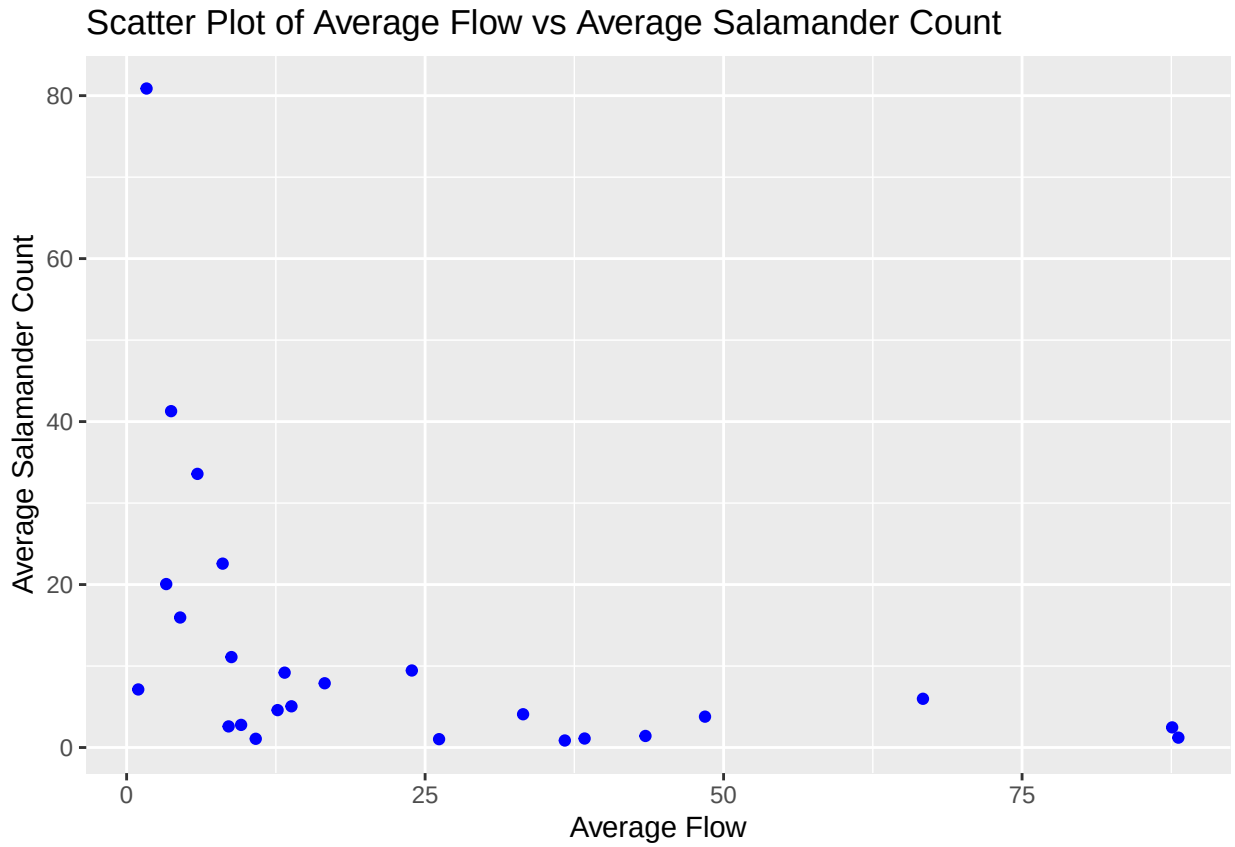
```r
# Load  libraries
library(openxlsx)
library(ggplot2)

# Load data
excel_file <- "~/Data-Science-G7/flow_1.xlsx"
flowRate <- read.xlsx(excel_file, sheet = 3)

#  scatter plot
scatter_plot <- ggplot(flowRate, aes(x = Average_flow, y = Average_salamander_count)) +
  geom_point(color = "blue") +
  labs(x = "Average Flow", y = "Average Salamander Count",
       title = "Scatter Plot of Average Flow vs Average Salamander Count")

# Display
print(scatter_plot)
```




```r
# Load necessary libraries
library(ggplot2)
library(mgcv)
```

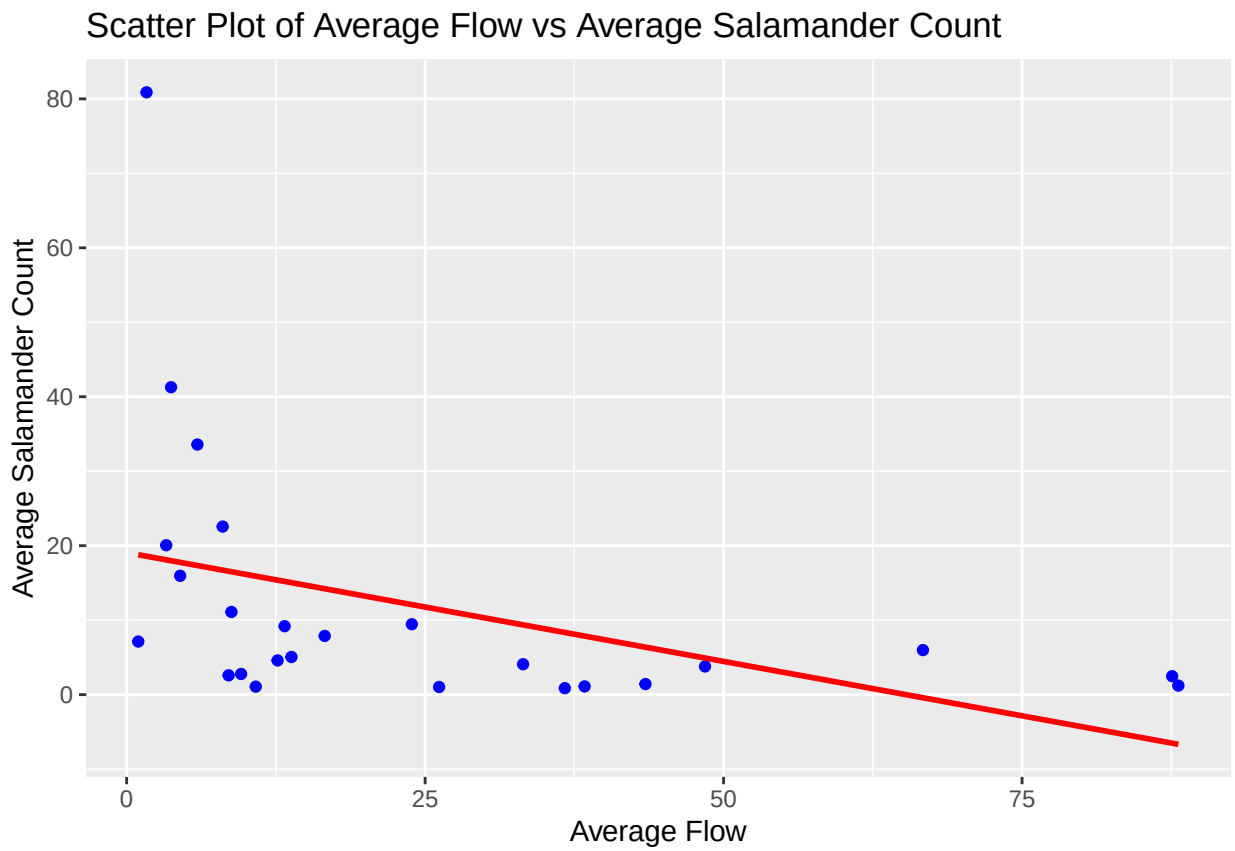
```
## Loading required package: nlme
```

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

```r
# Linear Regression
linear_model <- lm(Average_salamander_count ~ Average_flow, data = flowRate)
```

```r
# Quadratic Regression
quadratic_model <- lm(Average_salamander_count ~ poly(Average_flow, 2), data = flowRate)

# Logarithmic Regression
log_model <- lm(log(Average_salamander_count) ~ Average_flow, data = flowRate)

# GAM (Generalized Additive Model)
gam_model <- gam(Average_salamander_count ~ s(Average_flow), data = flowRate)

# Create a scatter plot
scatter_plot <- ggplot(flowRate, aes(x = Average_flow, y = Average_salamander_count)) +
  geom_point(color = "blue") +
  labs(x = "Average Flow", y = "Average Salamander Count",
       title = "Scatter Plot of Average Flow vs Average Salamander Count")

# 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(flowRate, aes(x = Average_flow, y = log(Average_salamander_count))) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "orange") +
  labs(x = "Average Flow", y = "Log(Average Salamander Count)",
       title = "Scatter Plot of Average Flow vs Log(Average Salamander Count)")

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

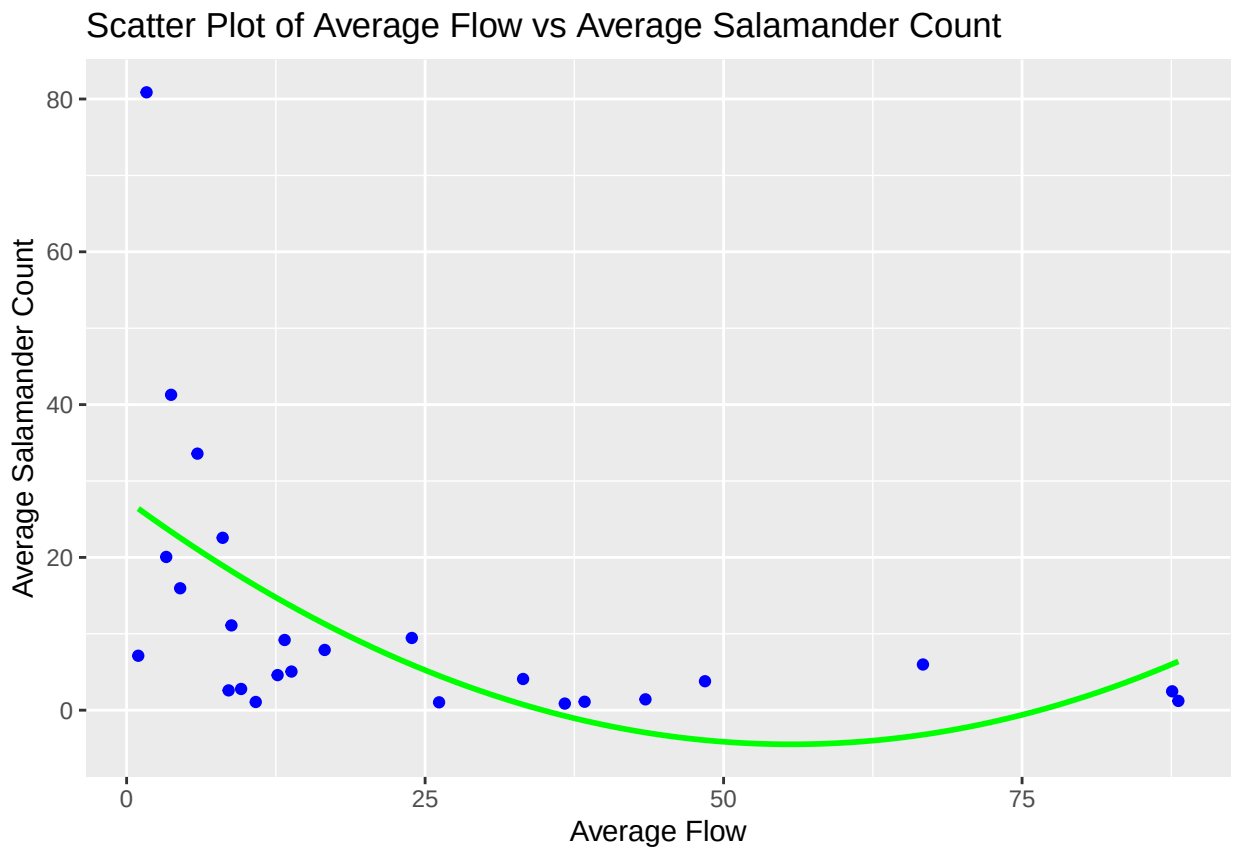
```r
# Display the plots
print(linear_plot)
```




```
summary(linear_model)
```

```
##
## Call:
## lm(formula = Average_salamander_count ~ Average_flow, data = flowRate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.835  -9.978  -5.282   5.848  62.293
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   19.0700     4.6303   4.118 0.000419 ***
## Average_flow  -0.2922     0.1324  -2.208 0.037523 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.47 on 23 degrees of freedom
## Multiple R-squared:  0.1748, Adjusted R-squared:  0.139
## F-statistic: 4.873 on 1 and 23 DF,  p-value: 0.03752
```

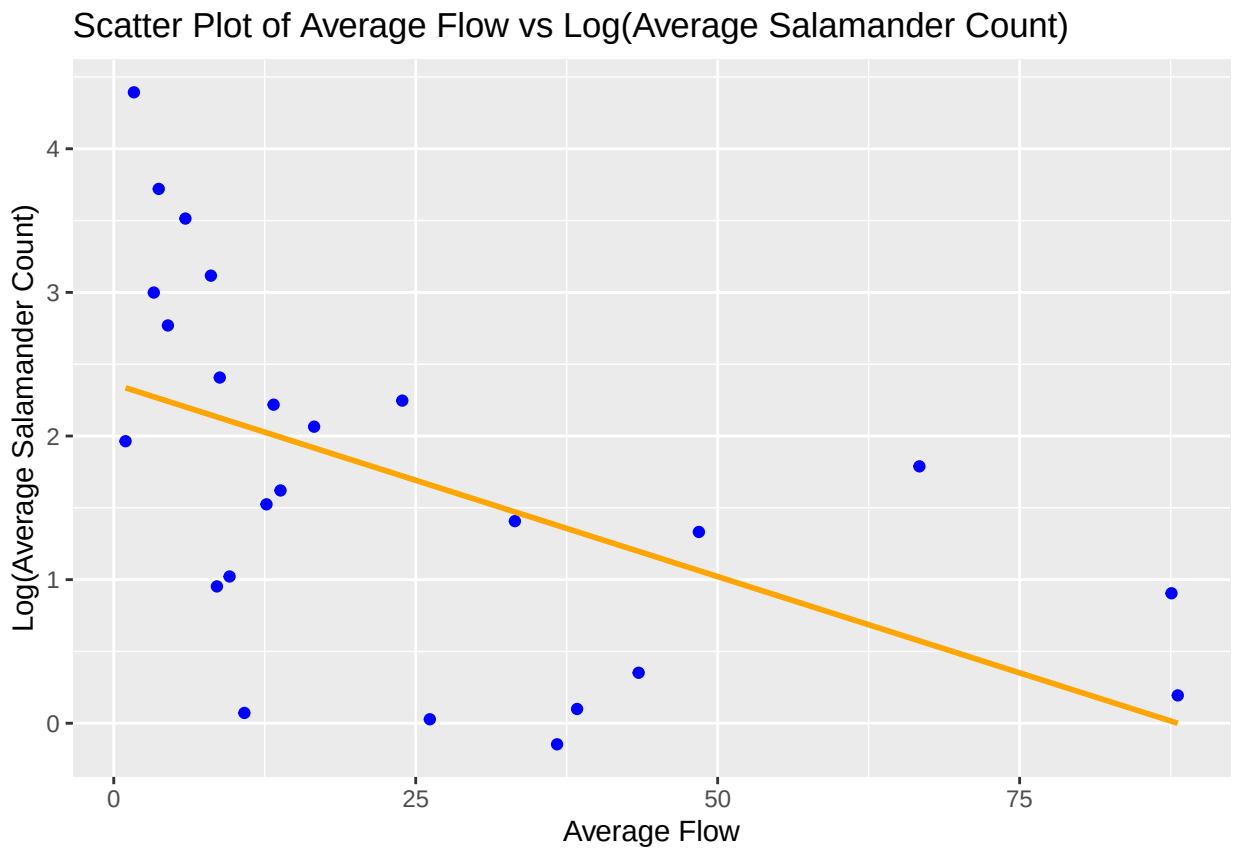
```
print(quadratic_plot)
```




```
summary(quadratic_model)
```

```
##
## Call:
## lm(formula = Average_salamander_count ~ poly(Average_flow, 2),
##     data = flowRate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.264  -7.113  -3.481   3.637  55.262
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               11.885      3.068   3.874 0.000821 ***
## poly(Average_flow, 2)1  -36.354     15.342  -2.370 0.027001 *
## poly(Average_flow, 2)2   32.547     15.342   2.121 0.045386 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.34 on 22 degrees of freedom
## Multiple R-squared:  0.315,  Adjusted R-squared:  0.2527
## F-statistic: 5.058 on 2 and 22 DF,  p-value: 0.01559
```

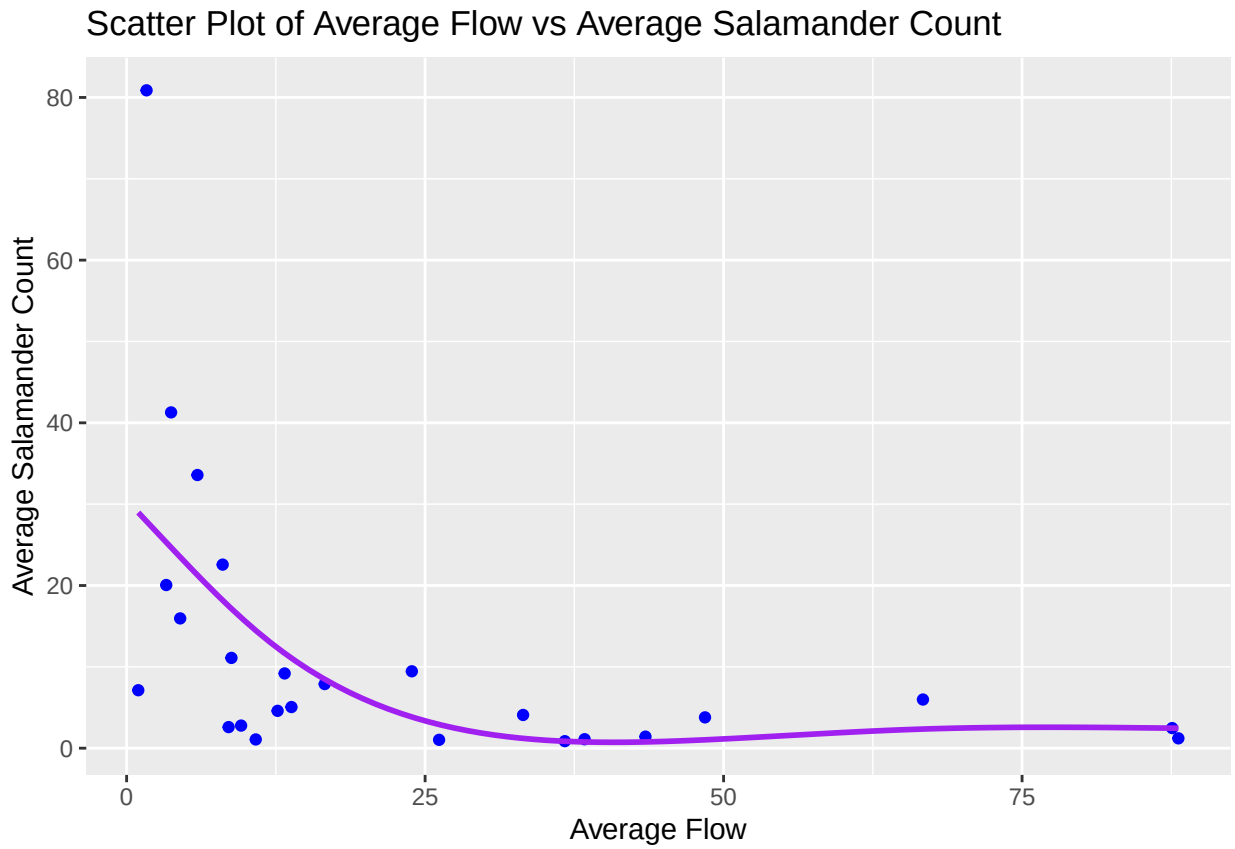
```
print(log_plot)
```




```
summary(log_model)
```

```
##
## Call:
## lm(formula = log(Average_salamander_count) ~ Average_flow, data = flowRate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0000 -0.8451  0.1941  0.7258  2.0758
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.361866   0.304772   7.750 7.37e-08 ***
## Average_flow -0.026812   0.008712  -3.078  0.00532 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.084 on 23 degrees of freedom
## Multiple R-squared:  0.2917, Adjusted R-squared:  0.2609
## F-statistic: 9.473 on 1 and 23 DF,  p-value: 0.005321
```

```
print(gam_plot)
```




```
summary(gam_model)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## Average_salamander_count ~ s(Average_flow)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.885      2.889   4.114 0.000498 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                   edf Ref.df     F p-value
## s(Average_flow) 3.089  3.796 3.685  0.0251 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.338   Deviance explained = 42.3%
## GCV = 249.45  Scale est. = 208.65    n = 25
```