Assignment 6 - Non Linear Regression

Ben Abraham Biju

2025-03-16

Loading all libraries

```
library(ggplot2)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(splines)
library(mgcv)
```

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

```
## This is mgcv 1.9-1. For overview type 'help("mgcv-package")'.
```

```
library(tidyr)
library(dplyr)
```

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

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

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

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

```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.4.3
```

```
##
## Attaching package: 'gridExtra'
```

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

The following dataset has been used to perform Non Linear Regression: Poverty data (https://online.stat.psu.edu/stat501/lesson/1 (https://online.stat.psu.edu/stat501/lesson/1))

The dataset is saved as a `.txt` file and loaded as a Dataframe

```
file_path="C:/Users/benab/OneDrive - iitkgp.ac.in/Desktop/Sem 6/SL Lab/Lab 7/STAT501_Lesson01/STAT501_Lesson01/poverty.txt"
poverty_df  <- read.table(file_path, header = TRUE, sep = "\t")
```

Displaying the first few rows of the dataset

```
head(poverty_df )
```

```
##     Location PovPct Brth15to17 Brth18to19 ViolCrime TeenBrth
## 1    Alabama   20.1       31.5       88.7      11.2     54.5
## 2     Alaska    7.1       18.9       73.7       9.1     39.5
## 3    Arizona   16.1       35.0      102.5      10.4     61.2
## 4   Arkansas   14.9       31.6      101.7      10.4     59.9
## 5 California   16.7       22.6       69.1      11.2     41.1
## 6   Colorado    8.8       26.2       79.1       5.8     47.0
```

```
str(poverty_df)
```

```
## 'data.frame':    51 obs. of  6 variables:
##  $ Location  : chr  "Alabama" "Alaska" "Arizona" "Arkansas" ...
##  $ PovPct    : num  20.1 7.1 16.1 14.9 16.7 8.8 9.7 10.3 22 16.2 ...
##  $ Brth15to17: num  31.5 18.9 35 31.6 22.6 26.2 14.1 24.7 44.8 23.2 ...
##  $ Brth18to19: num  88.7 73.7 102.5 101.7 69.1 ...
##  $ ViolCrime : num  11.2 9.1 10.4 10.4 11.2 5.8 4.6 3.5 65 7.3 ...
##  $ TeenBrth  : num  54.5 39.5 61.2 59.9 41.1 47 25.8 46.3 69.1 44.5 ...
```

```
summary(poverty_df)
```

```
##    Location             PovPct        Brth15to17      Brth18to19    
##  Length:51          Min.   : 5.30   Min.   : 8.10   Min.   : 39.00  
##  Class :character   1st Qu.:10.25   1st Qu.:17.25   1st Qu.: 58.30  
##  Mode  :character   Median :12.20   Median :20.00   Median : 69.40  
##                     Mean   :13.12   Mean   :22.28   Mean   : 72.02  
##                     3rd Qu.:15.80   3rd Qu.:28.10   3rd Qu.: 87.95  
##                     Max.   :25.30   Max.   :44.80   Max.   :104.30  
##    ViolCrime         TeenBrth    
##  Min.   : 0.900   Min.   :20.00  
##  1st Qu.: 3.900   1st Qu.:33.90  
##  Median : 6.300   Median :39.50  
##  Mean   : 7.855   Mean   :42.24  
##  3rd Qu.: 9.450   3rd Qu.:52.60  
##  Max.   :65.000   Max.   :69.10
```

```
sum(is.na(poverty_df)) # Checking for NaN values in the dataset
```

```
## [1] 0
```

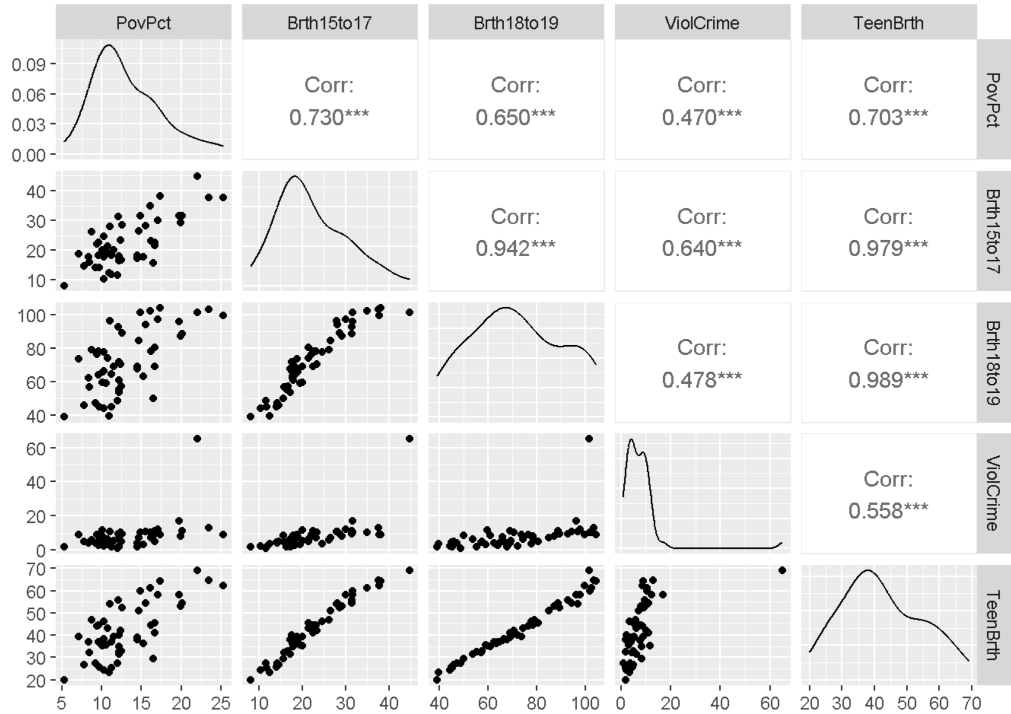
Since `Location` is of character type, it is encoded as a numerical value.

```
poverty_df$Location <- as.factor(poverty_df$Location)
```

Scatter plot

The scatter plot of all predictor variables is shown below

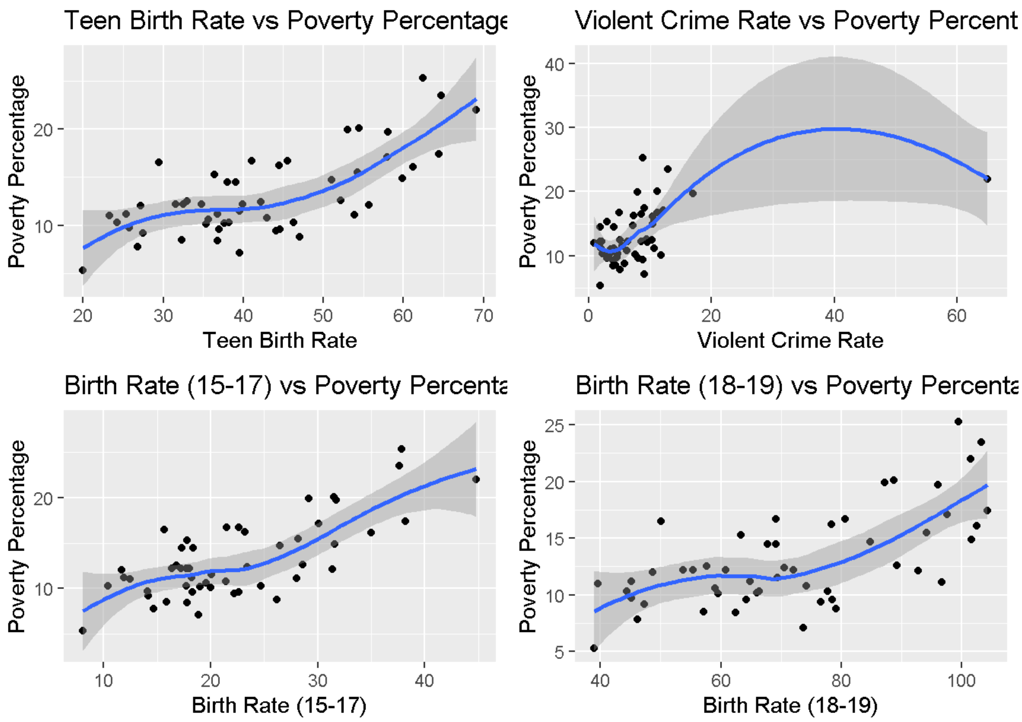
```
ggpairs(poverty_df[, 2:6])
```



The relationship of predictor variables with the response variable, `PovPct` is more clearly visualized with these graphs

```
p1 <- ggplot(poverty_df, aes(x = TeenBrth, y = PovPct)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Teen Birth Rate vs Poverty Percentage",
       x = "Teen Birth Rate", y = "Poverty Percentage")

p2 <- ggplot(poverty_df, aes(x = ViolCrime, y = PovPct)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Violent Crime Rate vs Poverty Percentage",
       x = "Violent Crime Rate", y = "Poverty Percentage")

p3 <- ggplot(poverty_df, aes(x = Brth15to17, y = PovPct)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Birth Rate (15-17) vs Poverty Percentage",
       x = "Birth Rate (15-17)", y = "Poverty Percentage")

p4 <- ggplot(poverty_df, aes(x = Brth18to19, y = PovPct)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Birth Rate (18-19) vs Poverty Percentage",
       x = "Birth Rate (18-19)", y = "Poverty Percentage")

grid.arrange(p1, p2, p3, p4, ncol = 2)
```

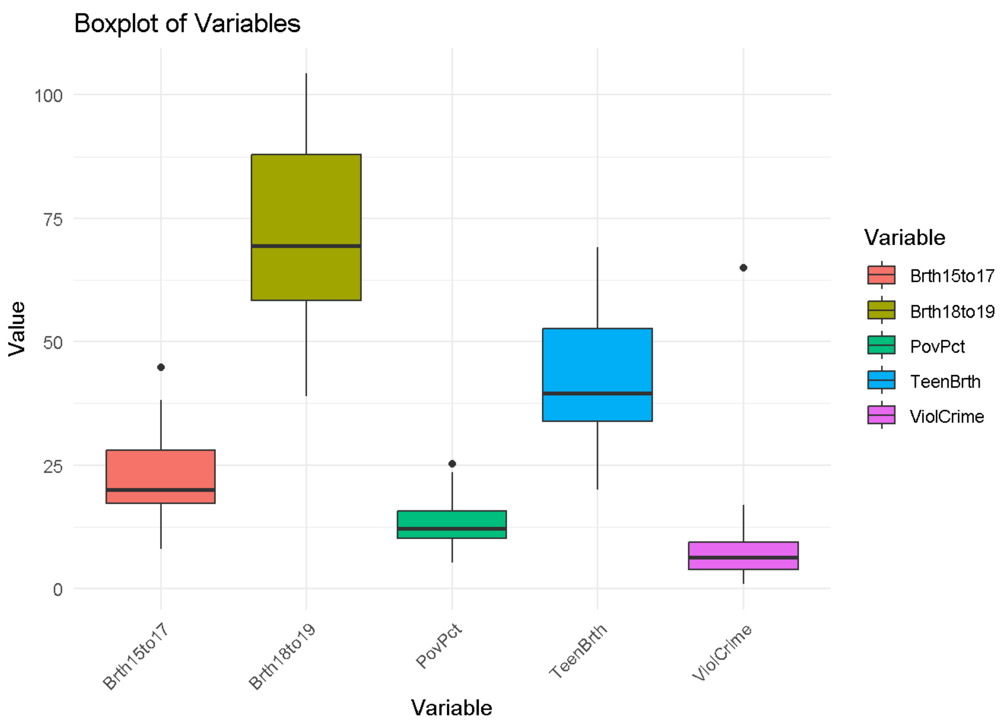
```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```



Box Plots

The box plots of all variables is given below.

```
boxplot_data <- poverty_df %>%
  pivot_longer(cols = -Location, names_to = "Variable", values_to = "Value")

ggplot(boxplot_data, aes(x = Variable, y = Value, fill = Variable)) +
  geom_boxplot() +
  theme_minimal() +
  labs(x = "Variable", y = "Value", title = "Boxplot of Variables") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Fitting a linear model

```
lm_model <- lm(PovPct ~ TeenBrth + ViolCrime + Brth15to17 + Brth18to19, data = poverty_df)
summary(lm_model)
```

```
##
## Call:
## lm(formula = PovPct ~ TeenBrth + ViolCrime + Brth15to17 + Brth18to19,
##     data = poverty_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.5239 -1.9763 -0.1048  1.6729  5.6012
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.22349    1.82549   3.409  0.00136 **
## TeenBrth     1.81957    0.66635   2.731  0.00893 **
## ViolCrime   -0.07786    0.06683  -1.165  0.24997
## Brth15to17  -0.45769    0.44681  -1.024  0.31102
## Brth18to19  -0.82144    0.27311  -3.008  0.00426 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.773 on 46 degrees of freedom
## Multiple R-squared:  0.6132, Adjusted R-squared:  0.5796
## F-statistic: 18.23 on 4 and 46 DF,  p-value: 4.916e-09
```

The R squared value of the linear model shows a decent fit, with the F statistic showing high significance of the predictor variables.

Fitting a non-linear model

A pure polynomial function of degree 2 is fitted on the dataset.

```
poly_model <- lm(PovPct ~ poly(Brth15to17, 2) + poly(Brth18to19, 2) +
                   poly(ViolCrime, 2) + poly(TeenBrth, 2), data = poverty_df)
summary(poly_model)
```

```
##
## Call:
## lm(formula = PovPct ~ poly(Brth15to17, 2) + poly(Brth18to19,
##     2) + poly(ViolCrime, 2) + poly(TeenBrth, 2), data = poverty_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.458 -1.797  0.063  1.322  5.407
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            13.1176     0.3832  34.233  < 2e-16 ***
## poly(Brth15to17, 2)1  -47.8460    28.9210  -1.654  0.10551
## poly(Brth15to17, 2)2   24.1395    20.5479   1.175  0.24670
## poly(Brth18to19, 2)1 -137.9250    45.9011  -3.005  0.00447 **
## poly(Brth18to19, 2)2    8.4431    23.0546   0.366  0.71604
## poly(ViolCrime, 2)1    -6.3508     5.0363  -1.261  0.21426
## poly(ViolCrime, 2)2   -11.1218     5.4067  -2.057  0.04593 *
## poly(TeenBrth, 2)1    204.6820    67.7293   3.022  0.00426 **
## poly(TeenBrth, 2)2    -24.3704    39.9370  -0.610  0.54500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.737 on 42 degrees of freedom
## Multiple R-squared:  0.6562, Adjusted R-squared:  0.5907
## F-statistic: 10.02 on 8 and 42 DF,  p-value: 1.127e-07
```

```
anova(poly_model)
```

```
## Analysis of Variance Table
##
## Response: PovPct
##                     Df Sum Sq Mean Sq F value    Pr(>F)
## poly(Brth15to17, 2)  2 505.35 252.675 33.7416 1.828e-09 ***
## poly(Brth18to19, 2)  2   1.79   0.895  0.1195   0.88766
## poly(ViolCrime, 2)   2  23.15  11.574  1.5455   0.22508
## poly(TeenBrth, 2)    2  69.93  34.965  4.6691   0.01476 *
## Residuals           42 314.52   7.489
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA results of the degree 2 polynomial suggests that `TeenBrth` and `Brth15to17` are significant.

Selecting the best fit polynomial function

Each predictor is modeled with varying polynomial degrees ranging from 2 to 5. The performance of each model is evaluated using AIC, BIC and R^2 values

```
degree_range <- 2:5

degree_combinations <- expand.grid(
  Brth15to17 = degree_range,
  Brth18to19 = degree_range,
  ViolCrime = degree_range,
  TeenBrth = degree_range
)
# Store AIC values and R² values
aic_values <- numeric(length(degree_range))
bic_values <- numeric(length(degree_range))
r2_values <- numeric(length(degree_range))

# Loop through each degree
for (i in 1:nrow(degree_combinations))  {

  d1 <- degree_combinations$Brth15to17[i]
d2 <- degree_combinations$Brth18to19[i]
d3 <- degree_combinations$ViolCrime[i]
d4 <- degree_combinations$TeenBrth[i]

# Fit polynomial regression model with different degrees for each predictor
model <- lm(PovPct ~ poly(Brth15to17, d1) + poly(Brth18to19, d2) +
poly(ViolCrime, d3) + poly(TeenBrth, d4), data = poverty_df)



  aic_values[i] <- AIC(model)
  bic_values[i] <- BIC(model)
  r2_values[i] <- summary(model)$r.squared
}
```

```
# Find the best degree based on minimum AIC
best_aic_index <- which.min(aic_values)
best_aic_combination <- degree_combinations[best_aic_index, ]


best_bic_index <- which.min(bic_values)
best_bic_combination <- degree_combinations[best_bic_index, ]

# Find the best degree based on maximum R²
best_r2_index <- which.max(r2_values)
best_r2_combination <- degree_combinations[best_r2_index, ]

# Print results
cat("Best polynomial based on AIC:", "\n")
```

```
## Best polynomial based on AIC:
```

```
print(best_aic_combination)
```

```
##    Brth15to17 Brth18to19 ViolCrime TeenBrth
## 73          2          4         2        3
```

```
cat("Best polynomial based on BIC:", "\n")
```

```
## Best polynomial based on BIC:
```

```
print(best_bic_combination)
```

```
##   Brth15to17 Brth18to19 ViolCrime TeenBrth
## 1          2          2         2        2
```

```
cat("Best polynomial based on R²:", "\n")
```

```
## Best polynomial based on R²:
```

```
print(best_r2_combination)
```

```
##     Brth15to17 Brth18to19 ViolCrime TeenBrth
## 256          5          5         5        5
```

- AIC-optimal model: Used polynomial degrees of 2 for `Brth15to17` and `ViolCrime` , 4 for `Brth18to19` , and 3 for `TeenBrth`. This model achieved an adjusted R^2 of 0.6197, improving upon the linear model.
- R^2-optimal model: Used 5th-degree polynomials for all predictors, which likely overfit the data but maximized the R^2 value.
- BIC-optimal model: Used 2nd-degree polynomials for all predictors, providing a more parsimonious model than the AIC-optimal one.

```
poverty_df <- poverty_df %>%
  arrange(Location)

# Fit the best model using AIC-optimal polynomial degrees
best_aic_model <- lm(PovPct ~ poly(Brth15to17, best_aic_combination$Brth15to17) +
                       poly(Brth18to19, best_aic_combination$Brth18to19) +
                       poly(ViolCrime, best_aic_combination$ViolCrime) +
                       poly(TeenBrth, best_aic_combination$TeenBrth),
                     data = poverty_df)
summary(best_aic_model)
```

```
##
## Call:
## lm(formula = PovPct ~ poly(Brth15to17, best_aic_combination$Brth15to17) +
##     poly(Brth18to19, best_aic_combination$Brth18to19) + poly(ViolCrime,
##     best_aic_combination$ViolCrime) + poly(TeenBrth, best_aic_combination$TeenBrth),
##     data = poverty_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6986 -1.7077  0.1834  1.0313  4.9252
##
## Coefficients:
##                                                      Estimate Std. Error t value
## (Intercept)                                          13.1176     0.3693  35.517
## poly(Brth15to17, best_aic_combination$Brth15to17)1  -53.0747    28.3269  -1.874
## poly(Brth15to17, best_aic_combination$Brth15to17)2   21.5255    20.8228   1.034
## poly(Brth18to19, best_aic_combination$Brth18to19)1 -116.7655    47.1544  -2.476
## poly(Brth18to19, best_aic_combination$Brth18to19)2    7.0348    24.0661   0.292
## poly(Brth18to19, best_aic_combination$Brth18to19)3  -14.4677     7.5843  -1.908
## poly(Brth18to19, best_aic_combination$Brth18to19)4   -5.5821     3.0249  -1.845
## poly(ViolCrime, best_aic_combination$ViolCrime)1    -16.1925     6.3906  -2.534
## poly(ViolCrime, best_aic_combination$ViolCrime)2    -14.7958     5.6335  -2.626
## poly(TeenBrth, best_aic_combination$TeenBrth)1      192.8275    67.4961   2.857
## poly(TeenBrth, best_aic_combination$TeenBrth)2      -12.5646    40.9520  -0.307
## poly(TeenBrth, best_aic_combination$TeenBrth)3       20.0158    10.0523   1.991
##                                                     Pr(>|t|)
## (Intercept)                                          < 2e-16 ***
## poly(Brth15to17, best_aic_combination$Brth15to17)1  0.06849 .
## poly(Brth15to17, best_aic_combination$Brth15to17)2  0.30762
## poly(Brth18to19, best_aic_combination$Brth18to19)1  0.01772 *
## poly(Brth18to19, best_aic_combination$Brth18to19)2  0.77160
## poly(Brth18to19, best_aic_combination$Brth18to19)3  0.06383 .
## poly(Brth18to19, best_aic_combination$Brth18to19)4  0.07259 .
## poly(ViolCrime, best_aic_combination$ViolCrime)1    0.01541 *
## poly(ViolCrime, best_aic_combination$ViolCrime)2    0.01227 *
## poly(TeenBrth, best_aic_combination$TeenBrth)1      0.00683 **
## poly(TeenBrth, best_aic_combination$TeenBrth)2      0.76062
## poly(TeenBrth, best_aic_combination$TeenBrth)3      0.05350 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.638 on 39 degrees of freedom
## Multiple R-squared:  0.7034, Adjusted R-squared:  0.6197
## F-statistic: 8.408 on 11 and 39 DF,  p-value: 2.523e-07
```

```
anova(best_aic_model)
```

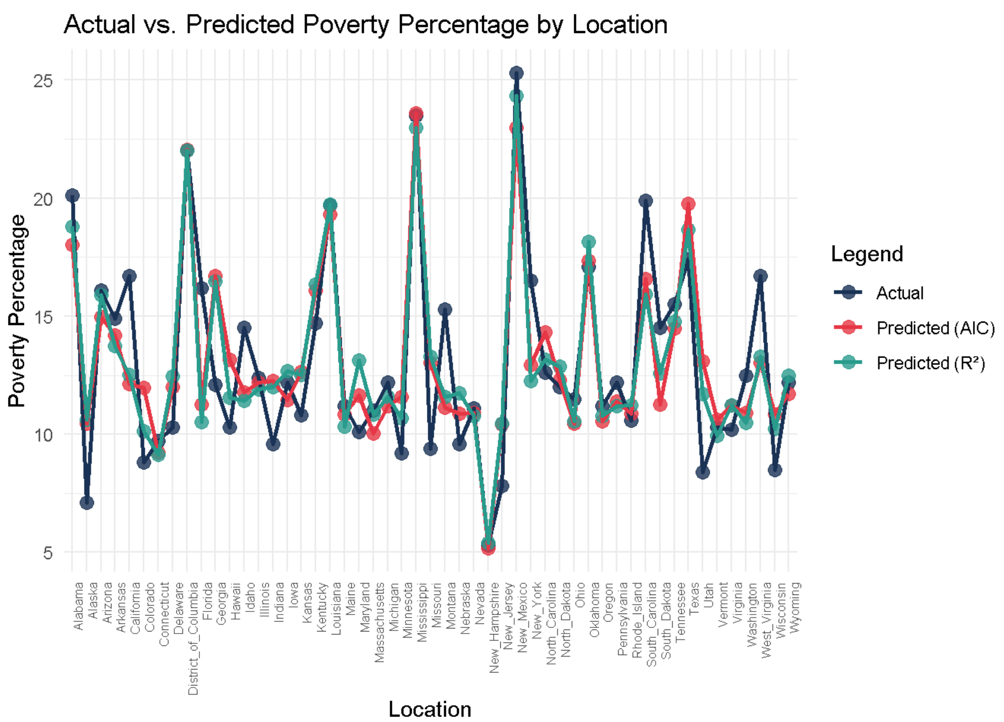
```
## Analysis of Variance Table
##
## Response: PovPct
##                                                   Df Sum Sq Mean Sq F value
## poly(Brth15to17, best_aic_combination$Brth15to17)  2 505.35 252.675 36.3209
## poly(Brth18to19, best_aic_combination$Brth18to19)  4  18.72   4.681  0.6729
## poly(ViolCrime, best_aic_combination$ViolCrime)    2  32.01  16.005  2.3006
## poly(TeenBrth, best_aic_combination$TeenBrth)      3  87.34  29.113  4.1848
## Residuals                                         39 271.31   6.957
##                                                      Pr(>F)
## poly(Brth15to17, best_aic_combination$Brth15to17) 1.239e-09 ***
## poly(Brth18to19, best_aic_combination$Brth18to19)   0.61478
## poly(ViolCrime, best_aic_combination$ViolCrime)     0.11364
## poly(TeenBrth, best_aic_combination$TeenBrth)       0.01162 *
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The AIC-optimal model showed that several polynomial terms were significant, particularly for `Brth15to17` and `TeenBrth` , indicating that the relationship between these predictors and poverty is indeed non-linear.

```
# Predict values using the best AIC model
poverty_df$Predicted_AIC <- predict(best_aic_model, newdata = poverty_df)



# Fit the best model using R²-optimal polynomial degrees
best_r2_model <- lm(PovPct ~ poly(Brth15to17, best_r2_combination$Brth15to17) +
                      poly(Brth18to19, best_r2_combination$Brth18to19) +
                      poly(ViolCrime, best_r2_combination$ViolCrime) +
                      poly(TeenBrth, best_r2_combination$TeenBrth),
                    data = poverty_df)

# Predict values using the best R² model
poverty_df$Predicted_R2 <- predict(best_r2_model, newdata = poverty_df)
```

```
# Plot the predicted values of best AIC and best R² model
plot_data <- poverty_df %>%
  select(Location, PovPct, Predicted_AIC, Predicted_R2) %>%
  pivot_longer(cols = c(PovPct, Predicted_AIC, Predicted_R2), names_to = "Type", values_to = "Value")

options(repr.plot.width=15, repr.plot.height=7)

ggplot(plot_data, aes(x = Location, y = Value, color = Type, group = Type)) +
  geom_point(size = 3, alpha = 0.8) +
  geom_line(size = 1) +
  scale_color_manual(values = c("PovPct" = "#1D3557",
                                "Predicted_AIC" = "#E63946",
                                "Predicted_R2" = "#2A9D8F"),
                     labels = c("Actual", "Predicted (AIC)", "Predicted (R²)")) +
  theme_minimal() +
  labs(title = "Actual vs. Predicted Poverty Percentage by Location",
       x = "Location",
       y = "Poverty Percentage",
       color = "Legend") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Using splines with varying knots

Spline models with 3 and 4 knots at different quantiles of the predictor variables are trained on the dataset.

Spline model with 3 knots

```
# Spline model with 3 knots
spline_model_3k <- lm(PovPct~ bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75))) +
                        bs(Brth18to19 , knots = quantile(poverty_df$Brth18to19 , probs = c(0.25, 0.5, 0.75))) +
                        bs(ViolCrime , knots = quantile(poverty_df$ViolCrime , probs = c(0.25, 0.5, 0.75)))+
                        bs(TeenBrth , knots = quantile(poverty_df$TeenBrth , probs = c(0.25, 0.5, 0.75))),
                      data = poverty_df)

# Summary of spline model
summary(spline_model_3k)
```

```
##
## Call:
## lm(formula = PovPct ~ bs(Brth15to17, knots = quantile(poverty_df$Brth15to17,
##     probs = c(0.25, 0.5, 0.75))) + bs(Brth18to19, knots = quantile(poverty_df$Brth18to19,
##     probs = c(0.25, 0.5, 0.75))) + bs(ViolCrime, knots = quantile(poverty_df$ViolCrime,
##     probs = c(0.25, 0.5, 0.75))) + bs(TeenBrth, knots = quantile(poverty_df$TeenBrth,
##     probs = c(0.25, 0.5, 0.75))), data = poverty_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.596 -1.228  0.001  1.018  5.528
##
## Coefficients:
##                                                                                 Estimate
## (Intercept)                                                                       6.1059
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))1 -45.5596
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))2 -19.6459
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))3 -25.8020
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))4 -44.9981
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))5 -38.6092
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))6 -25.7048
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))1 -29.7829
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))2 -15.9531
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))3 -27.1080
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))4 -37.5053
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))5 -50.2770
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))6 -62.0267
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))1    -0.5479
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))2    -4.6593
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))3    -3.2214
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))4     7.7576
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))5   -16.2502
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))6   -21.2738
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))1      78.5191
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))2      43.5881
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))3      61.4030
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))4      90.3844
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))5     115.7514
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))6     119.3456
##                                                                                 Std. Error
## (Intercept)                                                                         4.9126
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))1    56.2598
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))2    30.8076
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))3    33.6331
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))4    36.1107
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))5    41.9544
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))6    63.5785
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))1    30.6578
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))2    40.4471
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))3    36.8144
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))4    41.5919
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))5    40.1498
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))6    40.5406
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))1       5.8670
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))2       5.4013
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))3       5.8953
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))4      18.0845
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))5     101.9371
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))6      34.5585
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))1        81.5369
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))2        63.6543
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))3        64.5116
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))4        70.7076
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))5        69.2345
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))6        75.5823
##                                                                                 t value
## (Intercept)                                                                       1.243
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))1  -0.810
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))2  -0.638
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))3  -0.767
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))4  -1.246
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))5  -0.920
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))6  -0.404
```

```
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))1  -0.971
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))2  -0.394
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))3  -0.736
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))4  -0.902
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))5  -1.252
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))6  -1.530
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))1    -0.093
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))2    -0.863
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))3    -0.546
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))4     0.429
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))5    -0.159
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))6    -0.616
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))1       0.963
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))2       0.685
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))3       0.952
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))4       1.278
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))5       1.672
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))6       1.579
##                                                                                      Pr(>|t|)
## (Intercept)                                                                             0.225
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))1    0.425
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))2    0.529
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))3    0.450
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))4    0.224
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))5    0.366
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75)))6    0.689
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))1    0.340
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))2    0.696
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))3    0.468
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))4    0.375
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))5    0.222
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75)))6    0.138
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))1      0.926
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))2      0.396
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))3      0.589
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))4      0.671
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))5      0.875
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75)))6      0.544
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))1        0.344
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))2        0.500
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))3        0.350
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))4        0.212
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))5        0.107
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75)))6        0.126
## 
## Residual standard error: 3.049 on 26 degrees of freedom
## Multiple R-squared:  0.7358, Adjusted R-squared:  0.4919 
## F-statistic: 3.017 on 24 and 26 DF,  p-value: 0.003566
```

```
poverty_df$Spline_3k_Predicted <- predict(spline_model_3k)
```

Spline model with 4 knots

```
# Spline model with 3 knots
spline_model_4k <- lm(PovPct~ bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95))) +
                        bs(Brth18to19 , knots = quantile(poverty_df$Brth18to19 , probs = c(0.25, 0.5, 0.75, 0.95))) +
                        bs(ViolCrime , knots = quantile(poverty_df$ViolCrime , probs = c(0.25, 0.5, 0.75, 0.95)))+
                        bs(TeenBrth , knots = quantile(poverty_df$TeenBrth , probs = c(0.25, 0.5, 0.75, 0.95))),
                      data = poverty_df)

# Summary of spline model
summary(spline_model_4k)
```

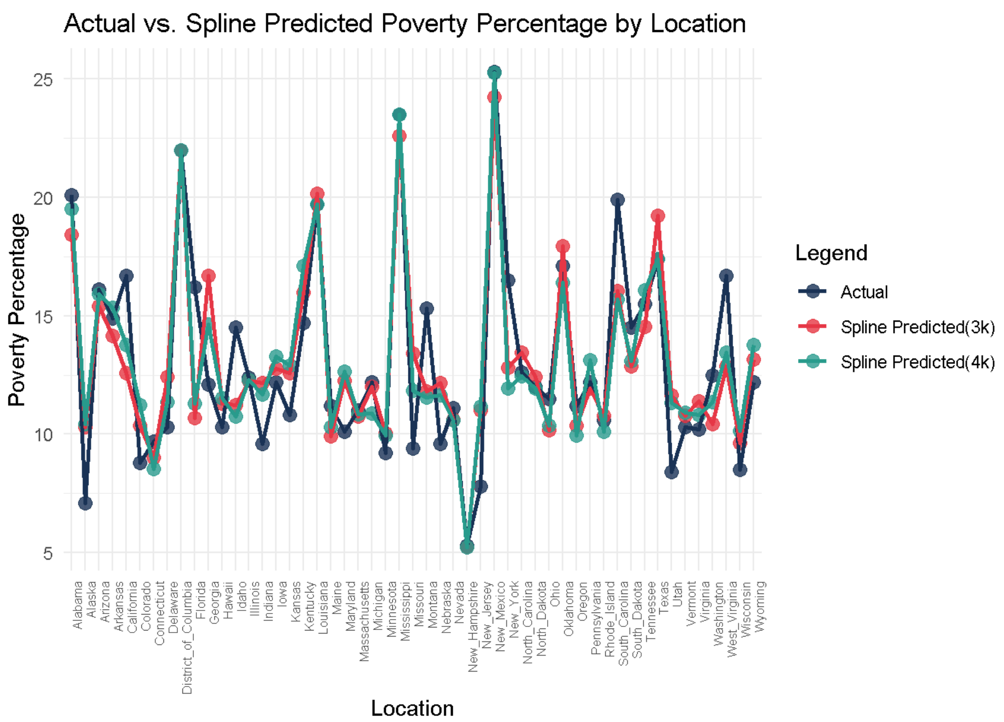
```
##
## Call:
## lm(formula = PovPct ~ bs(Brth15to17, knots = quantile(poverty_df$Brth15to17,
##     probs = c(0.25, 0.5, 0.75, 0.95))) + bs(Brth18to19, knots = quantile(poverty_df$Brth18to19,
##     probs = c(0.25, 0.5, 0.75, 0.95))) + bs(ViolCrime, knots = quantile(poverty_df$ViolCrime,
##     probs = c(0.25, 0.5, 0.75, 0.95))) + bs(TeenBrth, knots = quantile(poverty_df$TeenBrth,
##     probs = c(0.25, 0.5, 0.75, 0.95))), data = poverty_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3557 -1.4046  0.0005  1.0158  4.8988
##
## Coefficients:
##                                                                                       Estimate
## (Intercept)                                                                          4.717e+00
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))1 -4.191e+01
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))2 -2.243e+01
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))3 -3.142e+01
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))4 -3.984e+01
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))5 -6.088e+01
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))6 -2.613e+01
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))7 -2.996e+05
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))1 -3.127e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))2 -2.301e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))3 -3.403e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))4 -3.295e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))5 -5.595e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))6 -6.869e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))7  3.300e+01
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))1    2.086e+00
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))2   -5.048e+00
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))3    2.027e+00
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))4   -5.676e+00
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))5    1.297e+02
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))6   -6.444e+03
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))7    3.002e+05
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))1      7.542e+01
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))2      5.532e+01
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))3      7.083e+01
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))4      8.402e+01
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))5      1.305e+02
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))6      1.412e+02
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))7     -5.329e+02
##                                                                                     Std. Error
## (Intercept)                                                                          5.068e+00
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))1  5.724e+01
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))2  3.106e+01
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))3  3.408e+01
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))4  3.577e+01
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))5  4.707e+01
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))6  2.140e+02
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))7  2.319e+05
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))1  3.097e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))2  4.188e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))3  3.732e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))4  4.341e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))5  4.217e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))6  5.524e+01
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))7  8.237e+01
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))1    6.266e+00
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))2    5.489e+00
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))3    7.102e+00
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))4    6.169e+00
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))5    8.444e+01
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))6    4.917e+03
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))7    2.337e+05
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))1      8.283e+01
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))2      6.511e+01
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))3      6.531e+01
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))4      7.277e+01
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))5      7.973e+01
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))6      2.831e+02
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))7      5.100e+03
```

```
##                                                                                        t value
## (Intercept)                                                                              0.931
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))1  -0.732
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))2  -0.722
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))3  -0.922
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))4  -1.114
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))5  -1.293
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))6  -0.122
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))7  -1.292
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))1  -1.010
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))2  -0.549
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))3  -0.912
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))4  -0.759
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))5  -1.327
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))6  -1.244
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))7   0.401
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))1     0.333
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))2    -0.920
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))3     0.285
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))4    -0.920
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))5     1.535
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))6    -1.310
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))7     1.285
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))1       0.911
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))2       0.850
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))3       1.085
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))4       1.155
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))5       1.637
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))6       0.499
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))7      -0.104
##                                                                                        Pr(>|t|)
## (Intercept)                                                                               0.362
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))1    0.472
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))2    0.478
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))3    0.367
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))4    0.277
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))5    0.209
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))6    0.904
## bs(Brth15to17, knots = quantile(poverty_df$Brth15to17, probs = c(0.25, 0.5, 0.75, 0.95)))7    0.210
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))1    0.324
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))2    0.588
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))3    0.372
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))4    0.456
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))5    0.198
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))6    0.227
## bs(Brth18to19, knots = quantile(poverty_df$Brth18to19, probs = c(0.25, 0.5, 0.75, 0.95)))7    0.693
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))1      0.742
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))2      0.368
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))3      0.778
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))4      0.367
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))5      0.139
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))6      0.204
## bs(ViolCrime, knots = quantile(poverty_df$ViolCrime, probs = c(0.25, 0.5, 0.75, 0.95)))7      0.212
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))1        0.372
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))2        0.405
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))3        0.290
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))4        0.261
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))5        0.116
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))6        0.623
## bs(TeenBrth, knots = quantile(poverty_df$TeenBrth, probs = c(0.25, 0.5, 0.75, 0.95)))7        0.918
##
## Residual standard error: 3.07 on 22 degrees of freedom
## Multiple R-squared:  0.7734, Adjusted R-squared:  0.485
## F-statistic: 2.681 on 28 and 22 DF,  p-value: 0.01021
```

```
poverty_df$Spline_4k_Predicted <- predict(spline_model_4k)
```

Plot of actual vs. predicted values using splines

```
# Arrange data by Location
poverty_df <- poverty_df %>%
  arrange(Location)

plot_data <- poverty_df %>%
  select(Location, PovPct, Spline_3k_Predicted, Spline_4k_Predicted) %>%
  pivot_longer(cols = c(PovPct,Spline_3k_Predicted,Spline_4k_Predicted ),
               names_to = "Type",
               values_to = "Value")


# Create the plot
ggplot(plot_data, aes(x = Location, y = Value, color = Type, group = Type)) +
  geom_point(size = 3, alpha = 0.8) +  # Scatter points
  geom_line(size = 1) +  # Connecting lines
  scale_color_manual(values = c("PovPct" = "#1D3557",        # Actual values
                               "Spline_3k_Predicted" = "#E63946", # Predicted
                                "Spline_4k_Predicted" = "#2A9D8F"),
                     labels = c("Actual", "Spline Predicted(3k)", "Spline Predicted(4k)")) +
  theme_minimal() +
  labs(title = "Actual vs. Spline Predicted Poverty Percentage by Location",
       x = "Location",
       y = "Poverty Percentage",
       color = "Legend") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6))  # Rotate x-axis labels
```




```
# Print results
cat("AIC:", AIC(spline_model_3k), "\n")
```

```
## AIC: 276.075
```

```
cat("BIC:", BIC(spline_model_3k), "\n")
```

```
## BIC: 326.3025
```

```
cat("R²:",  summary(spline_model_3k)$r.squared, "\n")
```

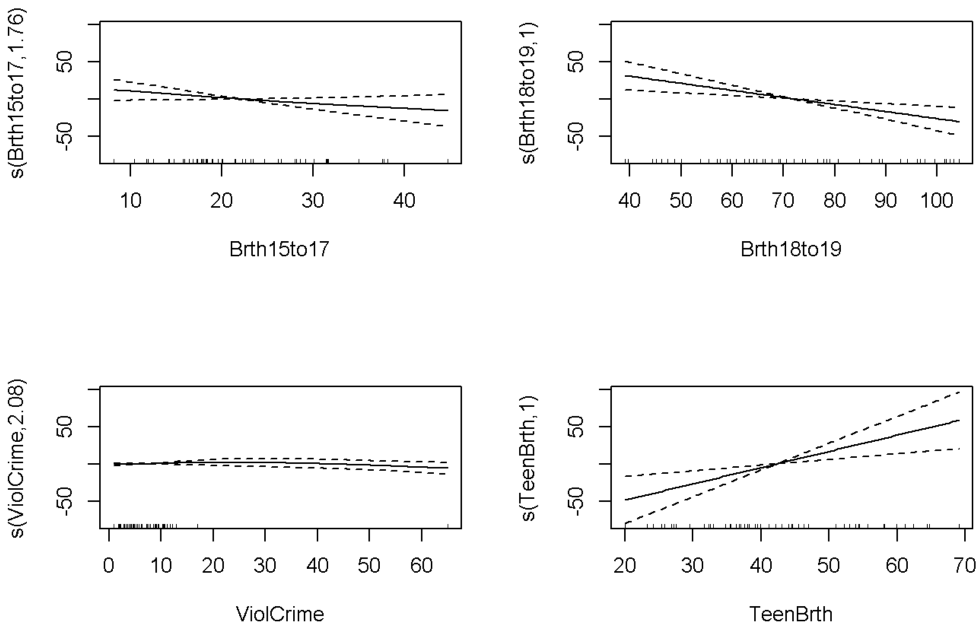
```
## R²: 0.7358023
```

Using Generalized Additive Model

```
# Fit a Generalized Additive Model (GAM)
gam_model <- gam(PovPct ~ s(Brth15to17) + s(Brth18to19) + s(ViolCrime) + s(TeenBrth), data = poverty_df)

summary(gam_model)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## PovPct ~ s(Brth15to17) + s(Brth18to19) + s(ViolCrime) + s(TeenBrth)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.1176     0.3777   34.73   <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(Brth15to17) 1.7630 2.2300  1.624 0.20315
## s(Brth18to19) 0.9999 0.9999 10.432 0.00235 **
## s(ViolCrime)  2.0816 2.1533  1.698 0.16608
## s(TeenBrth)   0.9999 0.9999  9.471 0.00359 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Rank: 36/37
## R-sq.(adj) =  0.602   Deviance explained = 64.9%
## GCV = 8.4029  Scale est. = 7.2752    n = 51
```

```
par(mfrow=c(2,2))
# Visualization of GAM effects
plot(gam_model)
```




```
AIC(lm_model, poly_model, spline_model_3k, spline_model_4k,gam_model)
```

```
##                         df      AIC
## lm_model         6.000000 255.5184
## poly_model      10.000000 257.5118
## spline_model_3k 26.000000 276.0750
## spline_model_4k 30.000000 276.2488
## gam_model        7.844385 254.2793
```

```
BIC(lm_model, poly_model, spline_model_3k, spline_model_4k,gam_model)
```

```
##                         df      BIC
## lm_model         6.000000 267.1094
## poly_model      10.000000 276.8300
## spline_model_3k 26.000000 326.3025
## spline_model_4k 30.000000 334.2036
## gam_model        7.844385 269.4333
```

Results

```
model_metrics <- data.frame(
  Model = c("Linear Regression", "Polynomial Regression", "Best AIC Model",
            "Best R² Model", "Spline (3 Knots)", "Spline (4 Knots)", "GAM"),
  AIC = c(AIC(lm_model), AIC(poly_model), AIC(best_aic_model),
          AIC(best_r2_model), AIC(spline_model_3k), AIC(spline_model_4k), AIC(gam_model)),
  BIC = c(BIC(lm_model), BIC(poly_model), BIC(best_aic_model),
          BIC(best_r2_model), BIC(spline_model_3k), BIC(spline_model_4k), BIC(gam_model)),
  Adjusted_R2 = c(summary(lm_model)$adj.r.squared, summary(poly_model)$adj.r.squared,
                  summary(best_aic_model)$adj.r.squared, summary(best_r2_model)$adj.r.squared,
                  summary(spline_model_3k)$adj.r.squared, summary(spline_model_4k)$adj.r.squared,
                  summary(gam_model)$r.sq)
)

# Print the data frame
print(model_metrics)
```

```
##                   Model      AIC      BIC Adjusted_R2
## 1     Linear Regression 255.5184 267.1094   0.5795512
## 2 Polynomial Regression 257.5118 276.8300   0.5906718
## 3        Best AIC Model 255.9755 281.0892   0.6197399
## 4         Best R² Model 268.4216 310.9217   0.5566678
## 5      Spline (3 Knots) 276.0750 326.3025   0.4919274
## 6      Spline (4 Knots) 276.2488 334.2036   0.4849705
## 7                   GAM 254.2793 269.4333   0.6023316
```

- The Generalized Additive Model (GAM) has the lowest AIC (254.2793) and BIC (269.4333), indicating it is the most parsimonious model with the best balance of goodness-of-fit and complexity.
- The AIC-optimal polynomial model with polynomial degrees (2, 4, 2, 3) for predictors achieved an adjusted R^2 of 0.6197, outperforming the linear model.
- The polynomial model with best AIC achieves the highest adjusted R^2 (0.6197399), indicating it explains the largest proportion of variance in the data while accounting for model complexity.
- The Linear Regression model has a decent performance with an adjusted R^2 of ~0.58, but non-linear models like GAM and Best AIC Model outperform it in terms of both adjusted R^2 and AIC/BIC.
- Both spline models (3 knots and 4 knots) perform poorly compared to other models, with higher AIC/BIC values and lower adjusted R^2 values.

Conclusion

The analysis demonstrates that non-linear relationships exist between poverty percentage (`PovPct`) and predictors like `TeenBrth` and `Brth18to19`. Among all tested models, GAM emerges as the best-performing model overall, combining flexibility with strong predictive power while avoiding overfitting.