Dissert-exp-dia-hora

Gabriel Robaina

2023-11-15

Day and time experiment

The goal of this experiment is to compare data from weekends and weekdays and see if there is any performance difference between them. Another goal is to compare business hours to off hours to see if there is any performance difference there. We will divide this into analysis of write and read operations.

All data in weekends are considered to be in off hours.

Write operations analysis

The first histogram was based on the latency of all data for write operations. We can see two modes: The one in the left is the interval (700,800] and is asymmetric. The right mode is symmetrical and is located in the interval (1000,1200]. For this reason, this is an asymmetric bimodal data distribution.

```
layout(matrix(1:2, nrow = 1))
hist(data_write$LATENCY_SECONDS * 1000, main = "Histogram of Write Latency", xlab = "Latency (ms)", brea

boxplot(data_write$LATENCY_SECONDS * 1000)
```
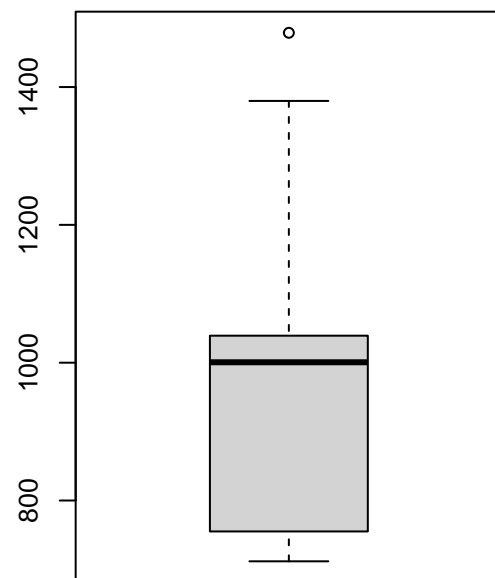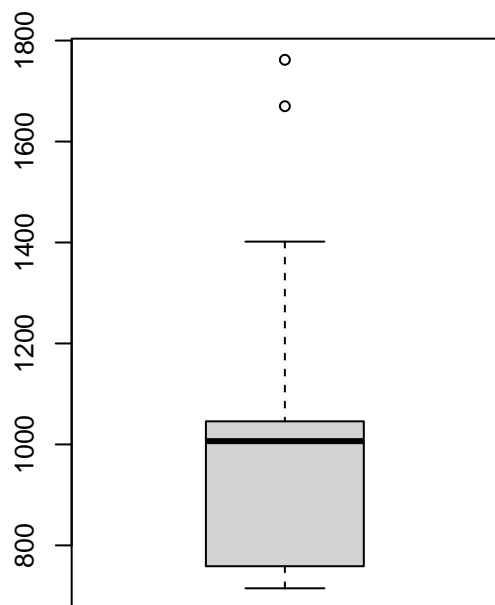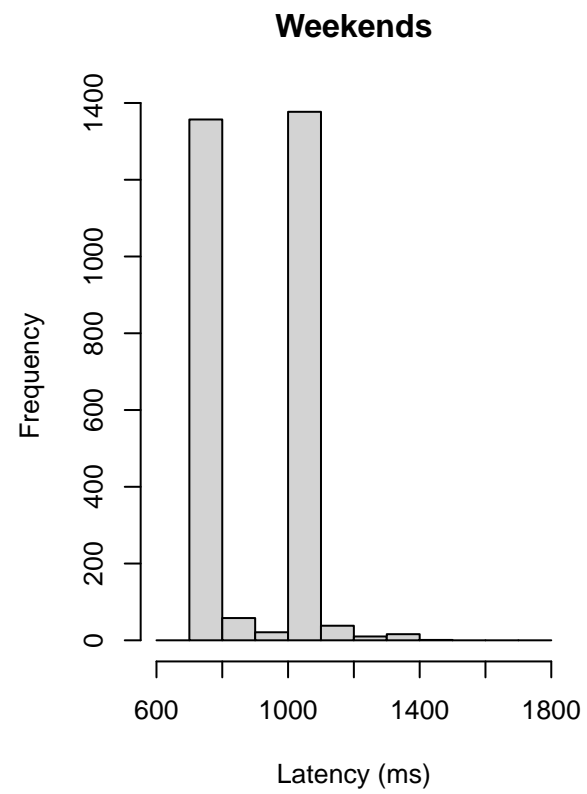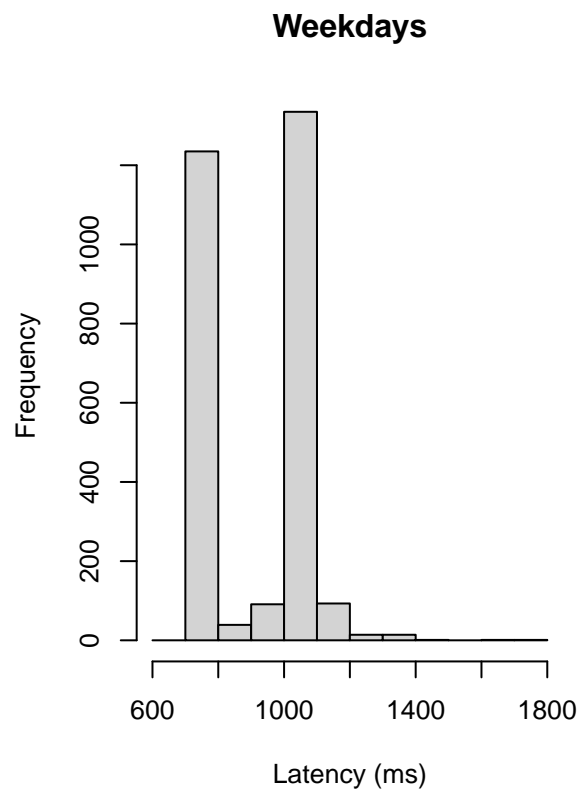
Histogram of Write Latency
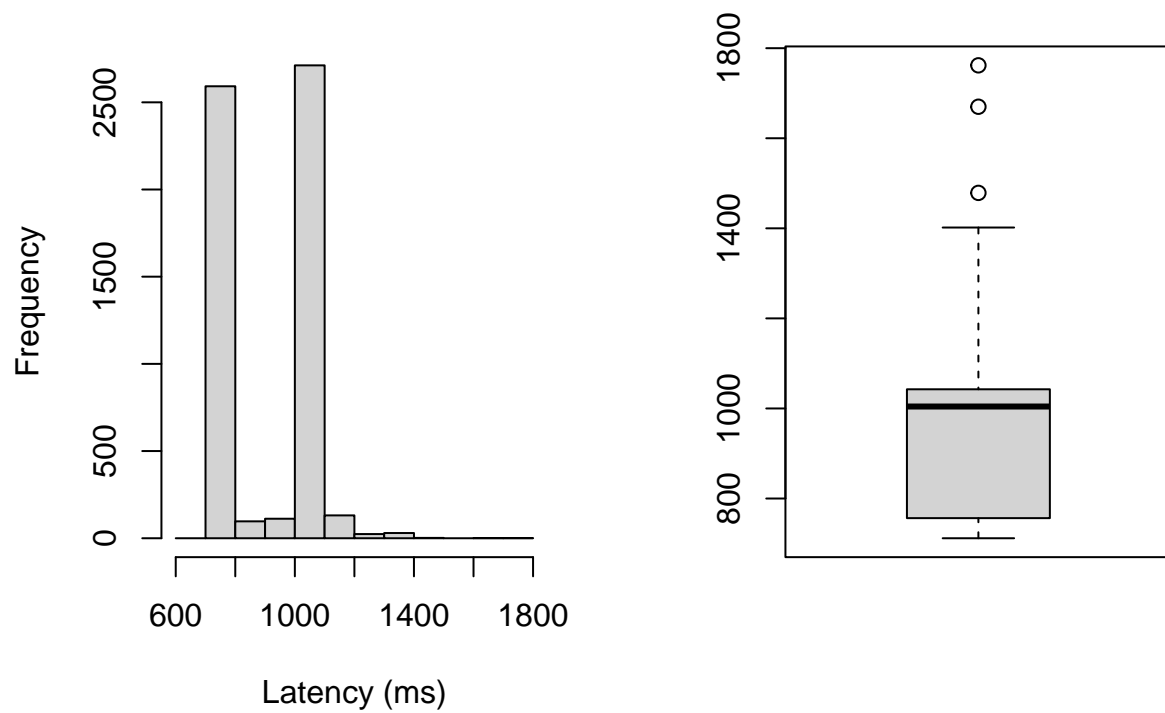

```
layout(matrix(1:1, nrow = 1))
```

Let's look for the source of both modes.

Modes in weekdays and weekends

First, lets plot histograms for both weekdays and weekends and see if they have a similar distribution.

Weekdays

Frequency

Latency (ms)

Weekends

Frequency

Latency (ms)

In both cases we have modes in the same intervals. While for weekdays the mode on the right is symmetrical, and for weekends its asymmetrical, in both cases we have asymmetric bimodal data distributions.
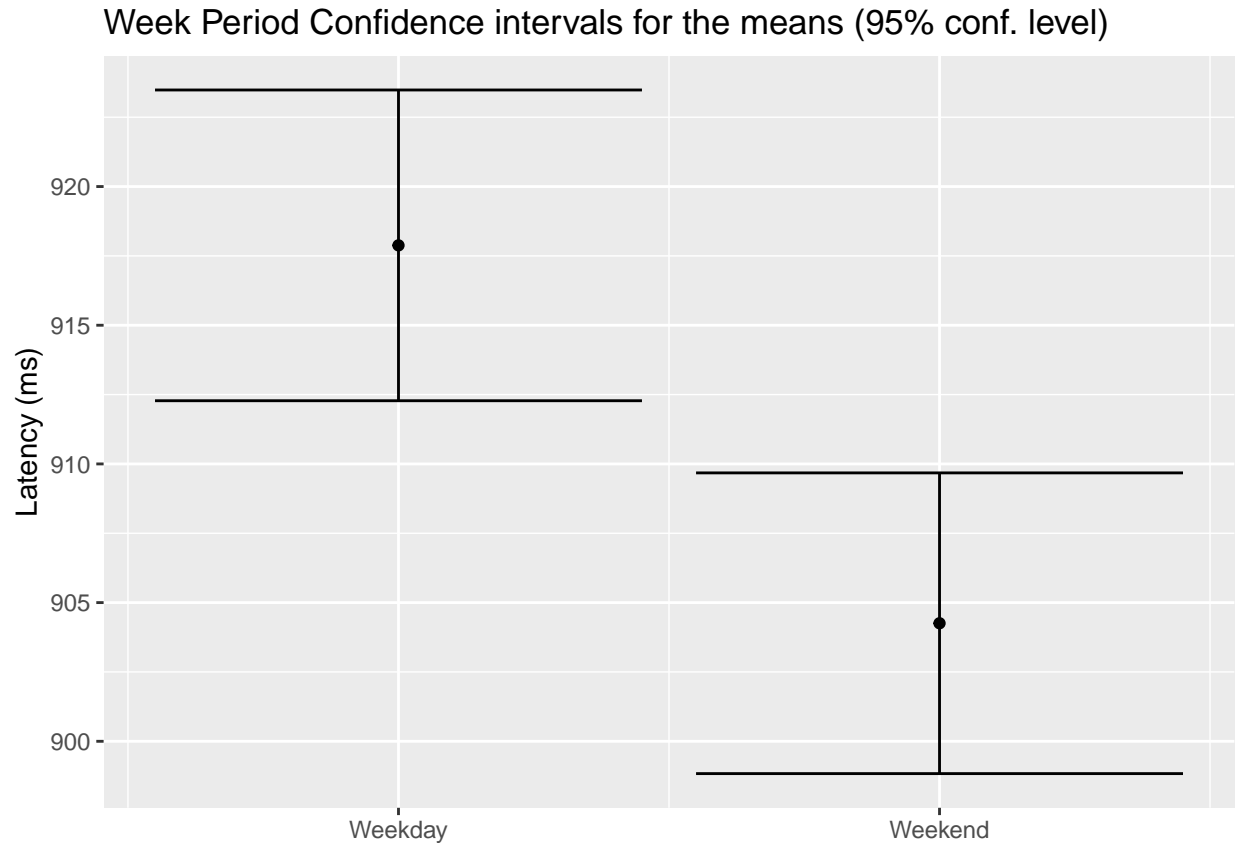
Modes business and off hours

Second, lets plot histograms for both business hours and weekends and see if they have a similar distribution.

Business hours

Off hours
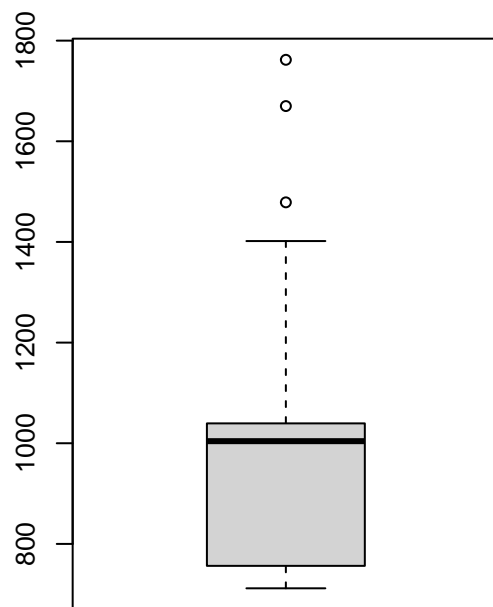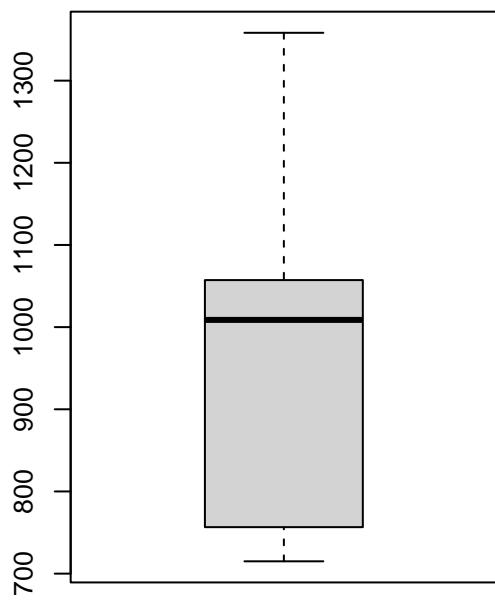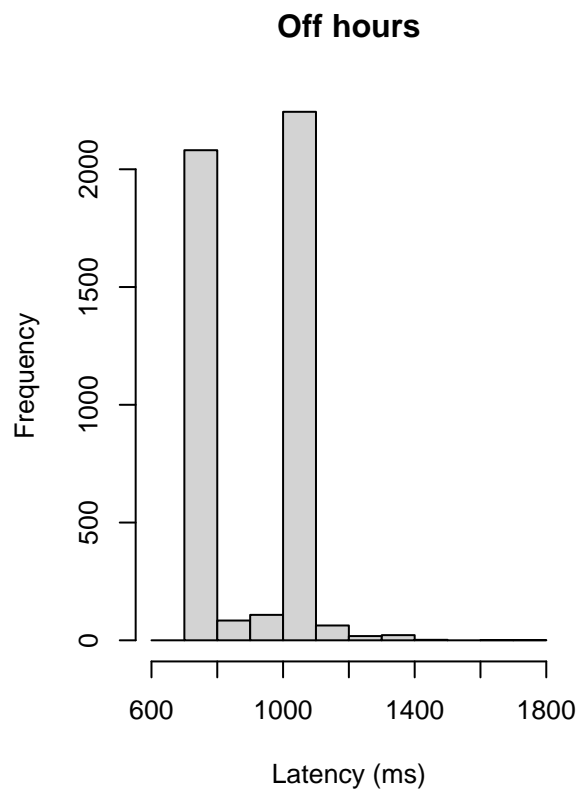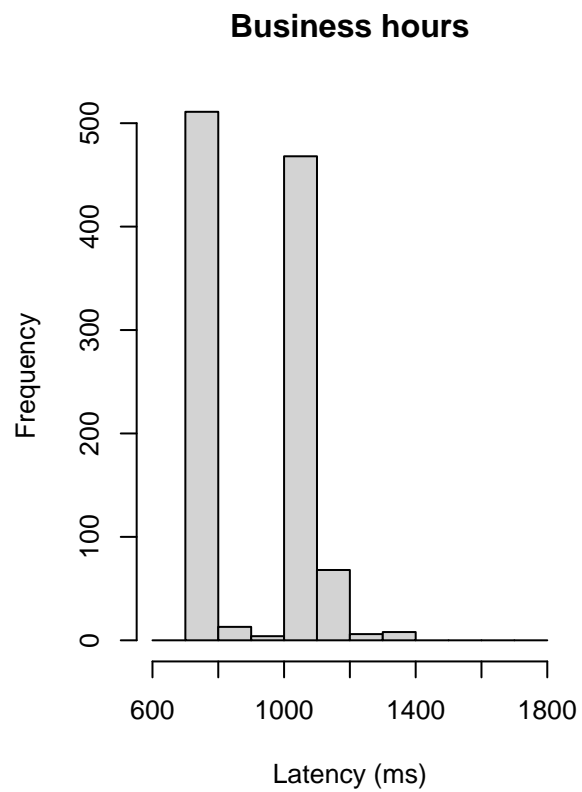
Confidence intervals

As we can see from the confidence intervals for the means on week periods, we cannot afirm there is no statistically significant difference between latencies measured during weekdays and weekends, since the confidence intervals have no overlap.

Week Period Confidence intervals for the means (95% conf. level)

Alternatively, the mean latency measured for off hours is included in the confidence interval for business hours, which means there is no statistically significant difference between latencies measured during business hours and off hours.

Time of Day Confidence intervals for the means (95% conf. level)