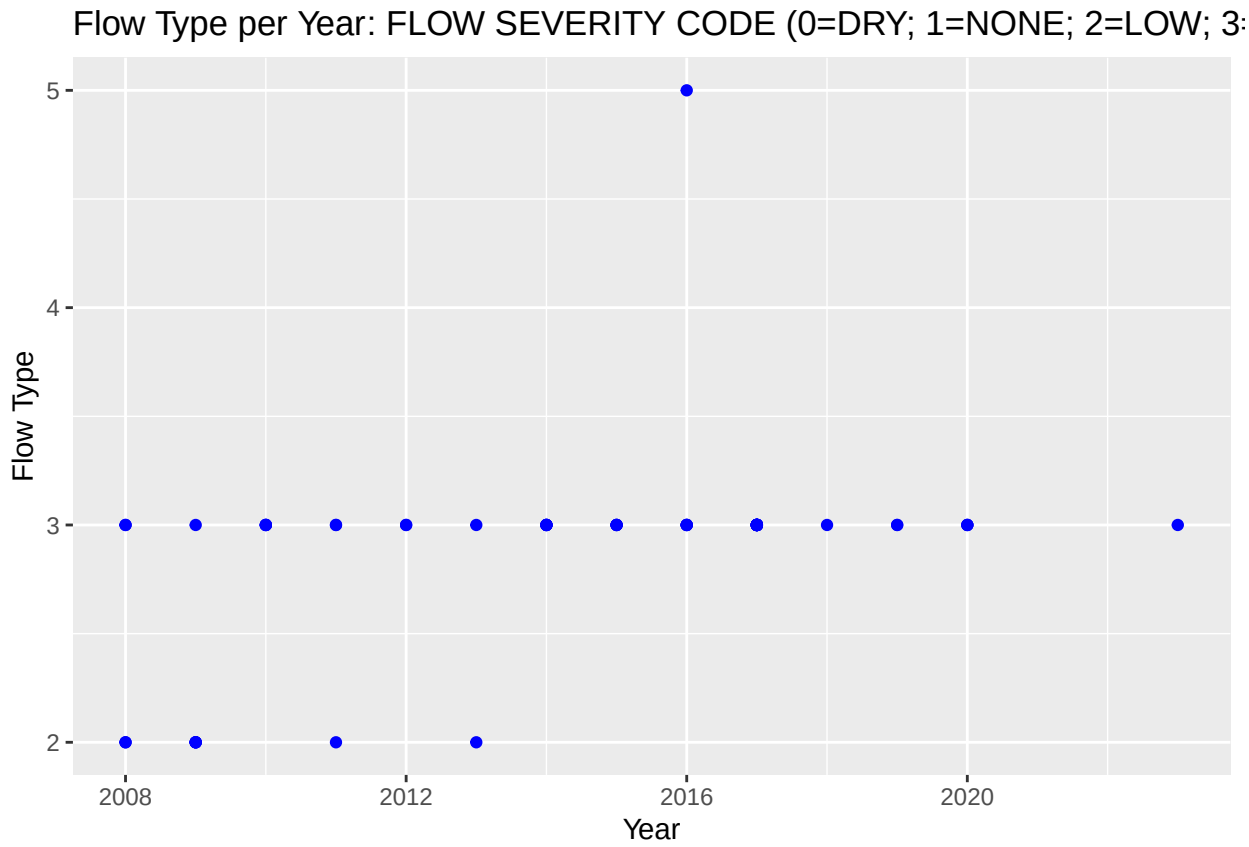
```r
library(openxlsx)
library(readxl)
library(ggplot2)
library(dplyr)
```

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

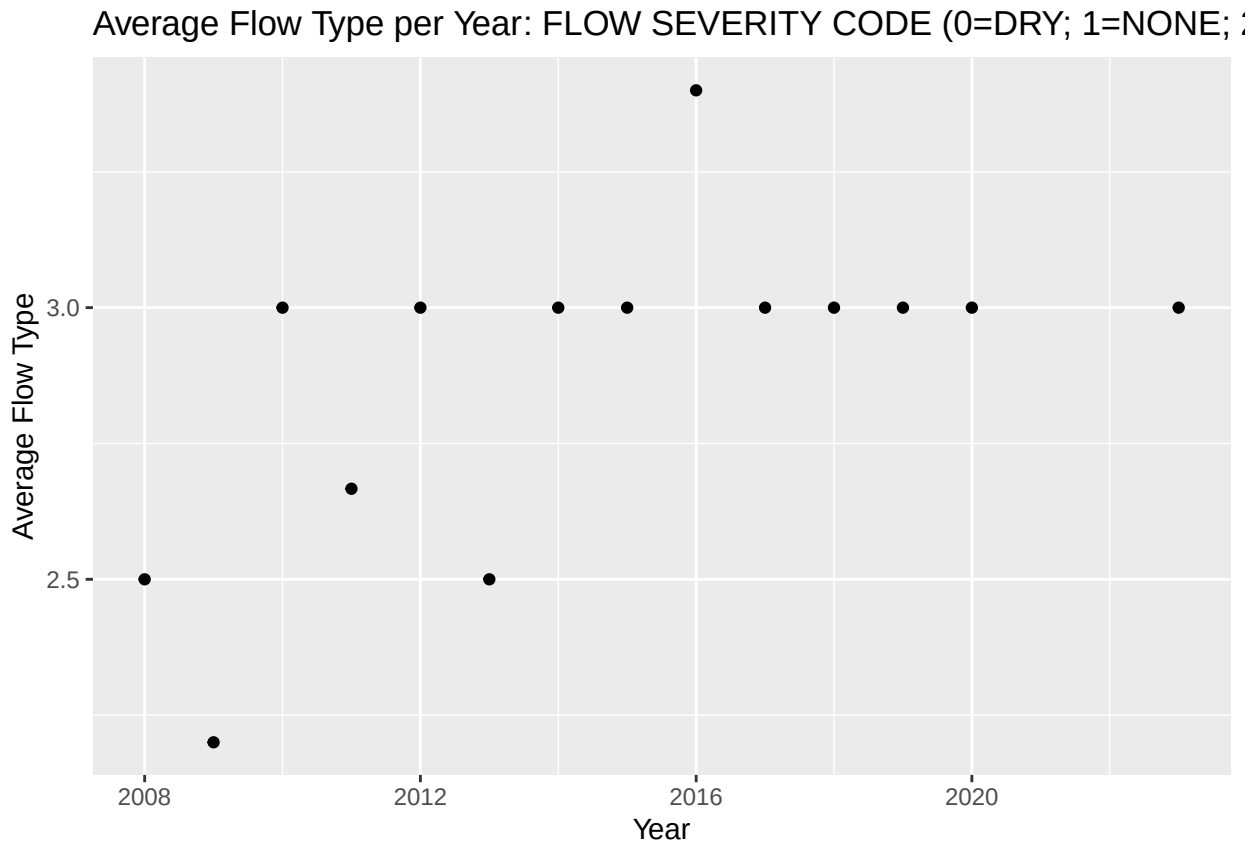
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

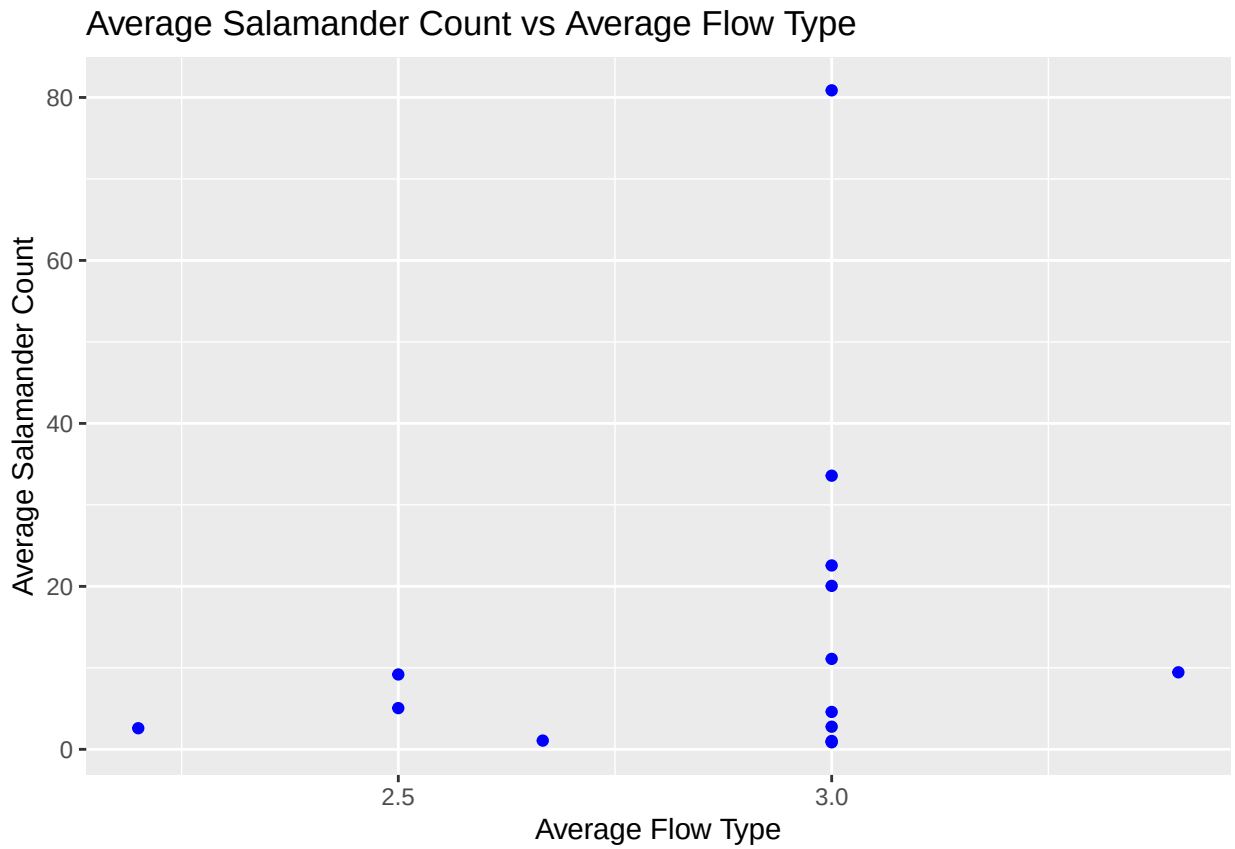
```r
excel_file <- "~/Data-Science-G7/DATA.xlsx"
sheet_names <- excel_sheets(excel_file)

flowData <- read_excel(excel_file, sheet = sheet_names[3])

# Replace values of 5 with 0 in the RESULT column
flowData$RESULT <- ifelse(flowData$RESULT == 6, 0, flowData$RESULT)

# Your plotting code
plot <- ggplot(flowData, aes(x = YEAR, y = RESULT)) +
  geom_point(color = "blue") +
  labs(x = "Year", y = "Flow Type",
       title = "Flow Type per Year: FLOW SEVERITY CODE (0=DRY; 1=NONE; 2=LOW; 3=NORM; 4=FLOOD; 5=HIGH)")

# Customize breaks on the y-axis
plot + scale_y_continuous(breaks = c(0, 1, 2, 3, 4, 5, 6))
```




```
# Calculate average flows per year
average_flows <- flowData %>%
  group_by(YEAR) %>%
  summarise(average_flow = mean(RESULT))

# Your plotting code
plot <- ggplot(average_flows, aes(x = YEAR, y = average_flow)) +
  geom_point(stat = "identity", fill = "blue") +
  labs(x = "Year", y = "Average Flow Type",
       title = "Average Flow Type per Year: FLOW SEVERITY CODE (0=DRY; 1=NONE; 2=LOW; 3=NORM; 4=FLOOD; 5

plot
```




```
salamanderData <- read_excel(excel_file, sheet = sheet_names[4])

averages <- merge(average_flows, salamanderData, by = "YEAR", all.x = TRUE)

plot <- ggplot(averages, aes(x = average_flow, y = Average_salamander)) +
  geom_point(color = "blue") +
  labs(x = "Average Flow Type", y = "Average Salamander Count",
       title = "Average Salamander Count vs Average Flow Type")

plot
```




```
plot <- ggplot(averages, aes(x = average_flow, y = Average_salamander)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", formula = y ~ exp(x), se = FALSE, color = "red") +
  labs(x = "Average Flow Type", y = "Average Salamander Count",
       title = "Average Salamander Count vs Average Flow Type")

# Fit exponential model
exp_model <- lm(log(Average_salamander) ~ average_flow, data = averages)

summary(exp_model)
```

```
##
## Call:
## lm(formula = log(Average_salamander) ~ average_flow, data = averages)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1469 -0.8529  0.0512  0.9502  2.3925
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.529      3.677  -0.416    0.685
## average_flow    1.177      1.272   0.925    0.373
##
## Residual standard error: 1.391 on 12 degrees of freedom
## Multiple R-squared:  0.06658,    Adjusted R-squared:  -0.01121
```

```
## F-statistic: 0.8559 on 1 and 12 DF,  p-value: 0.3731
```

```
plot
```