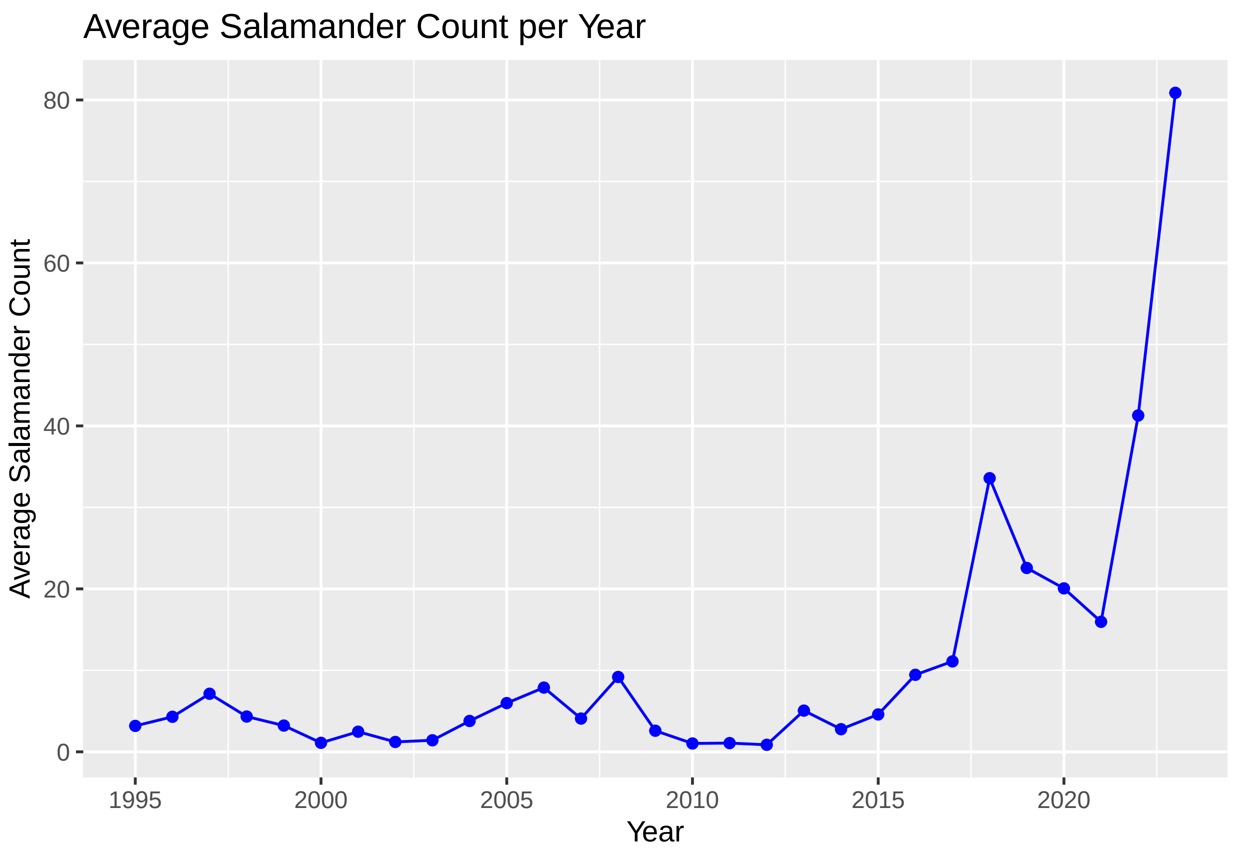
```r
library(openxlsx)
library(readxl)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

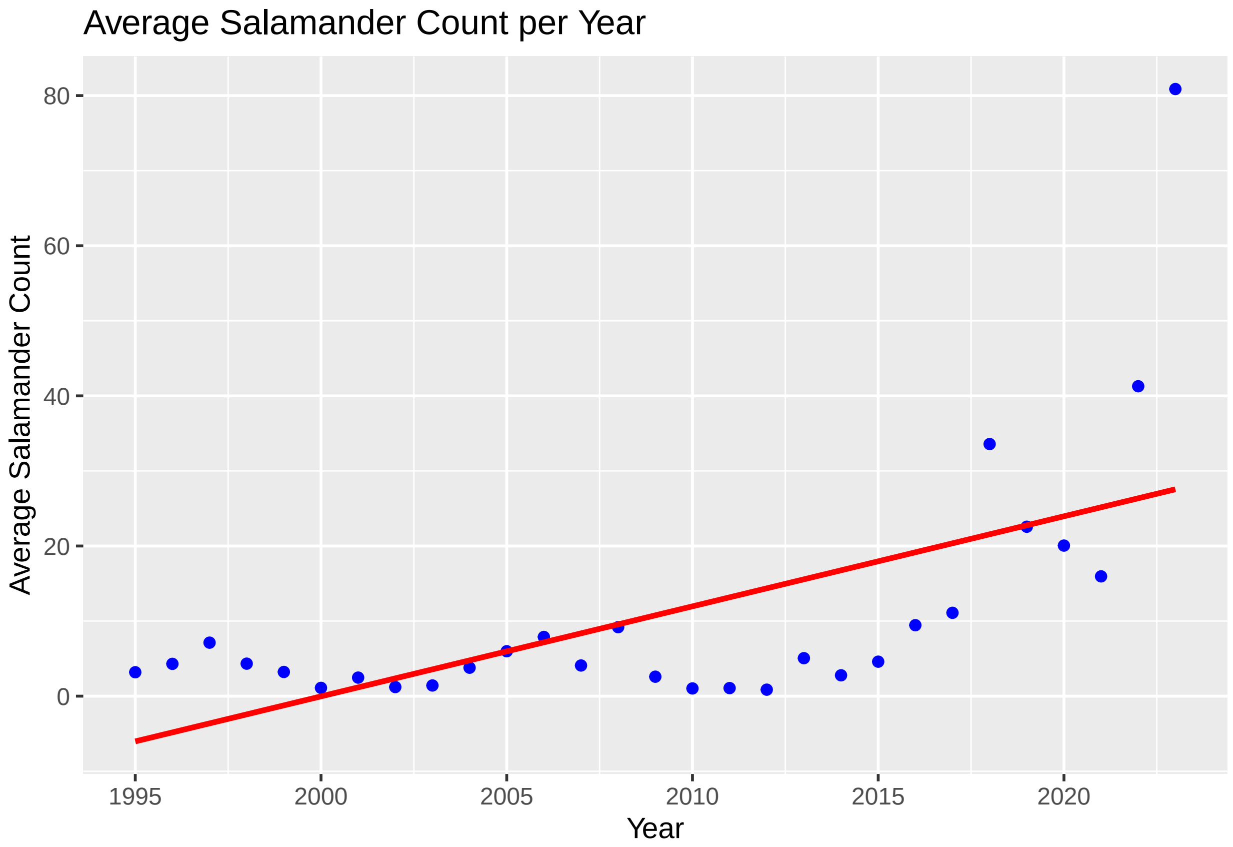
```r
excel_file <- "~/Data-Science-G7/DATA.xlsx"

#install.packages("readxl")
library(readxl)

sheet_names = excel_sheets(excel_file)
#sheet_names

avgSalpYear <- read_excel(excel_file, sheet = sheet_names[1])

ggplot(avgSalpYear, aes(x = YEAR, y = average_salamander)) +
  geom_line(color = "blue") +
  geom_point(color = "blue") +
  labs(x = "Year", y = "Average Salamander Count",
       title = "Average Salamander Count per Year ")
```



```r
# Fit a linear regression model
linear_model <- lm(average_salamander ~ YEAR, data = avgSalpYear)

# Display the summary of the linear model
summary(linear_model)
```

```
##
## Call:
## lm(formula = average_salamander ~ YEAR, data = avgSalpYear)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.983  -9.259  -0.979   4.452  53.318
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2398.888    598.705  -4.007 0.000435 ***
## YEAR            1.199      0.298   4.025 0.000414 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.43 on 27 degrees of freedom
## Multiple R-squared:  0.375,  Adjusted R-squared:  0.3518
## F-statistic:  16.2 on 1 and 27 DF,  p-value: 0.0004144
```

```r
# Visualize the linear regression line along with the data points
ggplot(avgSalpYear, aes(x = YEAR, y = average_salamander)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", se = FALSE, color = "red", formula = y ~ x) +
  labs(x = "Year", y = "Average Salamander Count",
       title = "Average Salamander Count per Year")
```

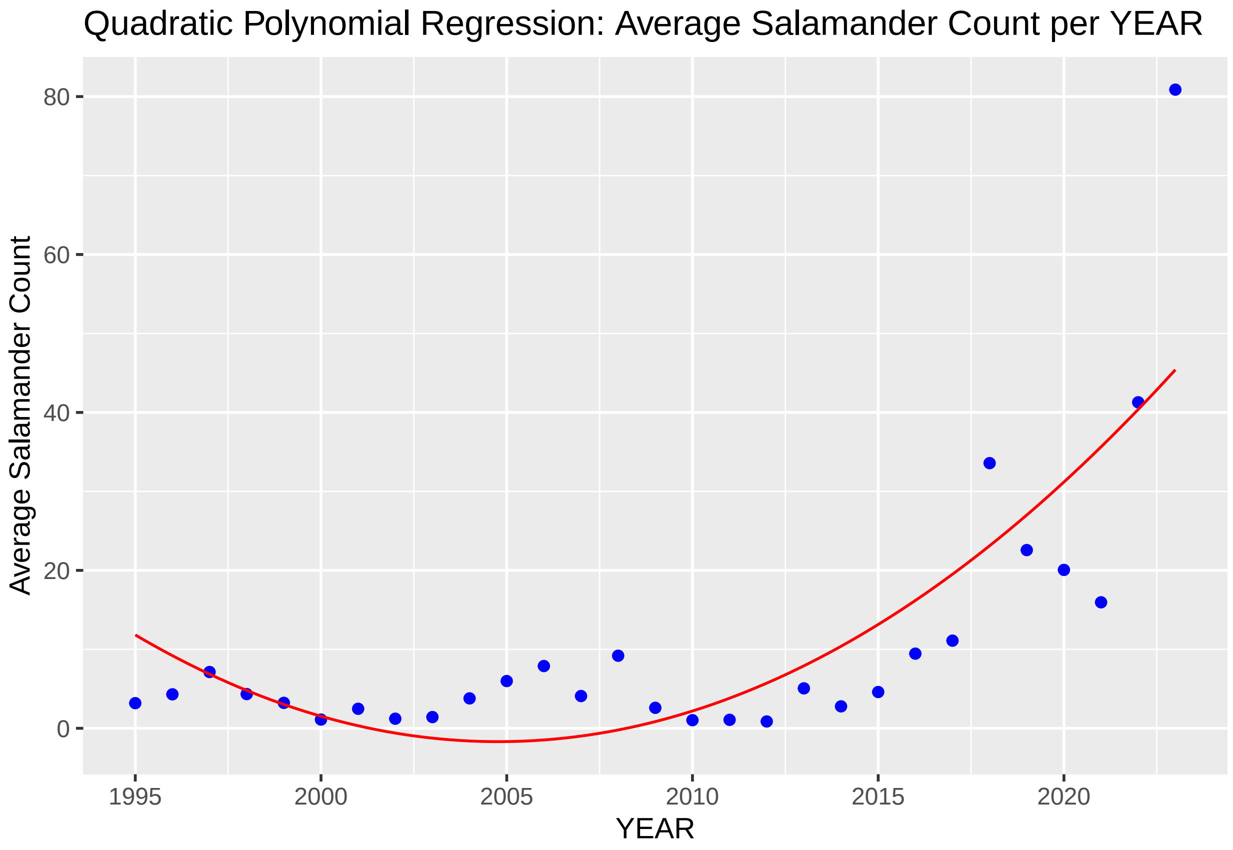


```
# Fit a quadratic polynomial regression model
quadratic_model <- lm(average_salamander ~ poly(YEAR, 2), data = avgSalpYear)

# Display the summary of the quadratic model
summary(quadratic_model)
```

```
##
## Call:
## lm(formula = average_salamander ~ poly(YEAR, 2), data = avgSalpYear)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.686  -4.896  -0.424   2.671  35.466
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      10.765      1.855   5.804 4.09e-06 ***
## poly(YEAR, 2)1   54.041      9.988   5.411 1.14e-05 ***
## poly(YEAR, 2)2   47.686      9.988   4.774 6.09e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 9.988 on 26 degrees of freedom
## Multiple R-squared:  0.667,  Adjusted R-squared:  0.6414
## F-statistic: 26.04 on 2 and 26 DF,  p-value: 6.198e-07
```

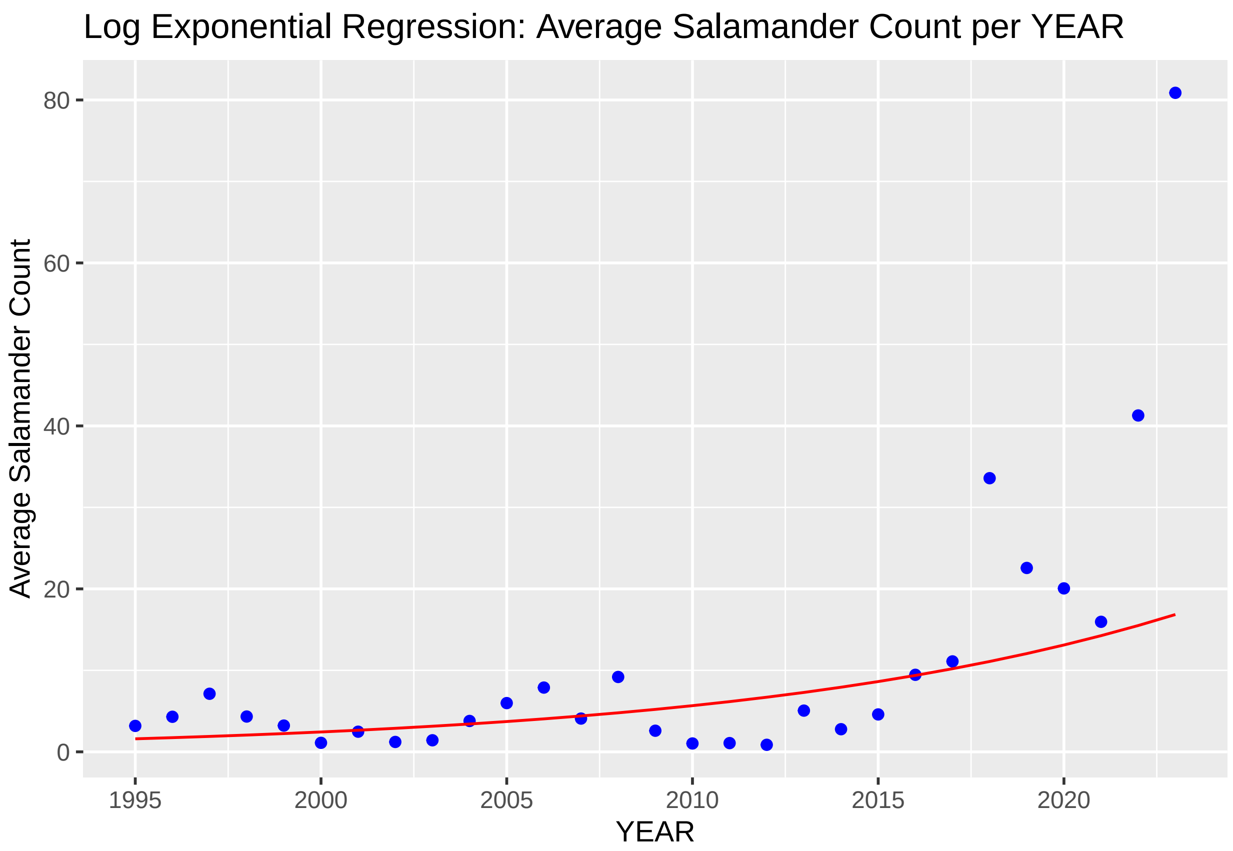
```
# Generate a sequence of values for plotting the regression curve
year_seq <- seq(min(avgSalpYear$YEAR), max(avgSalpYear$YEAR), length.out = 100)

# Predict the average dissolved oxygen using the quadratic model
predicted_values <- predict(quadratic_model, newdata = data.frame(YEAR = year_seq))

# Create a data frame for plotting
plot_data <- data.frame(YEAR = year_seq, predicted_values = predicted_values)

# Visualize the quadratic regression curve along with the data points
ggplot() +
  geom_point(data = avgSalpYear, aes(x = YEAR, y = average_salamander), color = "blue") +
  geom_line(data = plot_data, aes(x = YEAR, y = predicted_values), color = "red") +
  labs(x = "YEAR", y = "Average Salamander Count",
       title = "Quadratic Polynomial Regression: Average Salamander Count per YEAR")
```



```
#   exponential regression model
exp_model <- lm(log(average_salamander) ~ log(YEAR), data = avgSalpYear)

# summary of the exponential model
summary(exp_model)
```

```
##
## Call:
## lm(formula = log(average_salamander) ~ log(YEAR), data = avgSalpYear)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0491 -0.6976  0.1029  0.6673  1.5685
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1284.15     322.29  -3.984 0.000461 ***
## log(YEAR)     169.06      42.38   3.990 0.000455 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9504 on 27 degrees of freedom
## Multiple R-squared:  0.3709, Adjusted R-squared:  0.3476
## F-statistic: 15.92 on 1 and 27 DF,  p-value: 0.000455
```

```r
# Generate predicted values using the exponential model
predicted_values <- exp(predict(exp_model, newdata = avgSalpYear))

# exponential regression curve along with the data points
ggplot(avgSalpYear, aes(x = YEAR, y = average_salamander)) +
  geom_point(color = "blue") +
  geom_line(aes(x = YEAR, y = predicted_values), color = "red") +
  labs(x = "YEAR", y = "Average Salamander Count",
       title = "Log Exponential Regression: Average Salamander Count per YEAR")
```



```
library(mgcv)
```

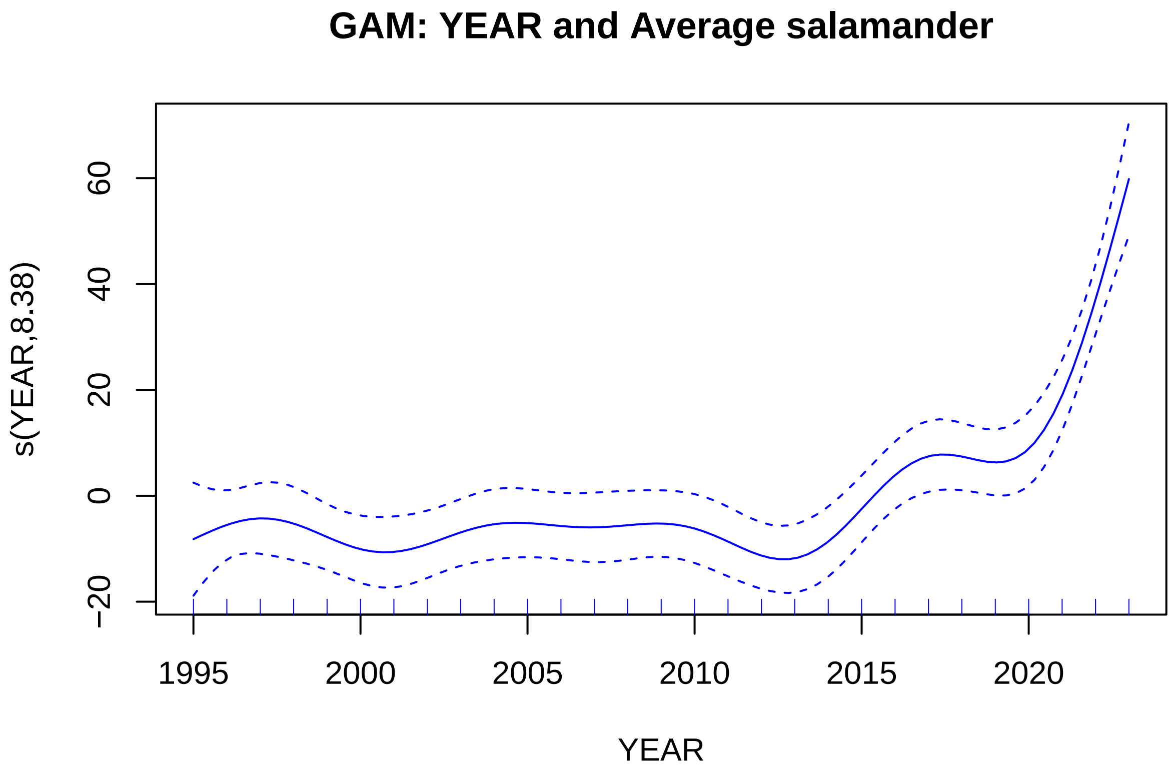
```
## Loading required package: nlme

##
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':
##
##     collapse

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

```
gam_model <- gam(average_salamander ~ s(YEAR), data = avgSalpYear)
summary(gam_model)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## average_salamander ~ s(YEAR)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   10.765      1.186   9.073 1.86e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Approximate significance of smooth terms:
##           edf Ref.df     F p-value    
## s(YEAR) 8.376  8.891 18.66  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## R-sq.(adj) =  0.853   Deviance explained = 89.7%
## GCV = 60.326  Scale est. = 40.823    n = 29
```

```
plot(gam_model, se = TRUE, col = "blue", main = "GAM: YEAR and Average salamander")
```


GAM: YEAR and Average salamander
s(YEAR,8.38)
YEAR