**Experiment 3.3**

**Student Name: Rahul Kumar UID: 20BCS7081**

**Branch: BE-CSE Section/Group:20BCS\_DM-716 B**

**Semester: 6 Date of Performance: 6 May 2023**

**Subject Name: DM LAB Subject Code: 20CSP\_376**

**AIM :-**

Outlier detection using R programming.

**Theory And Output :-**

## **What are outliers?**

Data points far from the dataset’s other points are considered outliers. This refers to the data values dispersed among other data values and upsetting the dataset’s general distribution.

**Effects of an outlier on model:**

* The format of the data appears to be skewed.
* Modifies the mean, variance, and other statistical characteristics of the data’s overall distribution.
* Leads to the model’s accuracy level being biased.

### boxplot() function:

Boxplots are created by using the [boxplot() function](https://www.geeksforgeeks.org/boxplots-in-r-language/) in the R programming language.

**Syntax:**boxplot(x, data, notch, varwidth, names, main)

**Parameters:**

* **x:**This parameter sets as a vector or a formula.
* **data:** This parameter sets the data frame.
* **notch:** This parameter is the label for horizontal axis.
* **varwidth:** This parameter is a logical value. Set as true to draw width of the box proportionate to the sample size.
* **main:** This parameter is the title of the chart.
* **names:** This parameter are the group labels that will be showed under each boxplot.

**Code And Output –**

data <- rnorm(500)

data[1:10] <- c(46,9,15,-90,

42,50,-82,74,61,-32)

data

data <- rnorm(500)

data[1:10] <- c(46,9,15,-90,

42,50,-82,74,61,-32)

boxplot(data)

data <- data[!data %in% boxplot.stats(data)$out]

data

data <- rnorm(500)

data[1:10] <- c(46,9,15,-90,42,50,-82,74,61,-32)

data <- data[!data %in% boxplot.stats(data)$out]

boxplot(data)

**# Plot the chart.**

**boxplot(mpg ~ cyl, data = mtcars,**

**xlab = "Number of Cylinders",**

**ylab = "Miles Per Gallon",**

**main = "Mileage Data")**

**set.seed(20000)**

**data <- data.frame( A = rpois(900, 3),**

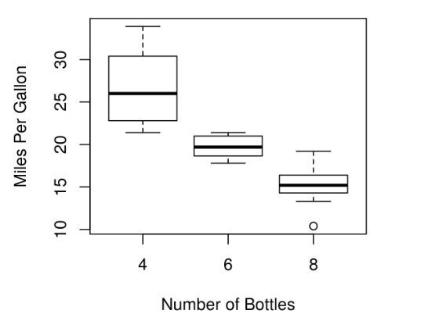
**B = rnorm(900),**

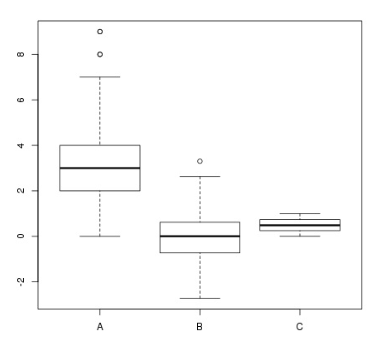
**C = runif(900)**

**)**

**# Applying boxplot function**

**boxplot(data)**

****

****

