Anomaly detection using surprisals

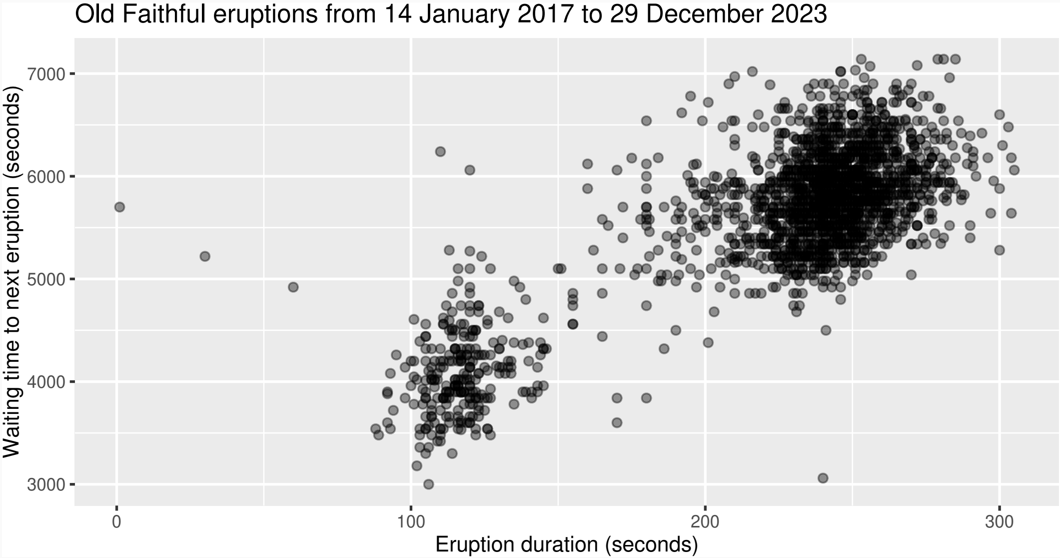
Rob J Hyndman

28 October 2025

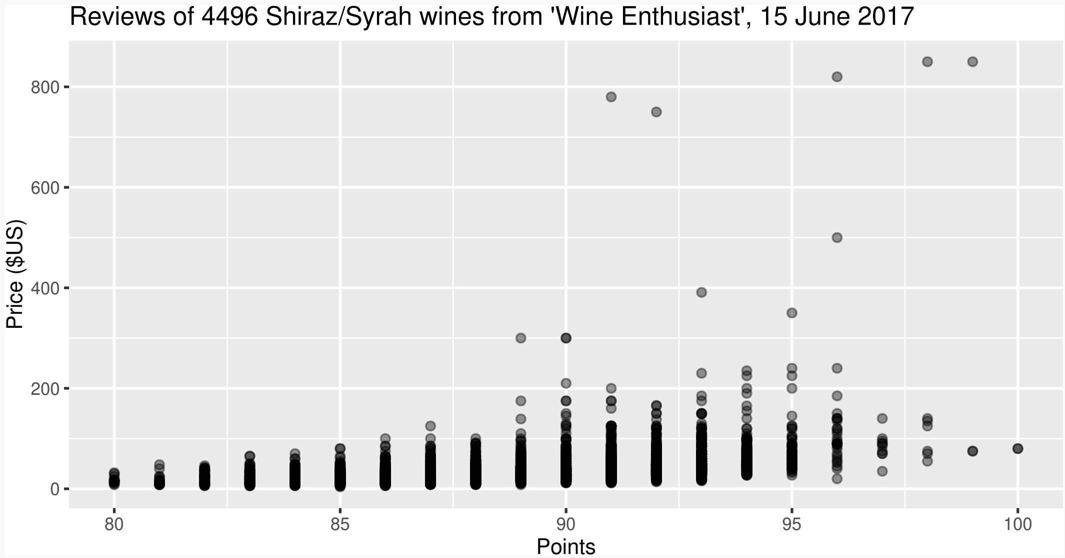
Outline

Outline

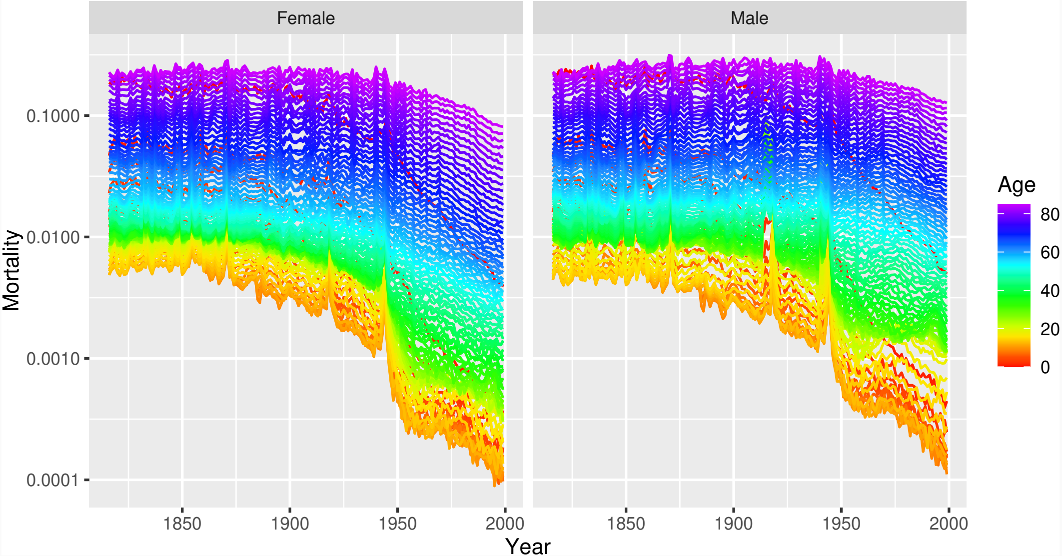
Old faithful eruptions



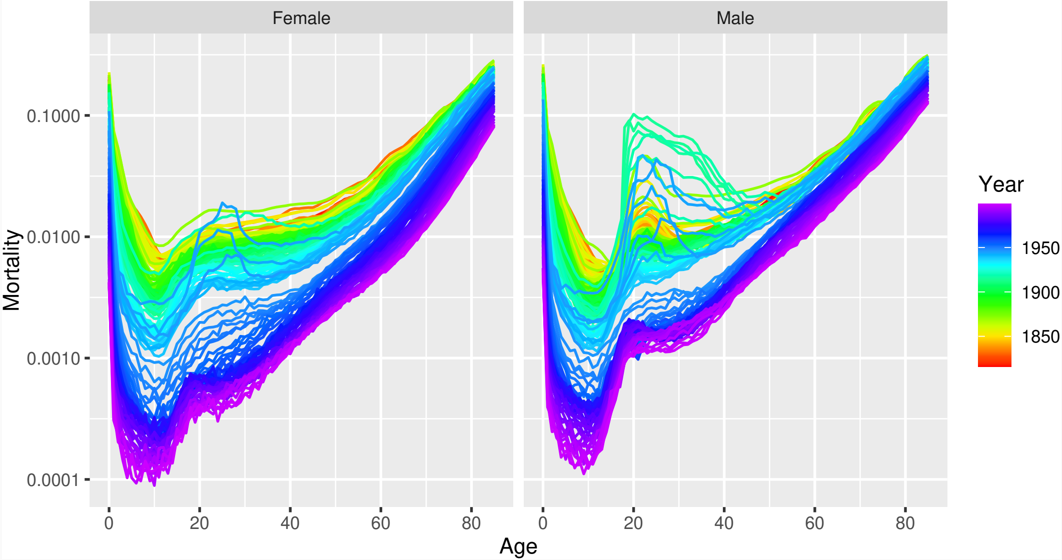
Wine quality and prices



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

> *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

$$\int \mathbb{1}(f(u) < f(y_i))du < \alpha,$$

where $\mathbb{1}$ is the indicator function, 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

$$\int \mathbb{1}(f(u) < f(y_i))du < \alpha,$$

where $\mathbb{1}$ is the indicator function, 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

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

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

The **surprisal score** is

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

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

Anomaly detection using surprisals

e.g., $F \sim N(\mu, \sigma^2)$

Outline

Extreme value theory

- GEV
- GPD

Application to French mortality

Application to Wine quality and prices

Outline

Bandwidth selection

Persistent homology

Application to Wine quality and prices

Outline