1(a)

a. # Replace 'path\_to\_file' with the actual file path if the file is not in the working directory

vinegar\_data <- vinegar

# Boxplot to visualize acidity differences between factory locations

boxplot(vinegar\_data$pH ~ vinegar\_data$Site, data = vinegar\_data,

xlab = "Factory Location", ylab = "Acidity",

main = "Acidity Levels across Factory Locations")A chart of a row of boxes

Description automatically generated

Interpretation

1. Variations in Median and Spread:Higher acidity level observed in Sydney location , lower acidity level observed in New York
2. iqr\_values <- tapply(vinegar\_data$pH, vinegar\_data$Site,IQR)

# Calculate IQR for each factory location

iqr\_values <- tapply(vinegar\_data$pH, vinegar\_data$Site,IQR)

View(iqr\_values)

A screenshot of a computer

Description automatically generated

1. Spread :Interquartile range highest in Sydney ,followed by Birmingham suggesting higher variability in acidity levels
2. Outliers: The outliers tending to the highest acidity values are observed in Paris, while tending to the minimum acidity values are observed in Birmingham.
3. Comparison of Acidity Levels Between Locations:Sydney has consistently higher acidity values , while New York has consistently lower acidity values.

1(b)

1. Hypothesis formulation : We will go with **Alternative Hypothesis (H1):** as There is at least one pair of factory locations with significantly different mean acidity levels.
2. Perform anova test
   1. # Perform one-way ANOVA
   2. anova\_result <- aov(vinegar\_data$pH ~ vinegar\_data$Site, data = vinegar\_data)
   3. # Summary of ANOVA results
   4. summary(anova\_result)

> summary(anova\_result)

A screenshot of a computer

Description automatically generated Df Sum Sq Mean Sq F value Pr(>F)

vinegar\_data$Site 4 24.57 6.143 6.682 0.000534 \*\*\*

Residuals 31 28.50 0.919

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Interpretation:**As P value is less than 0.05 , we reject the null hypothesis ,indicating significant differences in mean acidity levels among factory locations.As there are differences , we go for pairwise tests

> pairwise\_result <- pairwise.t.test(vinegar\_data$pH, vinegar\_data$Site, p.adj = "none")

> pairwise\_result

Pairwise comparisons using t tests with pooled SD

data: vinegar\_data$pH and vinegar\_data$Site

Birmingham London New York Paris

London 0.10843 - - -

New York 0.07349 0.88535 - -

Paris 0.66869 0.04539 0.02842 -

Sydney 0.01358 0.00017 7.6e-05 0.03654

P value adjustment method: none

A screenshot of a computer

Description automatically generated

the value 0.04539 for the comparison between Paris and London.

The p-value (0.04539) is below the conventional significance level of 0.05.

Interpretation: There's evidence to reject the null hypothesis for the pH levels between Paris and London. It suggests a statistically significant difference in mean pH levels between these two locations.