

Question 3

Sheen Thusoo

Write a function called `drawBoxPlot` that takes in a single input that is labelled `df`, an $N \times m$ data frame containing numeric data. This function should produce a single figure with m boxplots that share a common scale.

- Use the `boxplot.stats` to find the points for the figures.
- Use the first 4 columns of the Iris data-set (length and width of sepals and petals) to demonstrate the `drawBoxPlot` function.

First, getting the first 4 columns of the Iris dataset.

```
iris_data <- read.csv("Iris.csv")
iris_df <- iris_data[,c("SepalLength", "SepalWidth", "PetalLength", "PetalWidth")]
```

Function:

```
drawBoxPlot <- function(df) {

  # Empty plot & legend
  plot(1:ncol(df), type="n", xlab = "Variate", ylab = "", main = "Boxplots",
       xlim = c(0, ncol(df) + 0.5), ylim = c(0, max(df)), xaxt="n")

  legend(4, 6, legend=c("Outliers"),
        col=c("black"), pch=(19), cex=0.5)

  axis(1, at=1:ncol(df), labels=colnames(df))

  for (i in 1:ncol(df)) {
    col_stats <- boxplot.stats(df[, colnames(df)[i]], do.out = TRUE)

    # Plot Outliers
    if (isFALSE(length(col_stats$out) == 0)) {
      for (outlier in 1: length(col_stats$out)) {
        points(i, col_stats$out[outlier], pch = 19, cex = 0.5)
      }
    }

    # Boxplot
    rect(i-0.2, as.numeric(unlist(col_stats$stats[2])), i+0.2,
        as.numeric(unlist(col_stats$stats[3])), col = "#e0e0f3") # rect from Q1 to median
    rect(i-0.2, as.numeric(unlist(col_stats$stats[3])), i+0.2,
        as.numeric(unlist(col_stats$stats[4])), col= "#e0e0f3") # rect from median to Q3
  }
}
```

```

# Vertical Lines
lines(c(i,i), c(col_stats$stats[1], col_stats$stats[2]))
lines(c(i,i), c(col_stats$stats[4], col_stats$stats[5]))

# Whisker Lines
lines(c(i-0.1,i+0.1), c(col_stats$stats[1], col_stats$stats[1]))
lines(c(i-0.1,i+0.1), c(col_stats$stats[5], col_stats$stats[5]))

}
}

```

Call the function:

```
drawBoxPlot(iris_df)
```

