```r
#  required libraries
library(openxlsx)
library(ggplot2)
library(readxl)

# Read data from PH file
phFile <- "~/Data-Science-G7/PH.xlsx"
phData <- read_excel(phFile, sheet = 1)

# Read data from DATA file
excel_file <- "~/Data-Science-G7/DATA.xlsx"
timeVsSalamander <- read_excel(excel_file, sheet = 1)
OxyVsSalamander <- read_excel(excel_file, sheet = 2)

# Read data from flowRate file
flowRatefile <- "~/Data-Science-G7/flow_1.xlsx"
flowVsSalamander <- read_excel(flowRatefile, sheet = 3)

# Rename the column from "Year" to "YEAR" in flowVsSalamander
colnames(flowVsSalamander)[colnames(flowVsSalamander) == "Year"] <- "YEAR"

# Merge data frames based on the "YEAR" column
merged_data <- merge(merge(merge(phData, timeVsSalamander, by = "YEAR"), OxyVsSalamander, by = c("YEAR"

merged_data <- merged_data[, !(names(merged_data) %in% c("Average_salamander_count"))]

summary(merged_data)
```

```
##       YEAR       average_salamander   AVERAGE_PH    average_dissolved_oxygen
##  Min.   :2000   Min.   : 0.8637     Min.   :7.499   Min.   :4.303
##  1st Qu.:2006   1st Qu.: 2.2095     1st Qu.:7.585   1st Qu.:5.366
##  Median :2012   Median : 4.8251     Median :7.621   Median :6.054
##  Mean   :2012   Mean   :12.0837     Mean   :7.622   Mean   :5.944
##  3rd Qu.:2017   3rd Qu.:12.3150     3rd Qu.:7.667   3rd Qu.:6.577
##  Max.   :2023   Max.   :80.8750     Max.   :7.769   Max.   :7.340
##   Average_flow
##  Min.   : 1.670
##  1st Qu.: 8.418
##  Median :13.521
##  Mean   :25.574
##  3rd Qu.:37.117
##  Max.   :88.092
```

```r
# Assuming merged_data is your data frame
cor_matrix <- cor(merged_data[, c("average_dissolved_oxygen", "AVERAGE_PH", "average_salamander", "Avera

# Print the correlation matrix
print(cor_matrix)
```

```
##                          average_dissolved_oxygen  AVERAGE_PH
## average_dissolved_oxygen                1.00000000  0.04951398
## AVERAGE_PH                              0.04951398  1.00000000
## average_salamander                     -0.23635146  0.42222523
## Average_flow                            0.51226430 -0.33051236
##                          average_salamander Average_flow
```

```
## average_dissolved_oxygen           -0.2363515     0.5122643
## AVERAGE_PH                           0.4222252    -0.3305124
## average_salamander                   1.0000000    -0.4379192
## Average_flow                        -0.4379192     1.0000000
```

```r
# Increase the size of the plotting area
par(mar = c(5, 5, 4, 2) + 1)  # Adjust the margin to make room for variable names

# Plot a heatmap without axis labels
heatmap(cor_matrix,
        col = colorRampPalette(c("blue", "white", "green"))(50),
        main = "Correlation Heatmap",
        xlab = NULL,
        ylab = NULL,
        cexRow = 1.2,  # Increase the size of the row labels
        cexCol = .45   # Increase the size of the column labels
)
```
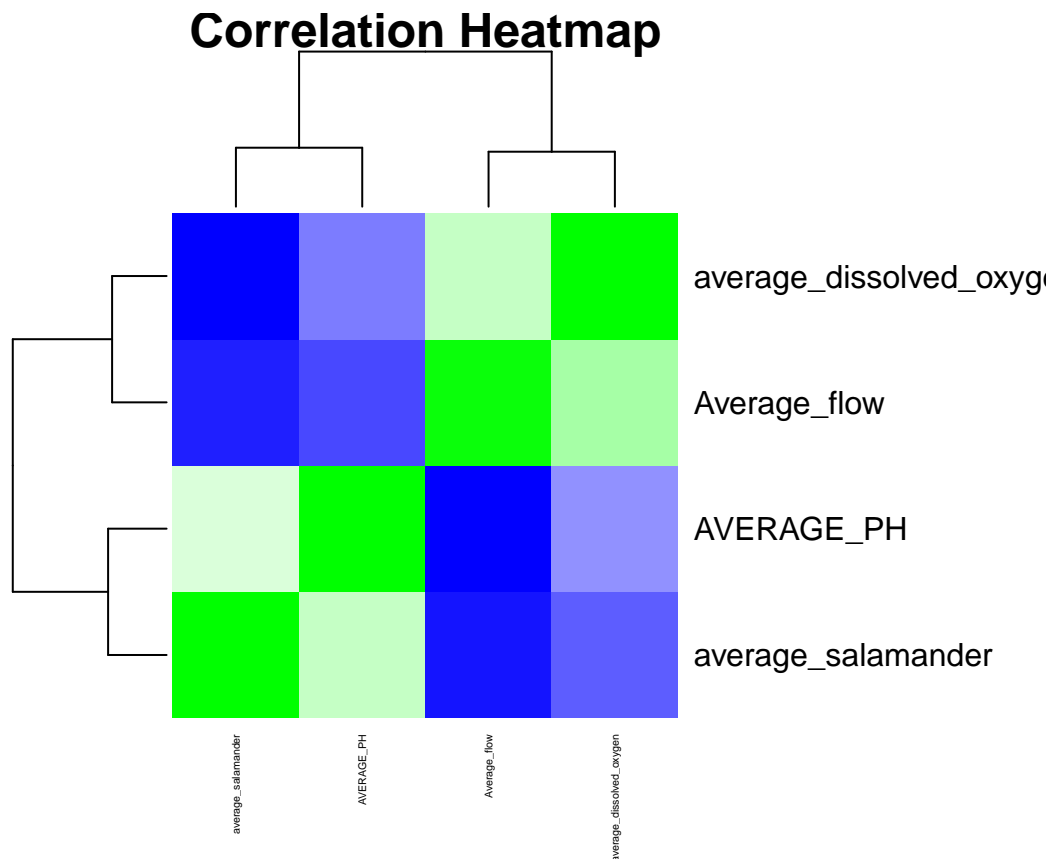


```r
model <- lm(average_salamander ~ AVERAGE_PH + average_dissolved_oxygen + Average_flow, data = merged_da

# summary of the regression model
summary(model)
```

```
##
## Call:
## lm(formula = average_salamander ~ AVERAGE_PH + average_dissolved_oxygen +
##     Average_flow, data = merged_data)
```

```
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.350  -7.316  -3.543   4.334  57.211
## 
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -714.0453   450.1544  -1.586    0.128
## AVERAGE_PH                98.0094    59.8092   1.639    0.117
## average_dissolved_oxygen  -2.7118     5.2780  -0.514    0.613
## Average_flow              -0.1887     0.1718  -1.099    0.285
## 
## Residual standard error: 16.38 on 20 degrees of freedom
## Multiple R-squared:  0.2876, Adjusted R-squared:  0.1808
## F-statistic: 2.692 on 3 and 20 DF,  p-value: 0.0737
```