Overview and objectives: This notebook shows how Harbinger’s utility functions for distance aggregation, thresholding, and grouping affect anomaly flags and decision thresholds. We compare Gaussian 3‑sigma vs. boxplot/IQR vs. ratio rules, and grouping strategies for contiguous detections.

# Install Harbinger (if needed)  
#install.packages("harbinger")

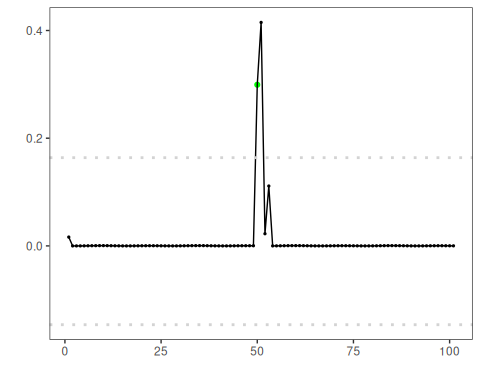
# Load required packages  
library(daltoolbox)  
library(harbinger)

# Instantiate utilities  
hutils <- harutils()

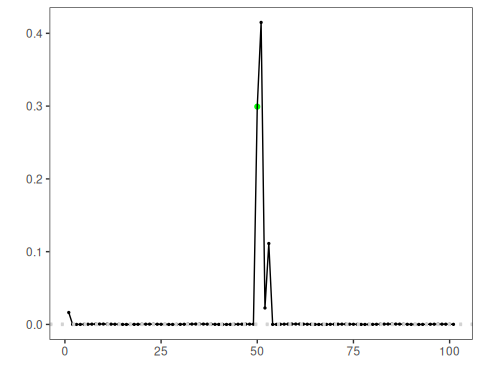
# Load a simple anomaly dataset and plot it  
data(examples\_anomalies)  
dataset <- examples\_anomalies$simple  
har\_plot(harbinger(), dataset$serie)



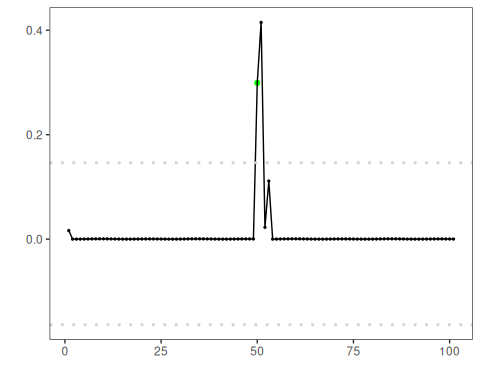
# Baseline: ARIMA with default distance (L2) and threshold (Gaussian 3-sigma)  
model <- hanr\_arima()  
model <- fit(model, dataset$serie)  
detection <- detect(model, dataset$serie)  
har\_plot(model, attr(detection, "res"), detection, dataset$event, yline = attr(detection, "threshold"))



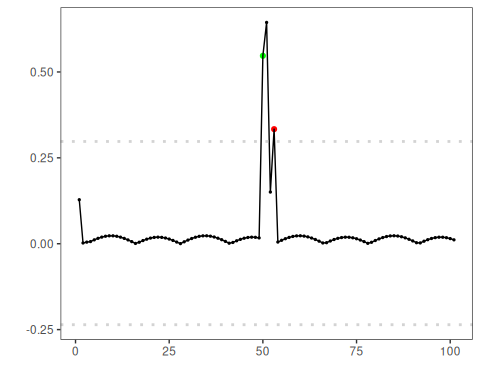
# Use Boxplot/IQR threshold instead of Gaussian  
model <- hanr\_arima()  
model$har\_outliers <- hutils$har\_outliers\_boxplot  
model <- fit(model, dataset$serie)  
detection <- detect(model, dataset$serie)  
har\_plot(model, attr(detection, "res"), detection, dataset$event, yline = attr(detection, "threshold"))



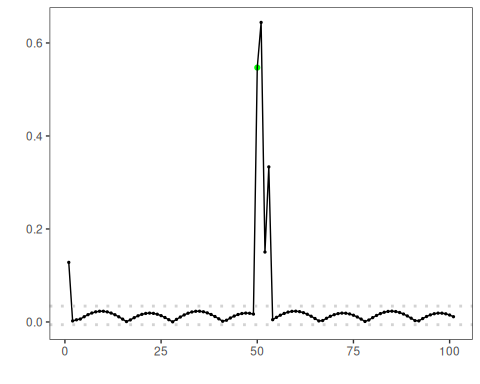
# Use ratio thresholding emphasizing relative deviation  
model <- hanr\_arima()  
model$har\_outliers <- hutils$har\_outliers\_ratio  
model <- fit(model, dataset$serie)  
detection <- detect(model, dataset$serie)  
har\_plot(model, attr(detection, "res"), detection, dataset$event, yline = attr(detection, "threshold"))



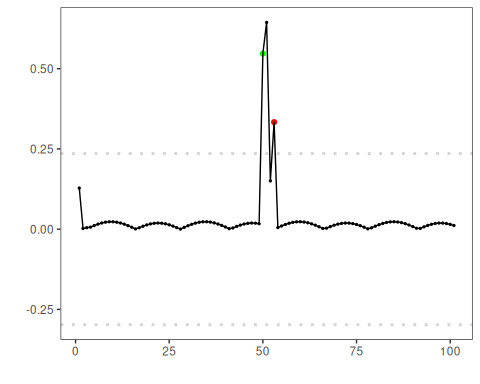
# Change distance to L1 (absolute deviation)  
model <- hanr\_arima()  
model$har\_distance <- hutils$har\_distance\_l1  
model <- fit(model, dataset$serie)  
detection <- detect(model, dataset$serie)  
har\_plot(model, attr(detection, "res"), detection, dataset$event, yline = attr(detection, "threshold"))



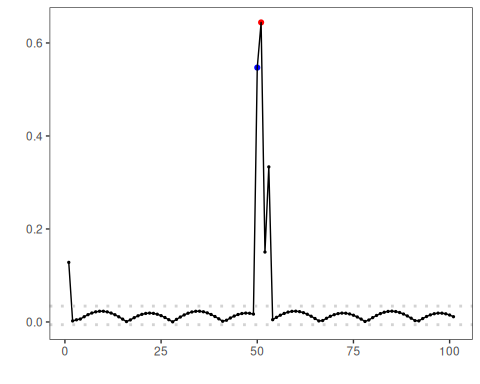
# L1 distance + Boxplot/IQR threshold  
model <- hanr\_arima()  
model$har\_distance <- hutils$har\_distance\_l1  
model$har\_outliers <- hutils$har\_outliers\_boxplot  
model <- fit(model, dataset$serie)  
detection <- detect(model, dataset$serie)  
har\_plot(model, attr(detection, "res"), detection, dataset$event, yline = attr(detection, "threshold"))



# L1 distance + ratio threshold  
model <- hanr\_arima()  
model$har\_distance <- hutils$har\_distance\_l1  
model$har\_outliers <- hutils$har\_outliers\_ratio  
model <- fit(model, dataset$serie)  
detection <- detect(model, dataset$serie)  
har\_plot(model, attr(detection, "res"), detection, dataset$event, yline = attr(detection, "threshold"))



# Keep only the highest-magnitude index in contiguous runs  
model <- hanr\_arima()  
model$har\_distance <- hutils$har\_distance\_l1  
model$har\_outliers <- hutils$har\_outliers\_boxplot  
model$har\_outliers\_check <- hutils$har\_outliers\_checks\_highgroup  
model <- fit(model, dataset$serie)  
detection <- detect(model, dataset$serie)  
har\_plot(model, attr(detection, "res"), detection, dataset$event, yline = attr(detection, "threshold"))



References - Tukey, J. W. (1977). Exploratory Data Analysis. Addison‑Wesley. (boxplot/IQR outlier rule) - Shewhart, W. A. (1931). Economic Control of Quality of Manufactured Product. D. Van Nostrand. (three‑sigma rule)