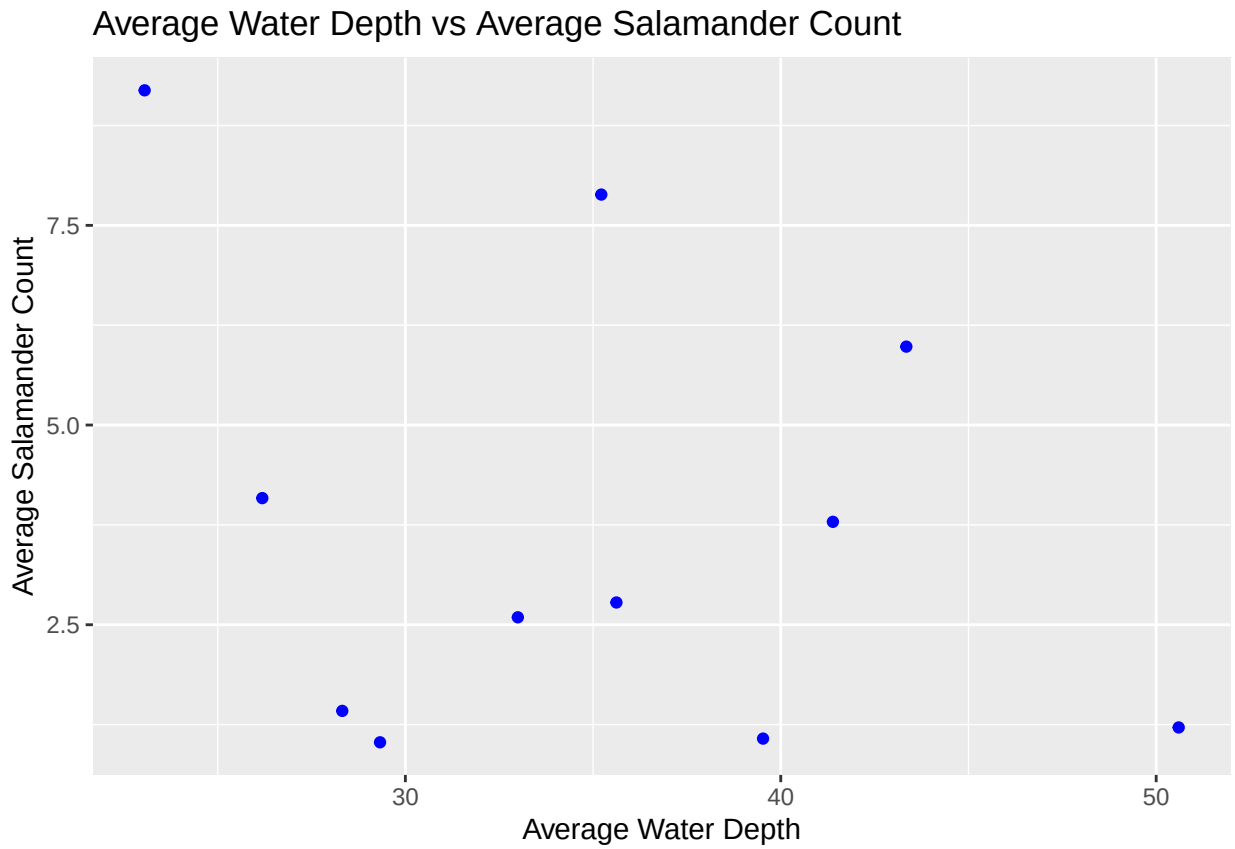
```r
waterDepth <- "~/Data-Science-G7/WaterDepth.xlsx"

#install.packages("readxl")
library(readxl)
library(ggplot2)
sheet_names = excel_sheets(waterDepth)
waterDepth <- read_excel(waterDepth, sheet = sheet_names[1])

excel_file2 <- "~/Data-Science-G7/DATA.xlsx"

#install.packages("readxl")

sheet_names2 = excel_sheets(excel_file2)
#sheet_names

salamanderCount <- read_excel(excel_file2, sheet = sheet_names2[1])
summary(salamanderCount)
```

```
##       YEAR      average_salamander
##  Min.   :1995   Min.   : 0.8637
##  1st Qu.:2002   1st Qu.: 2.5929
##  Median :2009   Median : 4.3313
##  Mean   :2009   Mean   :10.7647
##  3rd Qu.:2016   3rd Qu.: 9.4537
##  Max.   :2023   Max.   :80.8750
```

```r
# Merge data based on the common variable "YEAR"
merged_data <- merge(waterDepth, salamanderCount, by = "YEAR")


# Plot average PH on x-axis and average salamander count on y-axis
plot <- ggplot(merged_data, aes(x = WaterDepth, y = average_salamander)) +
  geom_point(color = "blue") +
  labs(x = "Average Water Depth", y = "Average Salamander Count",
       title = "Average Water Depth vs Average Salamander Count")
plot
```




```
# Install and load necessary libraries
# install.packages(c("ggplot2", "mgcv"))
library(ggplot2)
library(mgcv)
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.9-0. For overview type 'help("mgcv-package")'.
```

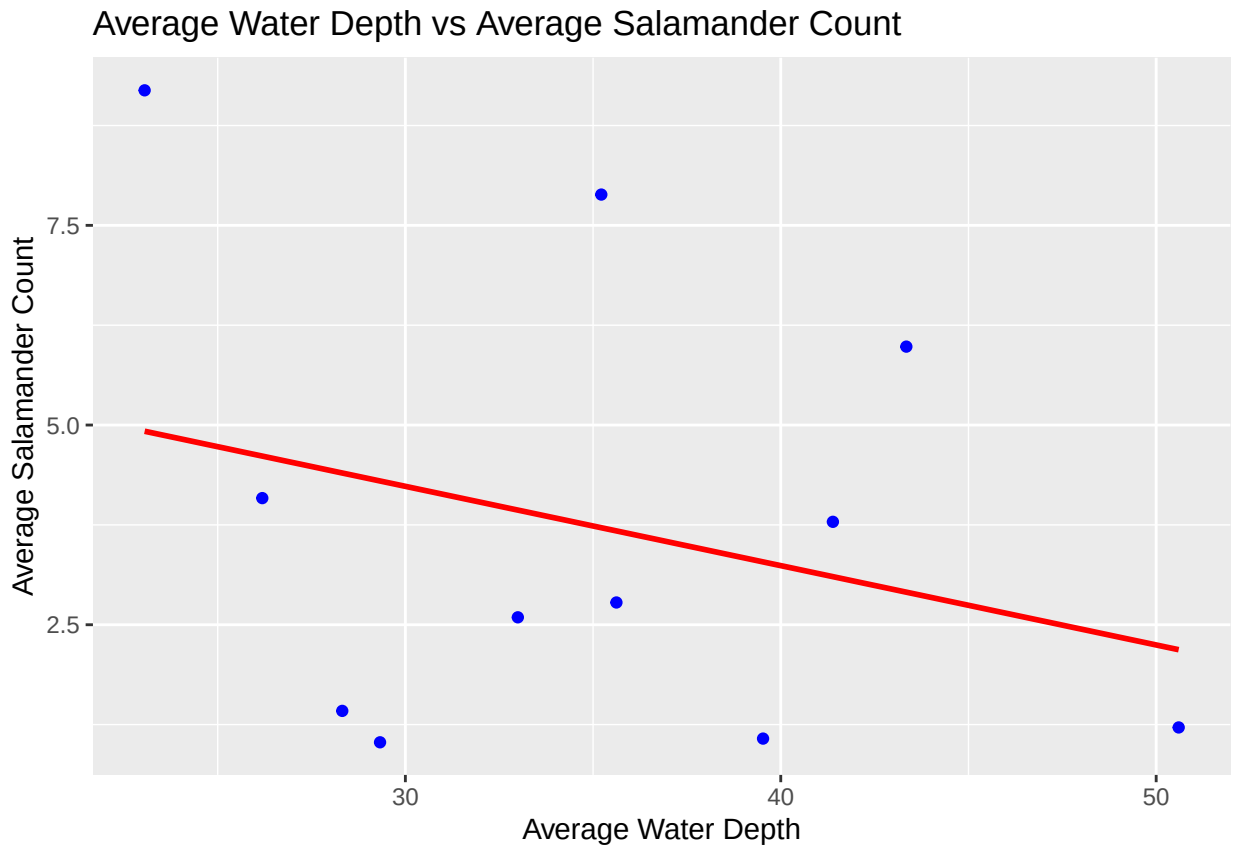
```
# Linear Regression
linear_model <- lm(average_salamander ~ WaterDepth, data = merged_data)

# Quadratic Regression
quadratic_model <- lm(average_salamander ~ poly(WaterDepth, 2), data = merged_data)

# Logarithmic Regression
log_model <- lm(log(average_salamander) ~ WaterDepth, data = merged_data)
# GAM (Generalized Additive Model)

gam_model <- gam(average_salamander ~ s(WaterDepth), data = merged_data)

# Create a scatter plot
scatter_plot <- ggplot(merged_data, aes(x = WaterDepth, y = average_salamander)) +
  geom_point(color = "blue") +
  labs(x = "Average Water Depth", y = "Average Salamander Count",
       title = "Average Water Depth vs Average Salamander Count")
```

```r
# Plot linear regression line
linear_plot <- scatter_plot +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "red")

# Plot quadratic regression line
quadratic_plot <- scatter_plot +
  geom_smooth(method = "lm", formula = y ~ poly(x, 2), se = FALSE, color = "green")

# Plot logarithmic regression line
log_plot <- ggplot(merged_data, aes(x = WaterDepth, y = log(average_salamander))) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "orange") +
  labs(x = "Average Water Depth", y = "Log(Average Salamander Count)",
       title = "Scatter Plot of Average Water Depth vs Log(Average Salamander Count)")

# Plot GAM
gam_plot <- scatter_plot +
  geom_smooth(method = "gam", formula = y ~ s(x), se = FALSE, color = "purple")
```

```r
summary(linear_model)
```

```
##
## Call:
## lm(formula = average_salamander ~ WaterDepth, data = merged_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2714 -1.7772 -0.8961  1.8803  4.2680
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.21137    3.95699   1.822    0.102
## WaterDepth  -0.09929    0.11016  -0.901    0.391
##
## Residual standard error: 2.864 on 9 degrees of freedom
## Multiple R-squared:  0.08279,    Adjusted R-squared:  -0.01913
## F-statistic: 0.8123 on 1 and 9 DF,  p-value: 0.3909
```

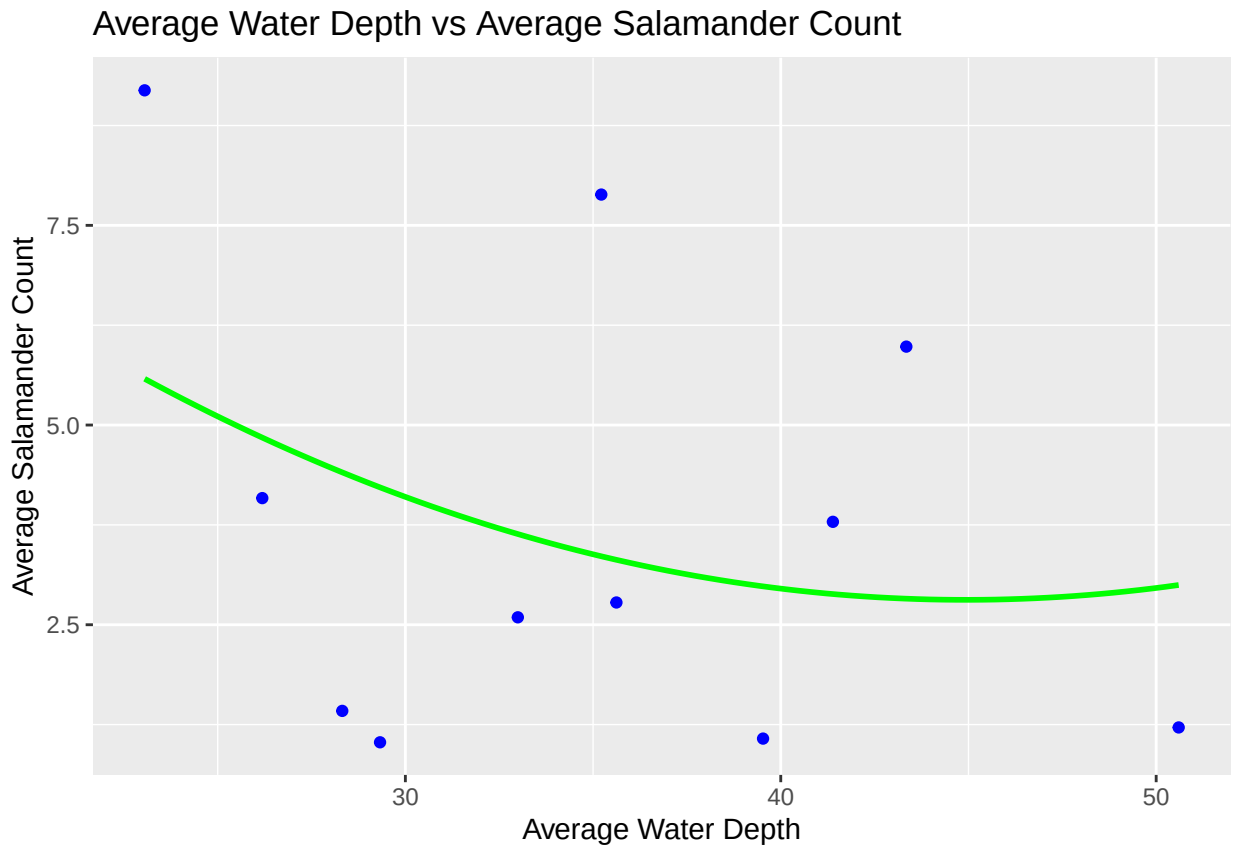
```r
print(linear_plot)
```




```
summary(quadratic_model)
```

```
##
## Call:
## lm(formula = average_salamander ~ poly(WaterDepth, 2), data = merged_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1914 -1.8445 -0.7575  2.0300  4.5284
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.7310     0.9057   4.119  0.00335 **
## poly(WaterDepth, 2)1  -2.5814     3.0038  -0.859  0.41515
## poly(WaterDepth, 2)2   1.2820     3.0038   0.427  0.68079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.004 on 8 degrees of freedom
## Multiple R-squared:  0.1032, Adjusted R-squared:  -0.121
## F-statistic: 0.4603 on 2 and 8 DF,  p-value: 0.6468
```

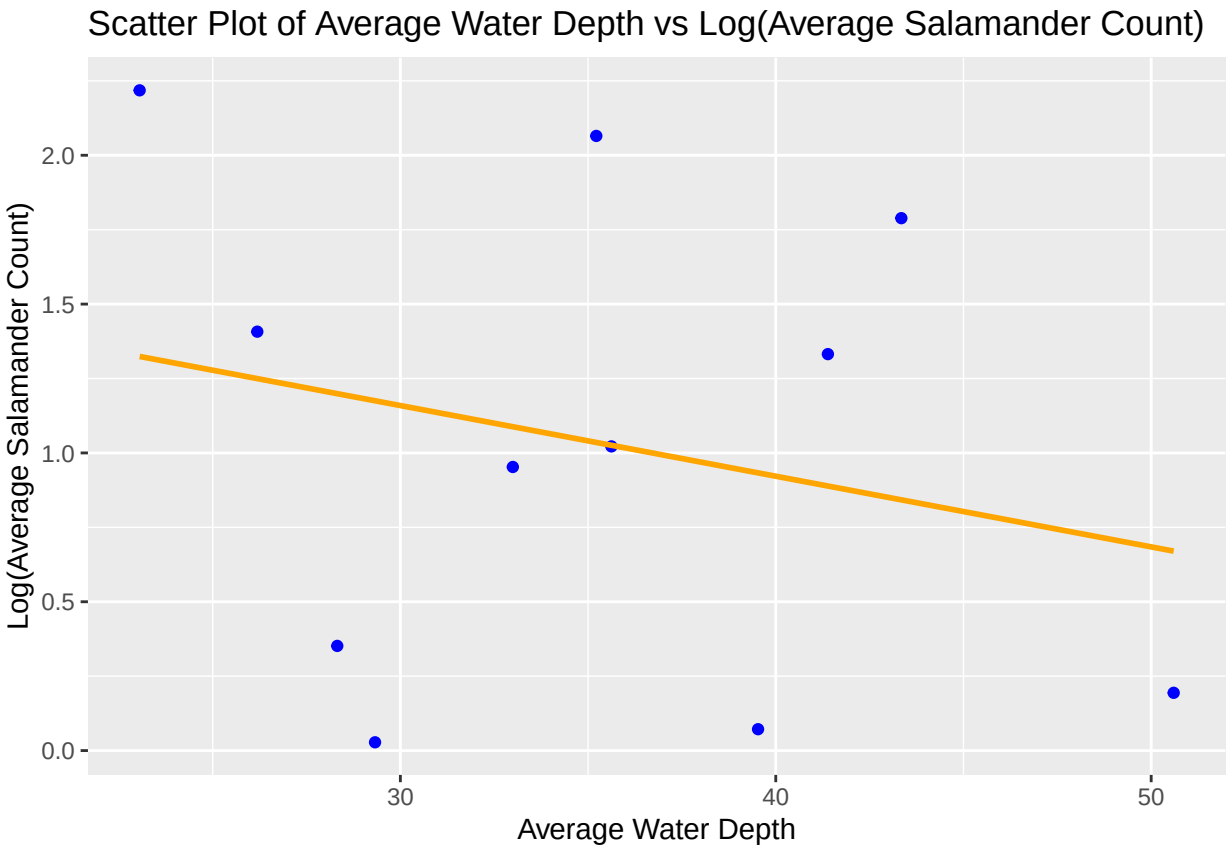
```
print(quadratic_plot)
```




```
summary(log_model)
```

```
##
## Call:
## lm(formula = log(average_salamander) ~ WaterDepth, data = merged_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.14737 -0.66175 -0.00382  0.66864  1.02974
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.87141    1.12785   1.659    0.131
## WaterDepth  -0.02374    0.03140  -0.756    0.469
##
## Residual standard error: 0.8163 on 9 degrees of freedom
## Multiple R-squared:  0.05973,    Adjusted R-squared:  -0.04475
## F-statistic: 0.5717 on 1 and 9 DF,  p-value: 0.4689
```

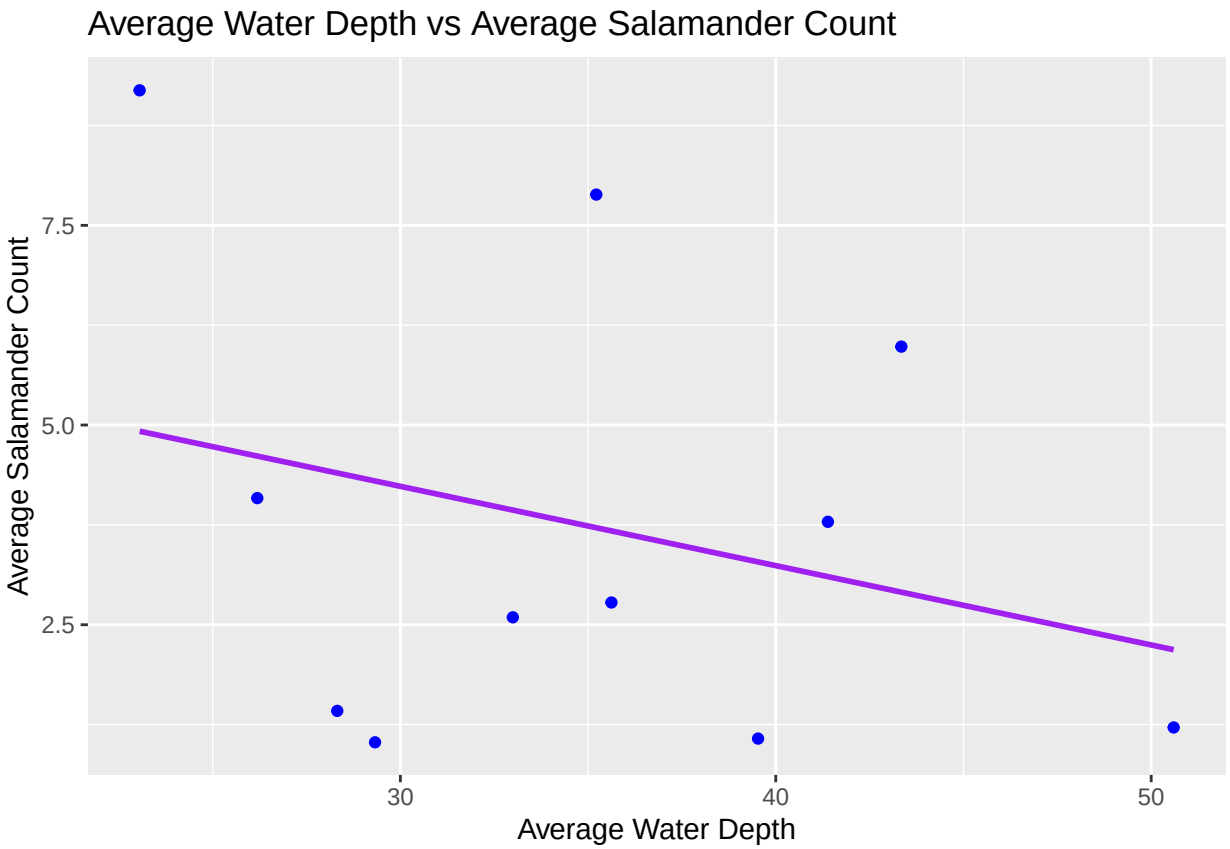
```
print(log_plot)
```




```
summary(gam_model)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## average_salamander ~ s(WaterDepth)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.7310     0.6454   5.781  0.00154 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                edf Ref.df     F p-value
## s(WaterDepth) 4.46  5.329 1.913   0.241
##
## R-sq.(adj) =  0.431   Deviance explained = 68.5%
## GCV = 9.0992  Scale est. = 4.5823    n = 11
```

```
print(gam_plot)
```


Average Water Depth vs Average Salamander Count
Average Salamander Count
7.5
5.0
2.5
30
40
50
Average Water Depth