

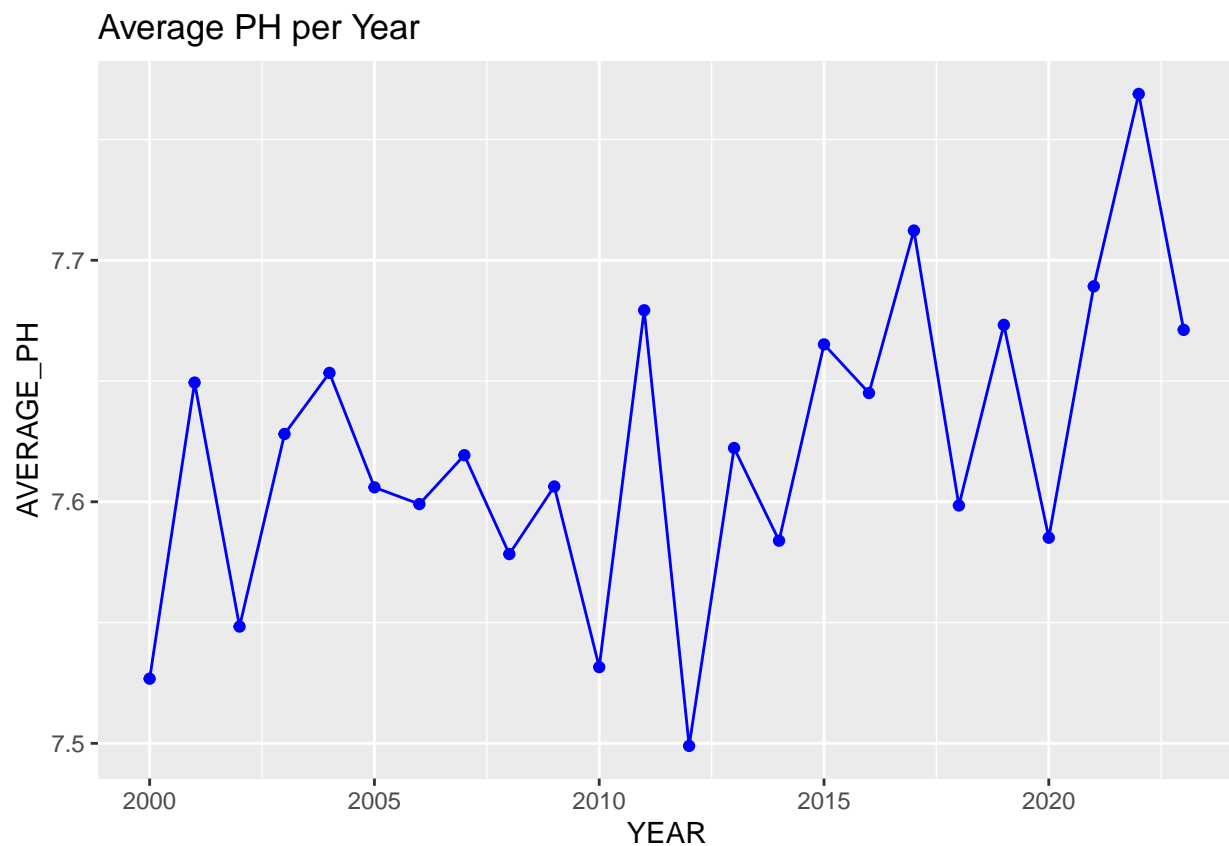
```
library(openxlsx)
library(ggplot2)

excel_file <- "~/Data-Science-G7/PH.xlsx"

#install.packages("readxl")
library(readxl)
sheet_names = excel_sheets(excel_file)

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

plot <- ggplot(phData, aes(x = YEAR, y = AVERAGE_PH)) +
  geom_line(color = "blue") +
  geom_point(color = "blue") +
  labs(x = "YEAR", y = "AVERAGE_PH",
       title = "Average PH per Year")
plot
```



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

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

sheet_names2 = excel_sheets(excel_file2)
#sheet_names
```

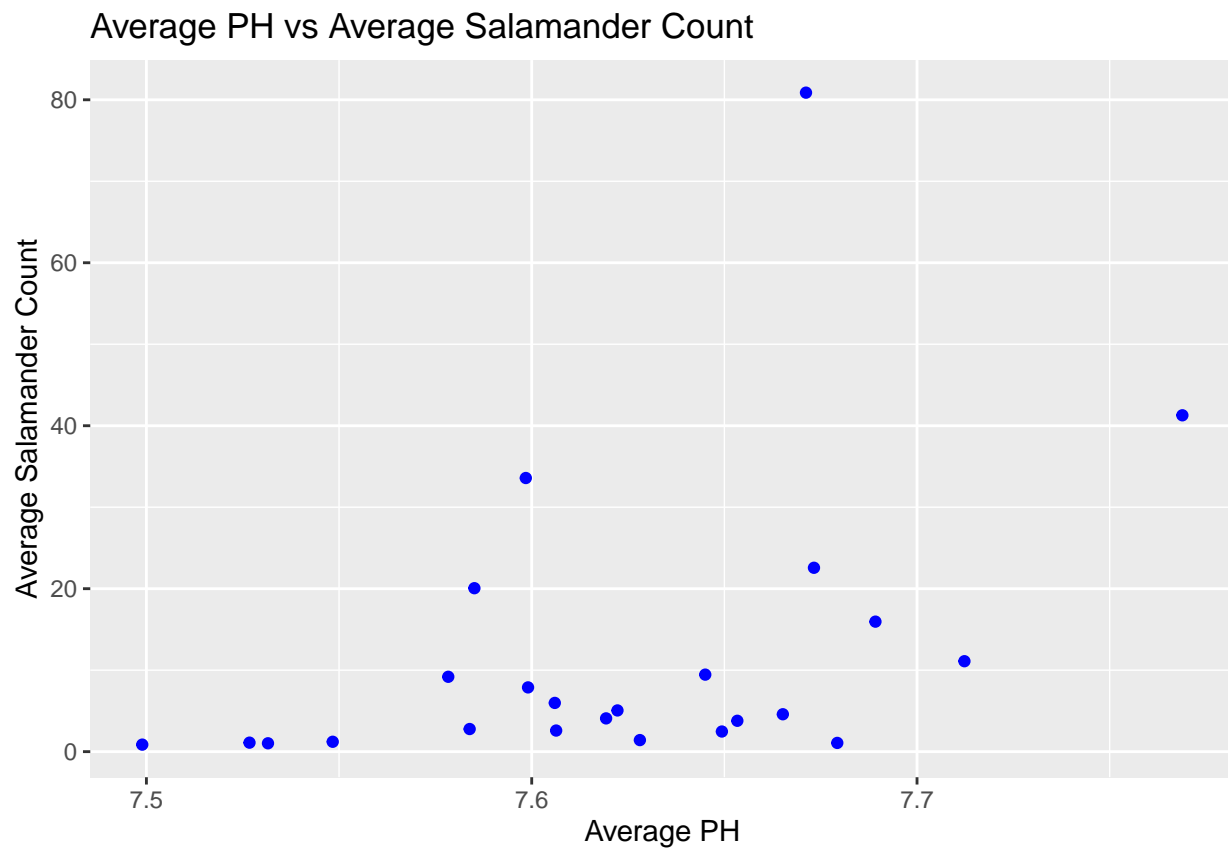
```

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

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

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

```



```

model <- lm(average_salamander ~ AVERAGE_PH, data = merged_data)

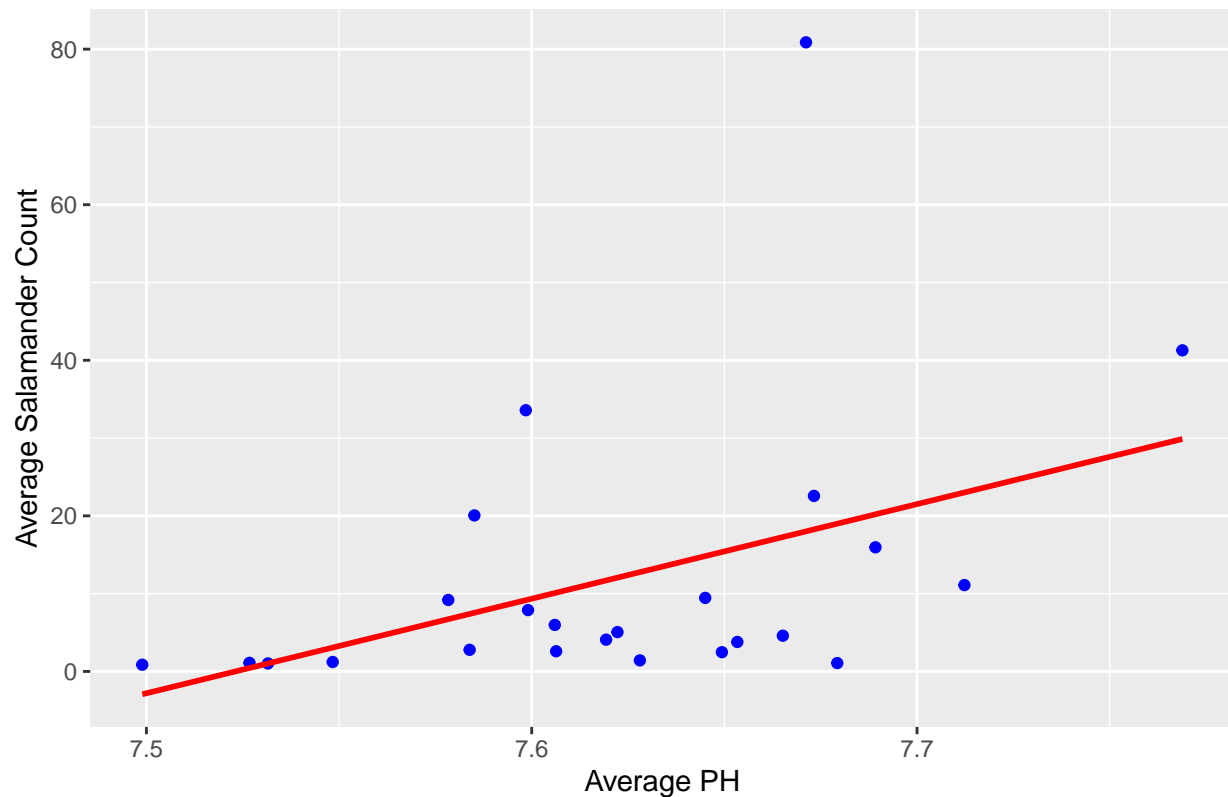
# Add regression line to the plot
plot <- plot + geom_smooth(method = "lm", se = FALSE, color = "red")

# Display the plot
print(plot)

## `geom_smooth()` using formula = 'y ~ x'

```

Average PH vs Average Salamander Count



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

```
##
## Call:
## lm(formula = average_salamander ~ AVERAGE_PH, data = merged_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.919  -8.545  -4.169   2.806  62.869
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -914.78     424.27  -2.156  0.0423 *
## AVERAGE_PH    121.60     55.66   2.185  0.0399 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.78 on 22 degrees of freedom
## Multiple R-squared:  0.1783, Adjusted R-squared:  0.1409
## F-statistic: 4.773 on 1 and 22 DF,  p-value: 0.03985
```

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

```

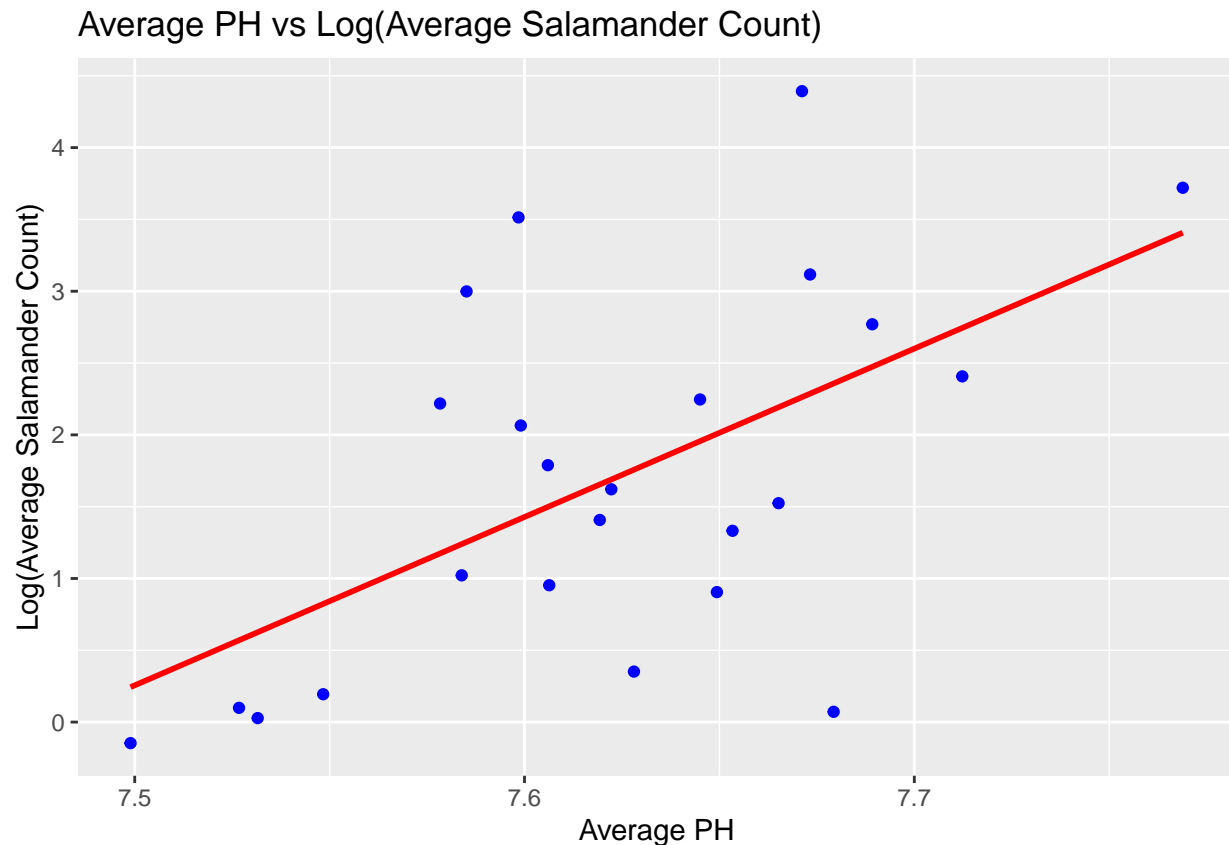
# Exponential regression
model <- lm(log(average_salamander) ~ AVERAGE_PH, data = merged_data)

# Add exponential regression line to the plot
plot <- plot + geom_smooth(method = "lm", se = FALSE, color = "red")

# Display the plot
print(plot)

## `geom_smooth()` using formula = 'y ~ x'

```



```

# Display the summary of the exponential regression model
summary(model)

##
## Call:
## lm(formula = log(average_salamander) ~ AVERAGE_PH, data = merged_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2859 -0.6057 -0.2323  0.3966  2.1303
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -87.658     27.278   -3.213  0.00400 **
## AVERAGE_PH    11.722      3.579    3.276  0.00346 **

```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.079 on 22 degrees of freedom
## Multiple R-squared:  0.3278, Adjusted R-squared:  0.2973
## F-statistic: 10.73 on 1 and 22 DF,  p-value: 0.003456

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

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

##
## Call:
## lm(formula = average_salamander ~ poly(AVERAGE_PH, 2), data = merged_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.545  -7.050  -3.260   0.709  63.570
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      12.084      3.475   3.478  0.00225 **
## poly(AVERAGE_PH, 2)1  36.652     17.022   2.153  0.04307 *
## poly(AVERAGE_PH, 2)2  10.375     17.022   0.609  0.54873
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.02 on 21 degrees of freedom
## Multiple R-squared:  0.1926, Adjusted R-squared:  0.1157
## F-statistic: 2.504 on 2 and 21 DF,  p-value: 0.1058

# Generate a sequence of values for plotting the regression curve
ph_seq <- seq(min(merged_data$AVERAGE_PH), max(merged_data$AVERAGE_PH), length.out = 100)

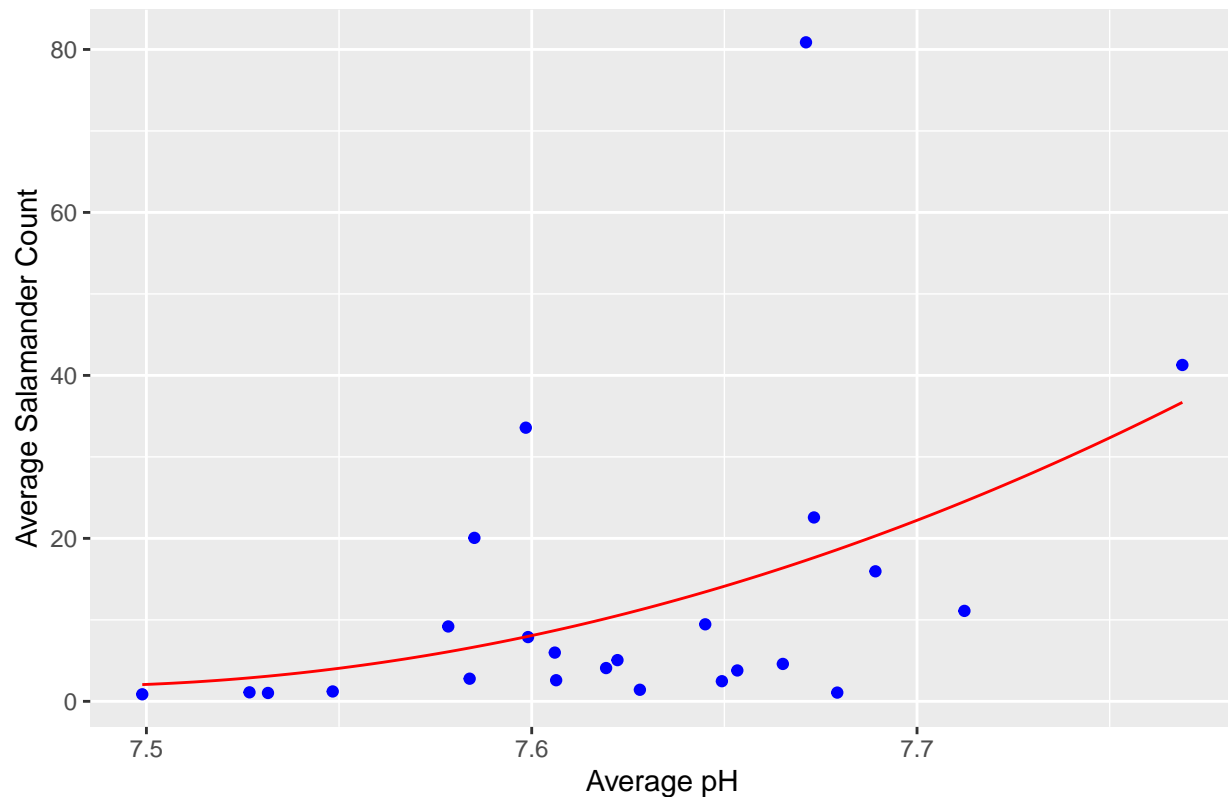
# Predict the average salamander count using the quadratic model
predicted_values <- predict(quadratic_model, newdata = data.frame(AVERAGE_PH = ph_seq))

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

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

```

Quadratic Polynomial Regression: Average Salamander Count per pH



```
library(mgcv)
```

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

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

```
gam_model <- gam(average_salamander ~ s(AVERAGE_PH), data = merged_data)
summary(gam_model)
```

```
##
```

```
## Family: gaussian
```

```
## Link function: identity
```

```
##
```

```
## Formula:
```

```
## average_salamander ~ s(AVERAGE_PH)
```

```
##
```

```
## Parametric coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.084      3.425   3.529 0.00189 **
```

```
## ---
```

```
## 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_PH)  1      1 4.773 0.0399 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## R-sq.(adj) = 0.141   Deviance explained = 17.8%  
## GCV = 307.05   Scale est. = 281.46   n = 24
```

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

