Dissolved Oxygen Analysis

Group 7

2023-10-04

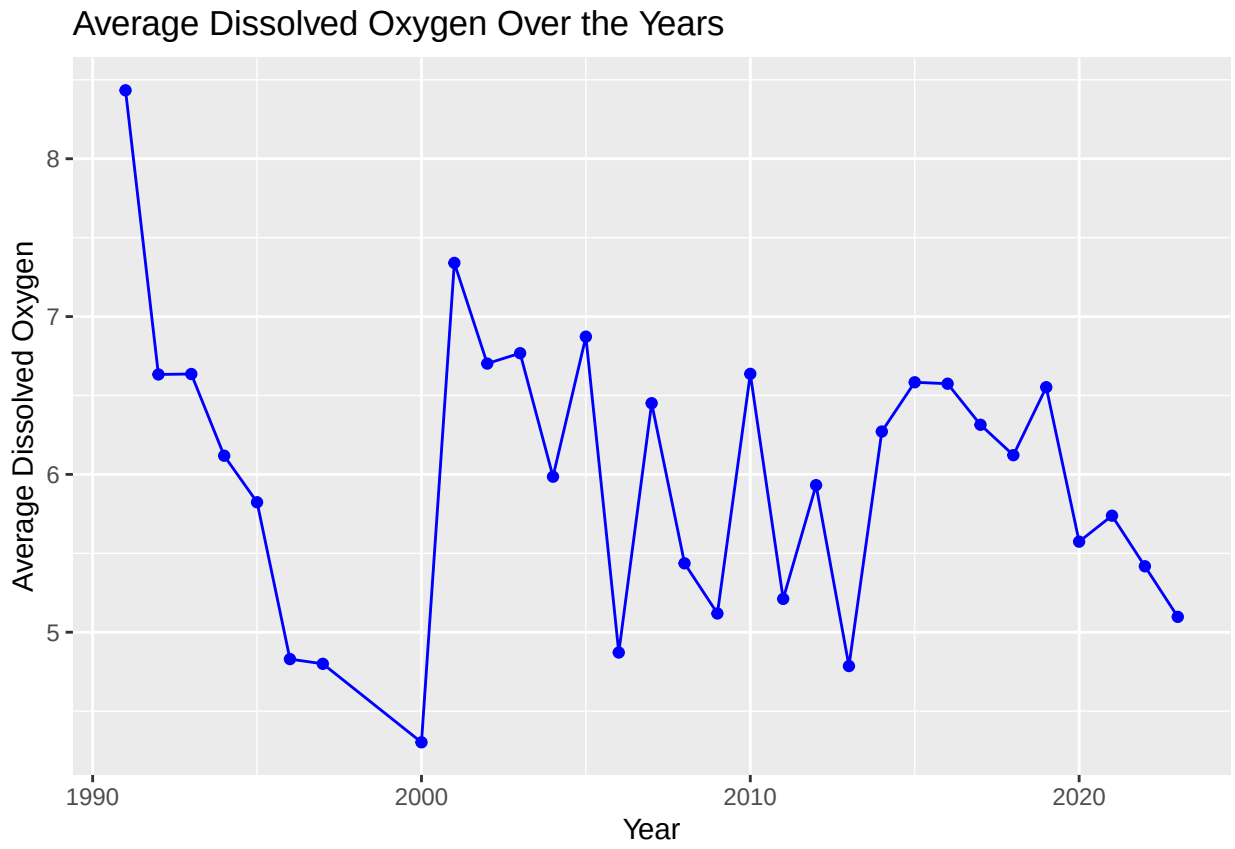
```
library(openxlsx)
library(readxl)
library(ggplot2)
library(dplyr)
```

```
##
## 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
```

```
library(openxlsx)

data2 = read.xlsx("~/Data-Science-G7/Average Dissolved Oxygen Per Year.xlsx")

library(ggplot2)

ggplot(data2, aes(x = YEAR, y = average_dissolved_oxygen)) +
  geom_line(color = "blue") +
  geom_point(color = "blue") +
  labs(x = "Year", y = "Average Dissolved Oxygen",
       title = "Average Dissolved Oxygen Over the Years")
```

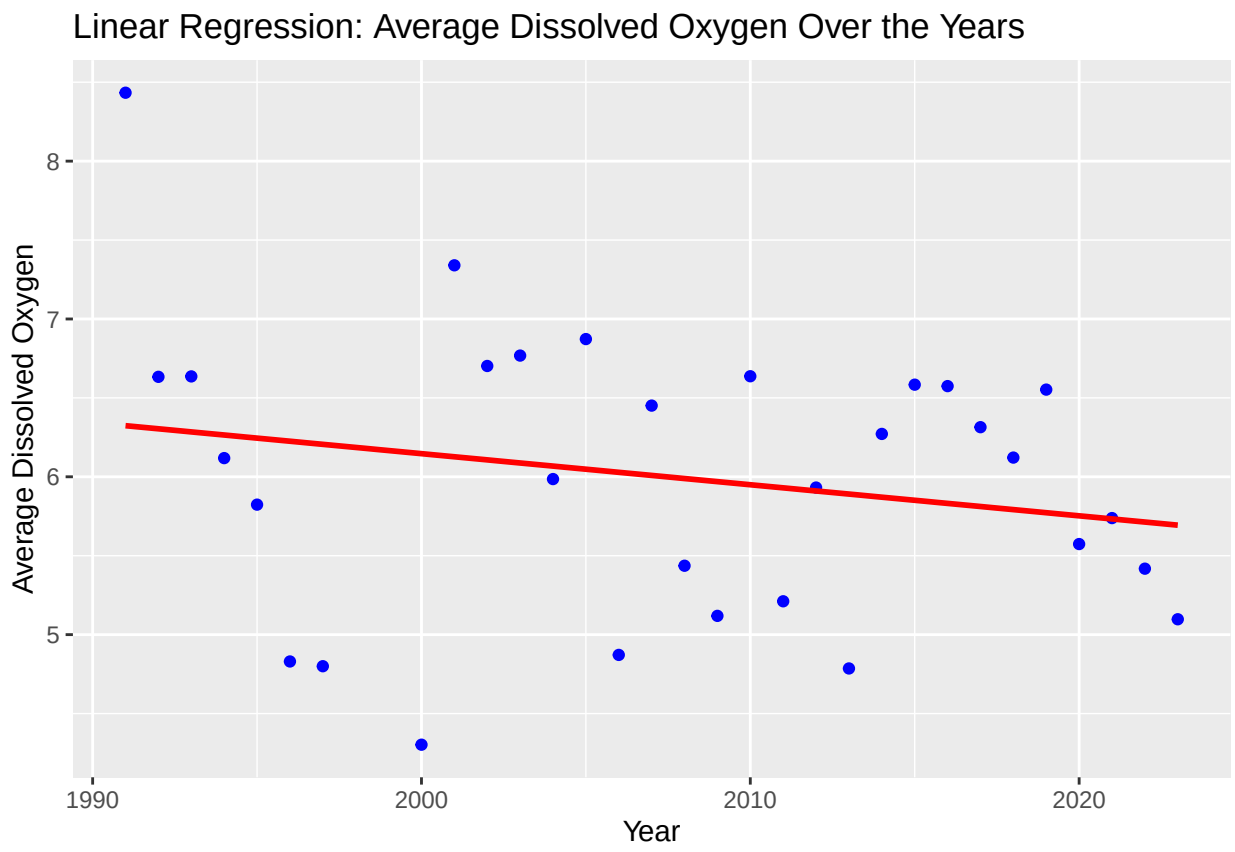



```
# Fit a linear regression model
linear_model <- lm(average_dissolved_oxygen ~ YEAR, data = data2)

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

```
##
## Call:
## lm(formula = average_dissolved_oxygen ~ YEAR, data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.84389 -0.57438  0.02156  0.63774  2.10948
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 45.53245   33.30256   1.367    0.182
## YEAR        -0.01969    0.01659  -1.187    0.245
##
## Residual standard error: 0.8837 on 29 degrees of freedom
## Multiple R-squared:  0.04634,    Adjusted R-squared:  0.01346
## F-statistic: 1.409 on 1 and 29 DF,  p-value: 0.2448
```

```
# Visualize the linear regression line along with the data points
ggplot(data2, aes(x = YEAR, y = average_dissolved_oxygen)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", se = FALSE, color = "red", formula = y ~ x) +
```

```
  labs(x = "Year", y = "Average Dissolved Oxygen",
       title = "Linear Regression: Average Dissolved Oxygen Over the Years")
```

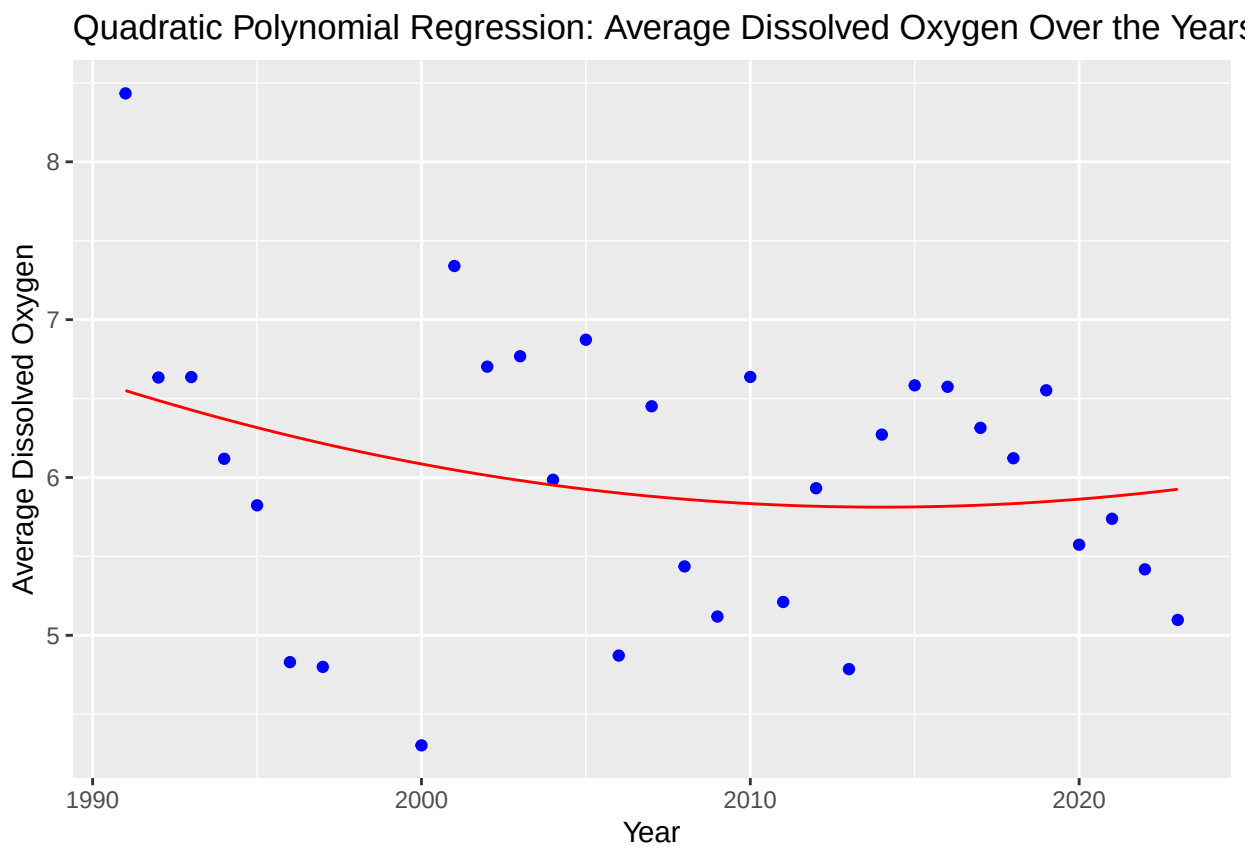



```
# Fit a quadratic polynomial regression model
quadratic_model <- lm(average_dissolved_oxygen ~ poly(YEAR, 2), data = data2)

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

```
##
## Call:
## lm(formula = average_dissolved_oxygen ~ poly(YEAR, 2), data = data2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7829 -0.5528  0.1142  0.6972  1.8828
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.9980     0.1600  37.484   <2e-16 ***
## poly(YEAR, 2)1  -1.0490     0.8909  -1.177    0.249
## poly(YEAR, 2)2   0.6486     0.8909   0.728    0.473
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8909 on 28 degrees of freedom
```

```
## Multiple R-squared:  0.06406,    Adjusted R-squared:  -0.002794
## F-statistic: 0.9582 on 2 and 28 DF,  p-value: 0.3958
```

```r
# Generate a sequence of values for plotting the regression curve
year_seq <- seq(min(data2$YEAR), max(data2$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 = data2, aes(x = YEAR, y = average_dissolved_oxygen), color = "blue") +
  geom_line(data = plot_data, aes(x = YEAR, y = predicted_values), color = "red") +
  labs(x = "Year", y = "Average Dissolved Oxygen",
       title = "Quadratic Polynomial Regression: Average Dissolved Oxygen Over the Years")
```

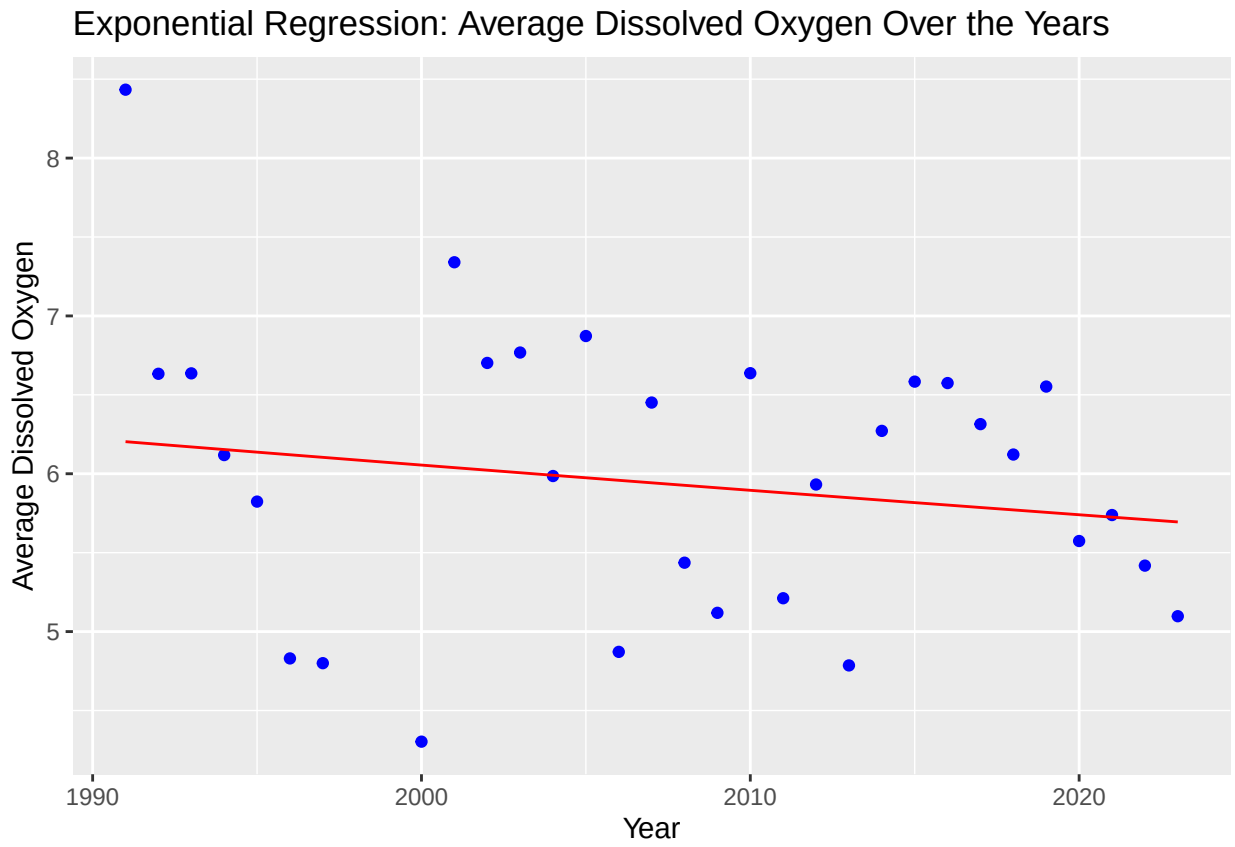



```r
#   exponential regression model
exp_model <- lm(log(average_dissolved_oxygen) ~ log(YEAR), data = data2)

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

```
##
## Call:
```

```
## lm(formula = log(average_dissolved_oxygen) ~ log(YEAR), data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34164 -0.09857  0.01154  0.11272  0.30710
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   42.577     42.747   0.996    0.327
## log(YEAR)     -5.365      5.621  -0.954    0.348
##
## Residual standard error: 0.1492 on 29 degrees of freedom
## Multiple R-squared:  0.03045,    Adjusted R-squared:  -0.002983
## F-statistic: 0.9108 on 1 and 29 DF,  p-value: 0.3478
```

```r
# Generate predicted values using the exponential model
predicted_values <- exp(predict(exp_model, newdata = data2))

# exponential regression curve along with the data points
ggplot(data2, aes(x = YEAR, y = average_dissolved_oxygen)) +
  geom_point(color = "blue") +
  geom_line(aes(x = YEAR, y = predicted_values), color = "red") +
  labs(x = "Year", y = "Average Dissolved Oxygen",
       title = "Exponential Regression: Average Dissolved Oxygen Over the Years")
```



```r
library(mgcv)
```

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

```
##
## Attaching package: 'nlme'
```

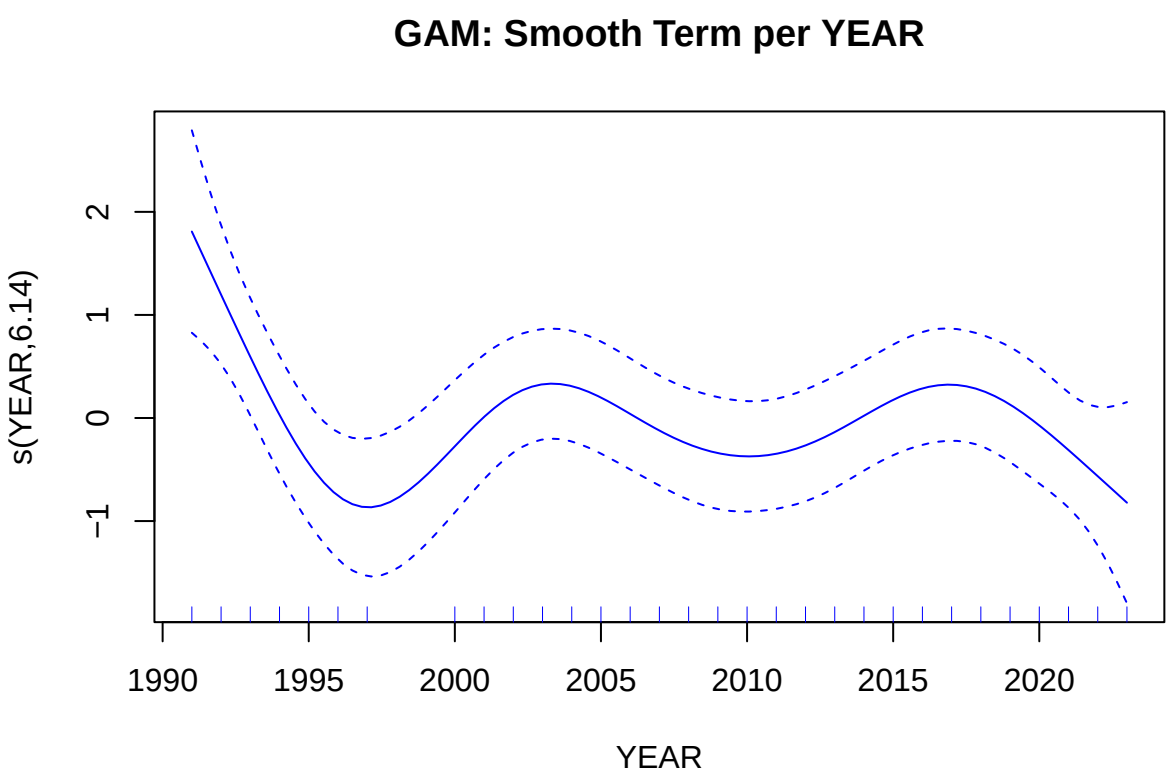
```
## The following object is masked from 'package:dplyr':
##
##     collapse
```

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

```r
gam_model <- gam(average_dissolved_oxygen ~ s(YEAR), data = data2)
summary(gam_model)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## average_dissolved_oxygen ~ s(YEAR)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.9980     0.1219   49.21   <2e-16 ***
## ---
## 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) 6.142  7.286 3.178  0.0152 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.418   Deviance explained = 53.7%
## GCV = 0.5983  Scale est. = 0.46046   n = 31
```

```r
plot(gam_model, se = TRUE, col = "blue", main = "GAM: Smooth Term per YEAR")
```


GAM: Smooth Term per YEAR
s(YEAR,6.14)
2
1
0
−1
1990
1995
2000
2005
2010
2015
2020
YEAR