STAT 3011 Discussion 015

Introduction to R: Week 4

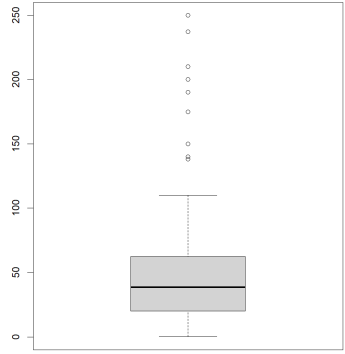
Talha Hamza
University of Minnesota

Spring 2025

Boxplot

Vanilla Boxplot: `boxplot(vector_name)` will draw a single boxplot for one distribution.

`For instance: boxplot("m_new$Budget")`



Side-by-Side Boxplot

- Compare multiple distributions in one graph.
- Two scenarios:
  1. Divide one distribution into different levels.
  2. Compare multiple distributions.

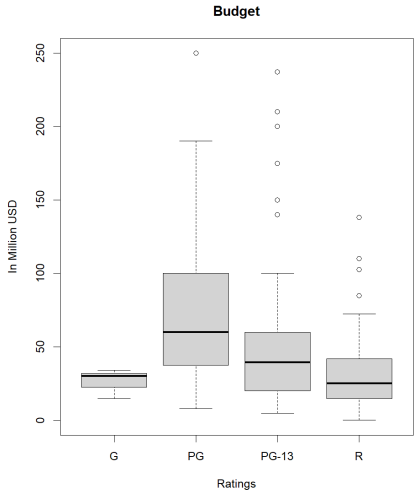
Scenario 1: Dividing a Distribution Across Levels

- A single quantitative variable is split into levels based on a categorical variable.
- In R, ~ (called tilde) represents a response-explanatory relationship.
- Format: `response_variable ~ explanatory_variable`
- In lab manual: quantitative variable with respect to categorical variable

Example:

```
boxplot(m_new$Budget ~ m_new$Rating,
        xlab = "Ratings", ylab = "In Million USD",
        main = "Budget")
```

Scenario 1: Output



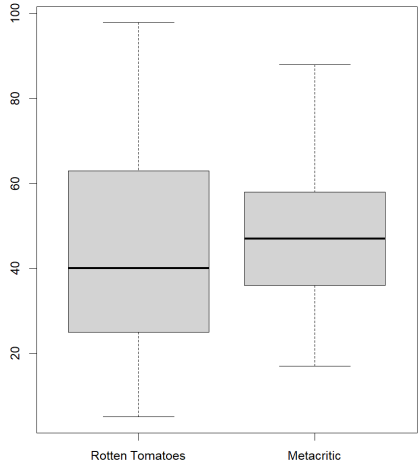
Scenario 2: Comparing Multiple Distributions

- We use a , (comma) to add multiple variables whose distributions we want to compare within a single boxplot visualization.
- In lab manual: quantitative vs. quantitative

Example Code:

```
boxplot(m_new$RT, m_new$Metacritic,
        names = c("Rotten Tomatoes", "Metacritic"),
        main = "Review scores")
```

Scenario 2: Output



Boxplot Customization

Optional Arguments:

- `xlab`: X-axis label.
- `ylab`: Y-axis label.
- `names`: Labels for each boxplot.
- `main`: Main title.

Note: Use `c()` to create a vector of labels.

Probability Formulas

General Addition Property of Probability:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Independence Rule:
If events A and B are independent, then:

$$P(A \cap B) = P(A) \cdot P(B)$$

Conditional Probability Formula:
Read as Probability of event A occuring GIVEN that B has already occurred

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{where } P(B) > 0$$

Understanding $P(A \cup B)$ and $P(A \cap B)$

$P(A \cup B)$: The probability of **either A or B** occurring.

- This includes the possibility of only A occurring, only B occurring, or both A and B occurring.
- Formula:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$P(A \cap B)$: The probability of **both A and B** occurring.

- Formula:

$$P(A \cap B) = P(A) \cdot P(B) \quad \text{(if A and B are independent)}$$

Questions? Let's Discuss!