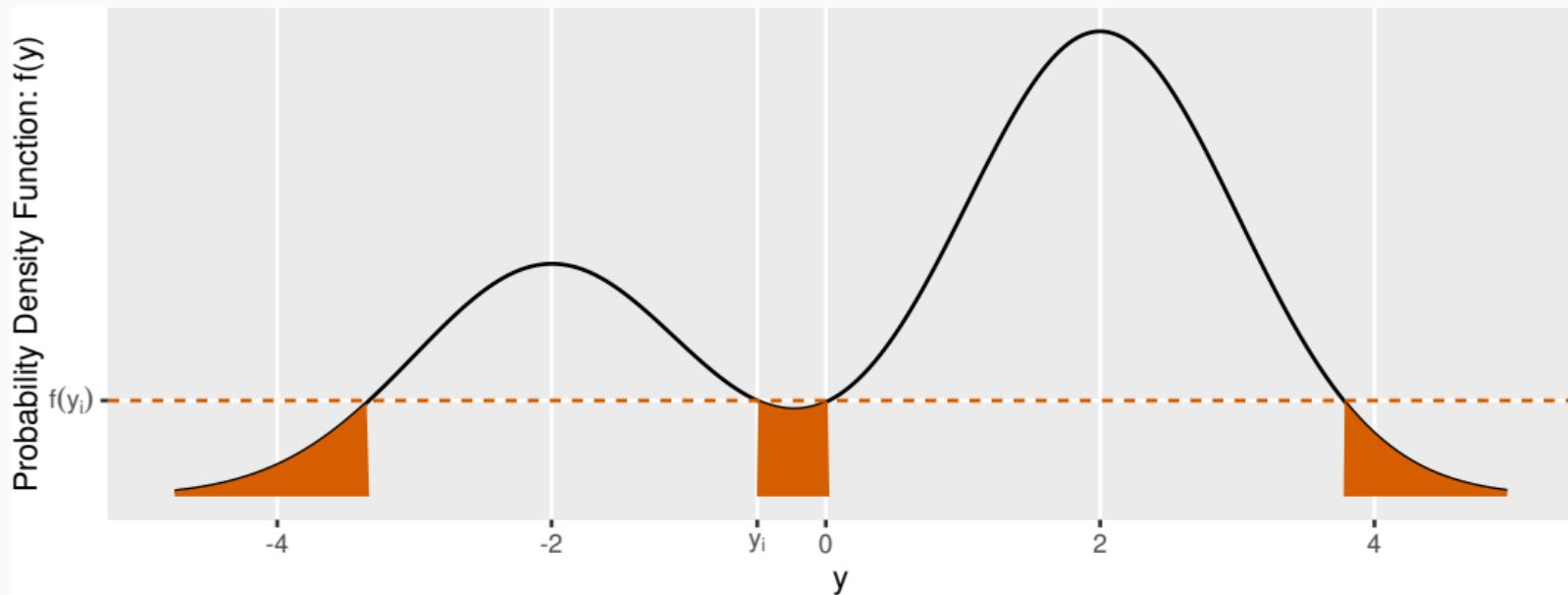
If  $F \sim N(\mu, \sigma^2)$ , then  $p_i = 2 [1 - \Phi(|y_i - \mu|/\sigma)]$

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  $c_\alpha$  is largest constant s.t.  $\Pr(Y \in R_\alpha) \geq 1 - \alpha$ .  
An observation is an anomaly if  $y_i \notin R_\alpha$ .



# Surprises

## 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

- 1 Anomalies and surprisals
- 2 Extreme value theory and surprisals
- 3 Lookout algorithm
- 4 Conclusions

# Fisher-Tippett-Gnedenko theorem

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

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

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)
- $\xi \rightarrow 0$ : 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):

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

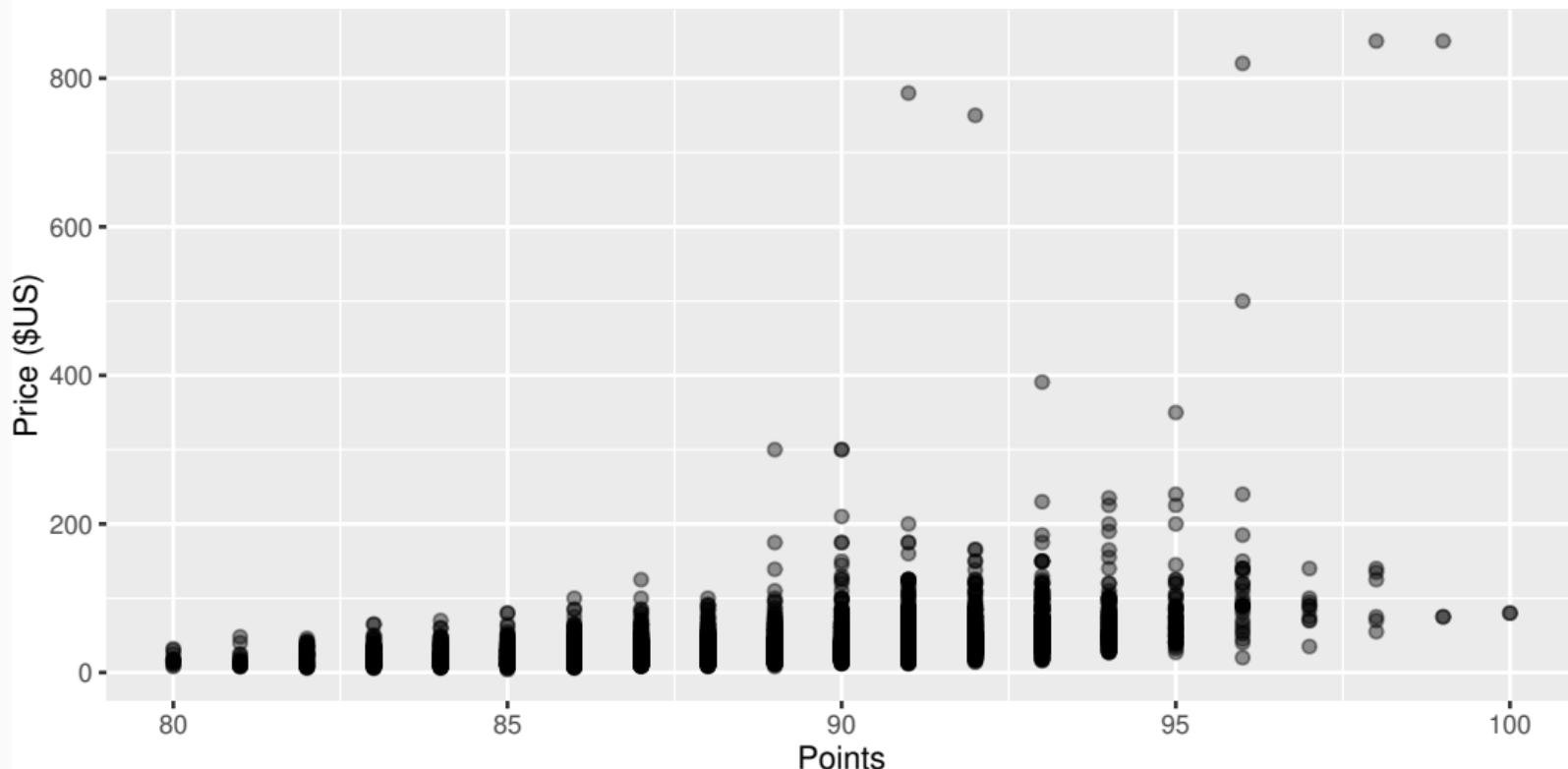
for large enough  $u$ , where  $\sigma_u = \sigma + \xi(u - \mu)$ .

# Surprises and EVT

- Suppose we have  $n$  iid observations  $Y_1, \dots, Y_n$  from distribution  $F$  with density  $f$ .
- Let  $S_i = -\log f(Y_i)$  be the surprisal of  $Y_i$ .
- Then  $S_1, \dots, S_n$  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.

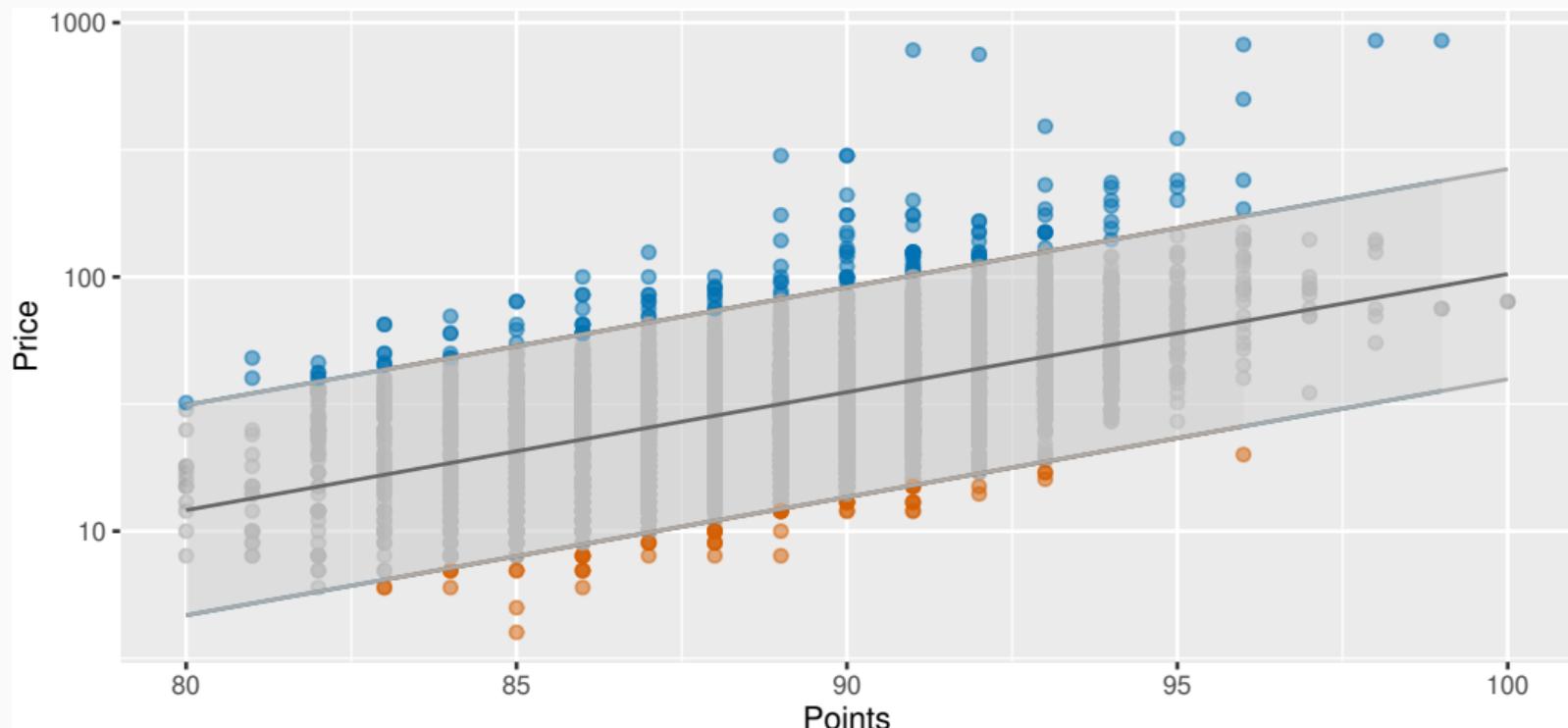
# Application to wine quality and prices

Reviews of 4496 Shiraz/Syrah wines from 'Wine Enthusiast', 15 June 2017



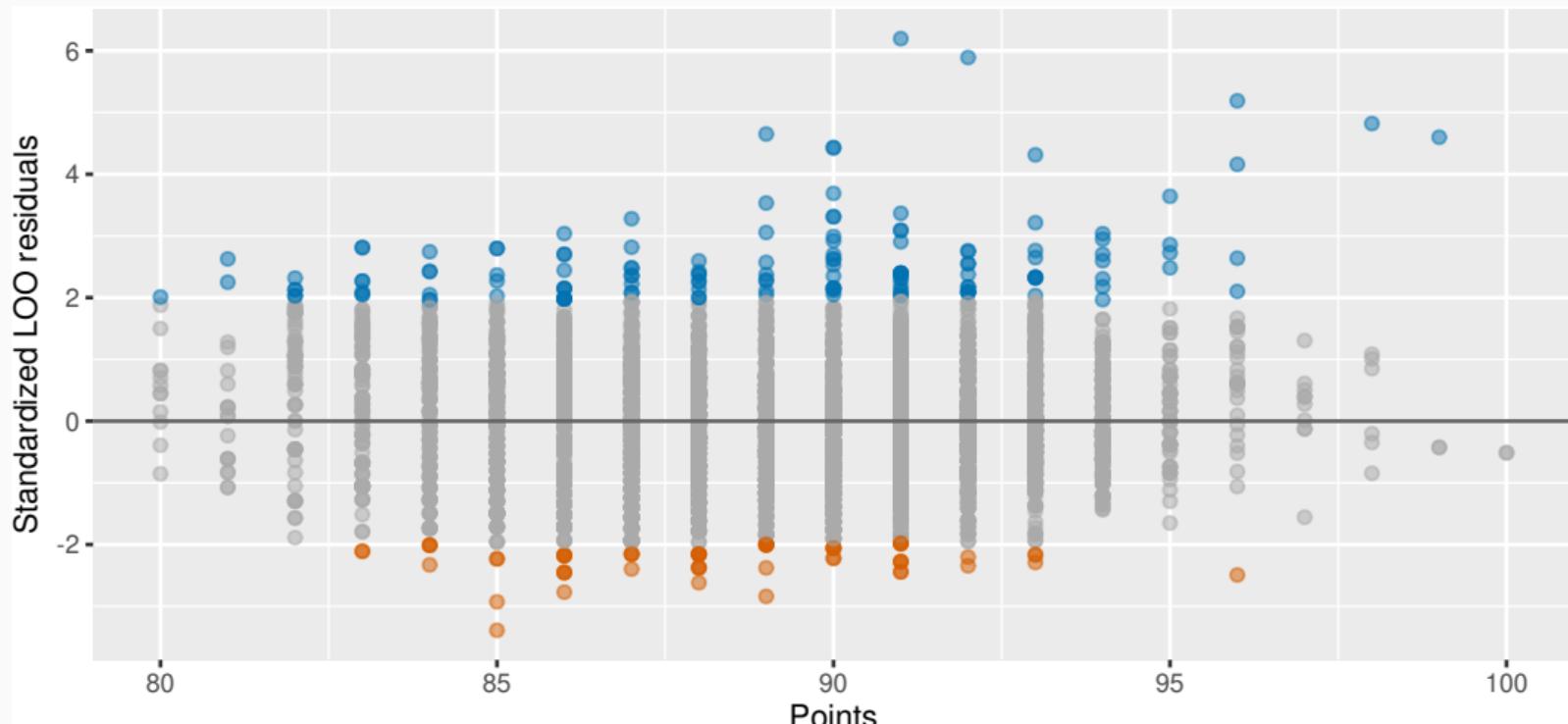
# Application to wine quality and prices

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



# Application to wine quality and prices

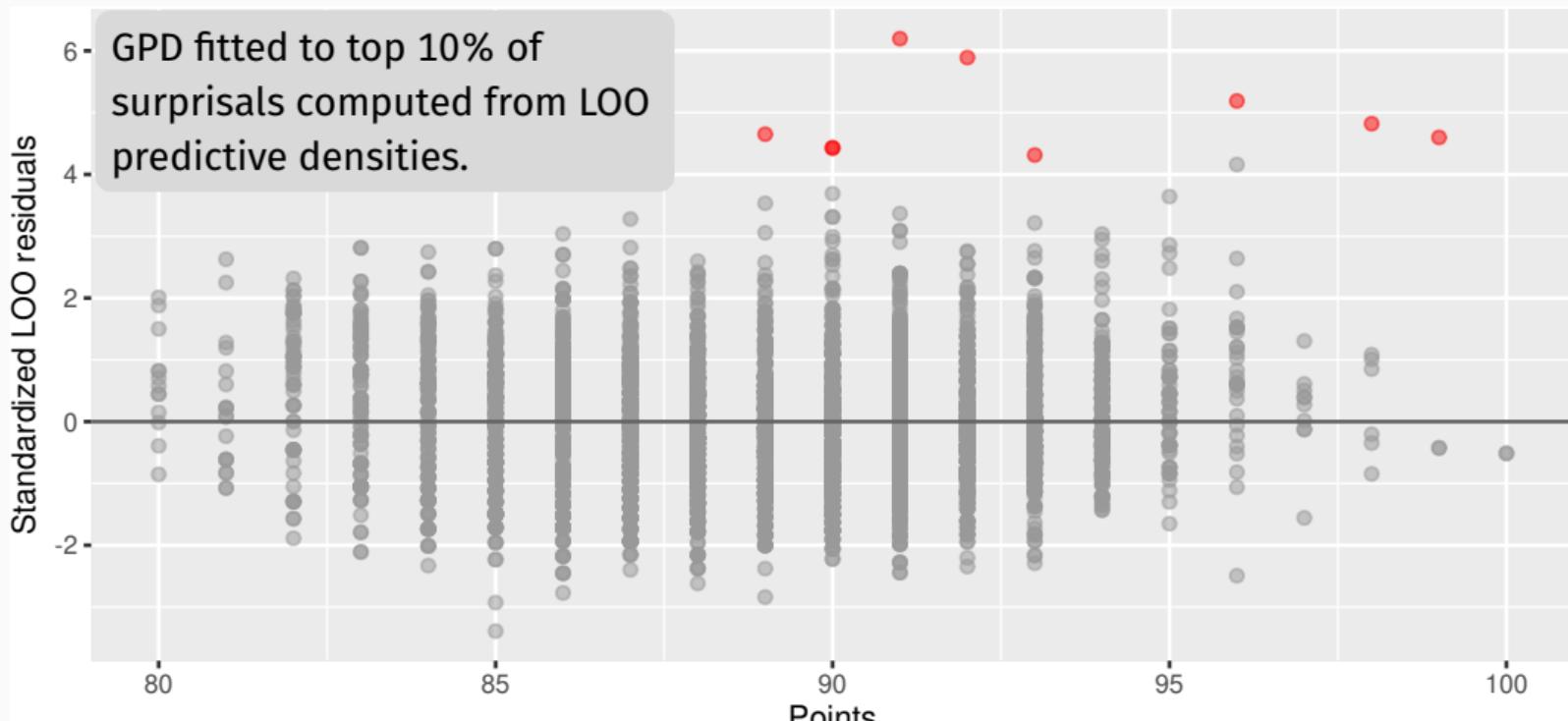
Proposed model:  $\log \text{Price} | \text{Points} \sim N(a + b\text{Points}, \sigma^2)$ .



# Application to wine quality and prices

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

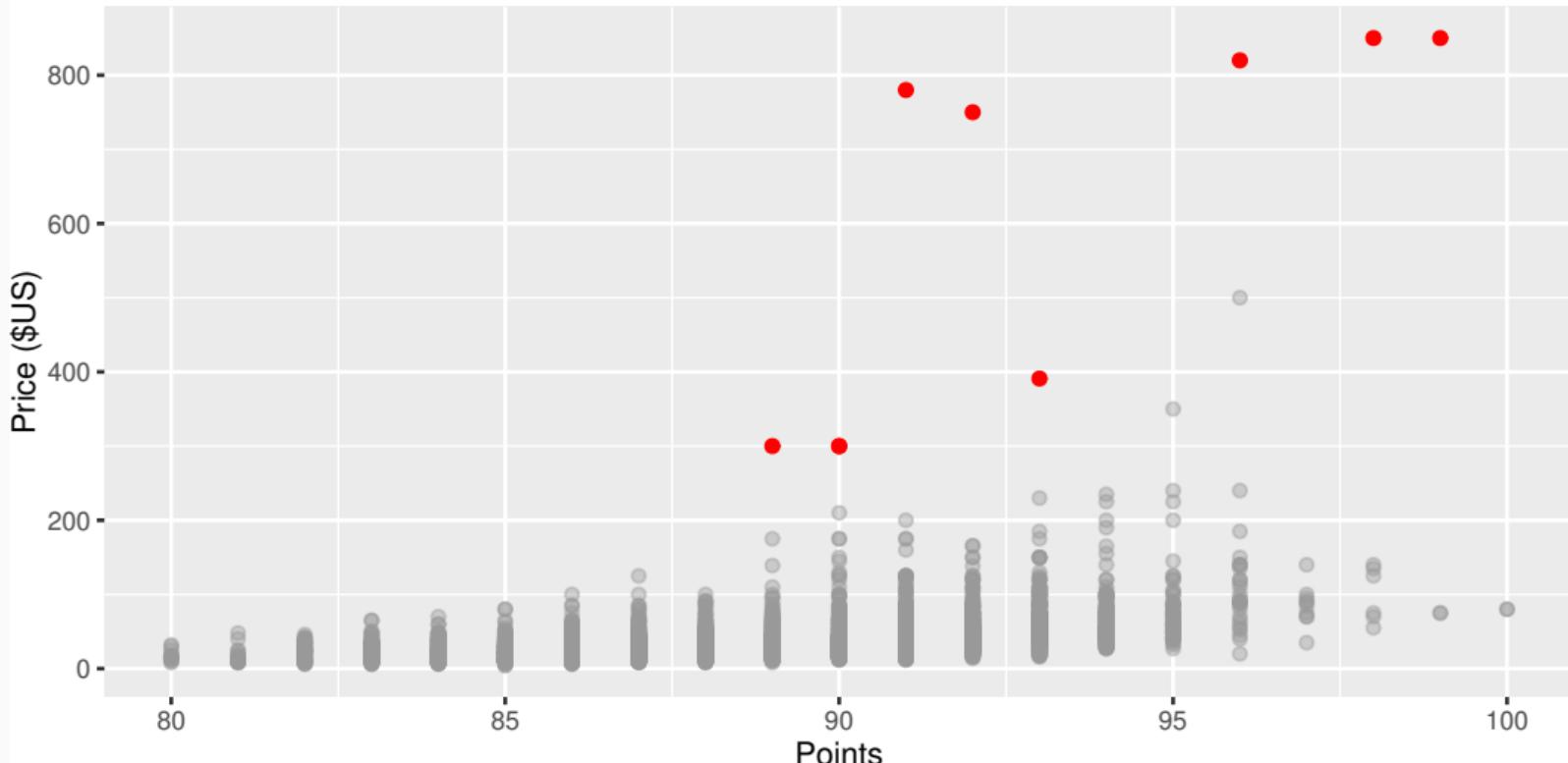
$\alpha = 0.001$



# Application to wine quality and prices

Reviews of 4496 Shiraz/Syrah wines from 'Wine Enthusiast', 15 June 2017

$\alpha = 0.001$



# Application to wine quality and prices

Anomalies detected ( $\alpha = 0.001$ ):

Area	Winery	Year	Points	Price
South Australia	Henschke	2009	91	780
California	Law	2013	92	750
South Australia	Henschke	2010	96	820
South Australia	Penfolds	2008	98	850
Tuscany	Tua Rita	2011	89	300
South Australia	Penfolds	2010	99	850
Tuscany	Tua Rita	2013	90	300
Tuscany	Tua Rita	2012	90	300
Rhône Valley	Domaine Jean-Michel Gerin	2013	93	391

# Application to French mortality

# Outline

- 1 Anomalies and surprisals
- 2 Extreme value theory and surprisals
- 3 Lookout algorithm
- 4 Conclusions



# Bandwidth selection

# Persistent homology

# Application to Wine quality and prices

# Outline

- 1 Anomalies and surprisals
- 2 Extreme value theory and surprisals
- 3 Lookout algorithm
- 4 Conclusions