FACULTY OF COMPUTER AND INFORMATION TECHNOLOGY

SEPTEMBER 2023

TEB 2043

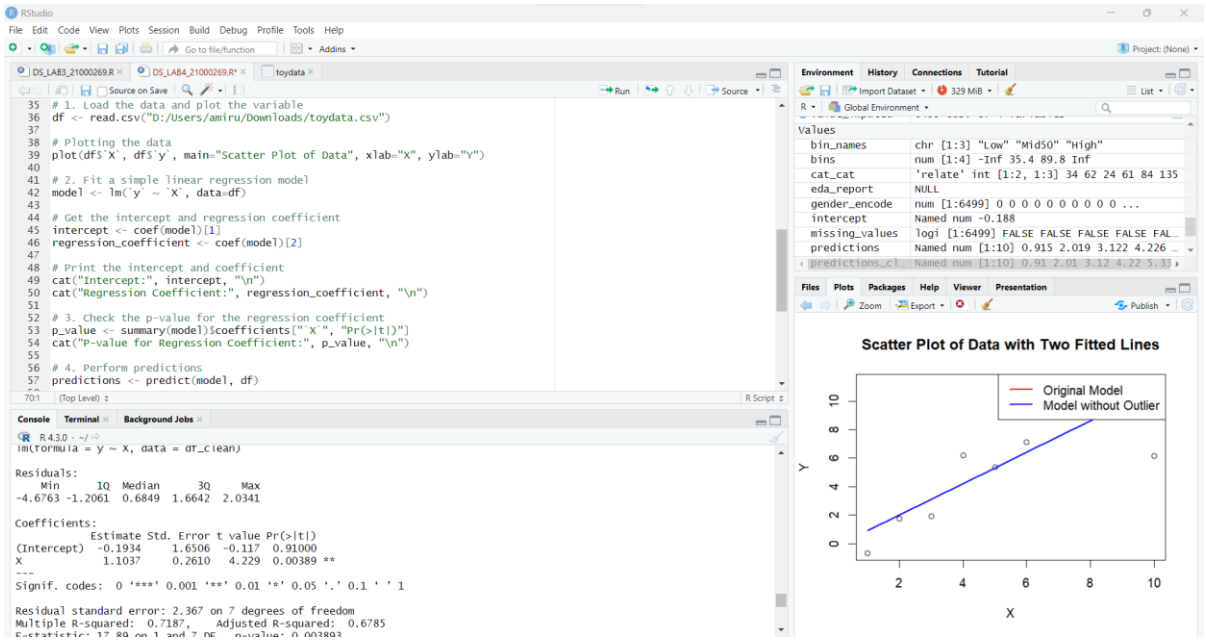
DATA SCIENCE

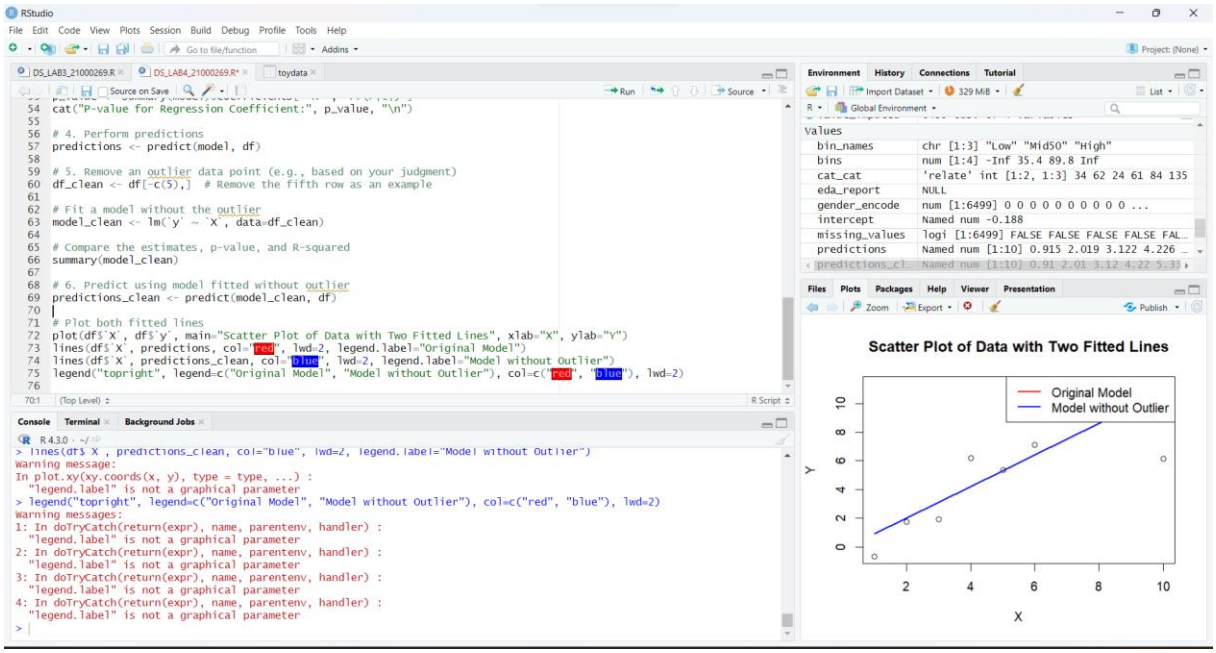
LAB 4

NO.	NAME	STUDENT ID	PROGRAM
1.	Amirul Hakim Muhamad Hafidz	21000269	Undergraduate of Science in Computer Science (Hons)

3.Regression coefficient's P-value:

- The null hypothesis that the coefficient is zero (there is no relationship between x and y) is tested by the p-value for the regression coefficient.
- A low p-value denotes a significant correlation between the coefficient and the target variable y.





#ACTIVITY 3

```
# 1. Load the data and plot the variable
df <- read.csv("D:/Users/amiru/Downloads/toydata.csv")

# Plotting the data
plot(df$`X`, df$`y`, main="Scatter Plot of Data", xlab="X", ylab="Y")

# 2. Fit a simple linear regression model
model <- lm(`y` ~ `X`, data=df)

# Get the intercept and regression coefficient
intercept <- coef(model)[1]
regression_coefficient <- coef(model)[2]

# Print the intercept and coefficient
cat("Intercept:", intercept, "\n")
cat("Regression Coefficient:", regression_coefficient, "\n")

# 3. Check the p-value for the regression coefficient
p_value <- summary(model)$coefficients["`X`", "Pr(>|t|)"]
cat("P-value for Regression Coefficient:", p_value, "\n")

# 4. Perform predictions
predictions <- predict(model, df)

# 5. Remove an outlier data point
df_clean <- df[-c(5),]  # Remove the fifth row as example

# Fit a model without the outlier
model_clean <- lm(`y` ~ `X`, data=df_clean)
```

```
# Compare the estimates, p-value, and R-squared
summary(model_clean)

# 6. Predict using model fitted without outlier
predictions_clean <- predict(model_clean, df)

# Plot both fitted lines
plot(df$`X`, df$`y`, main="Scatter Plot of Data with Two Fitted Lines", xlab="X", ylab="Y")
lines(df$`X`, predictions, col="red", lwd=2, legend.label="Original Model")
lines(df$`X`, predictions_clean, col="blue", lwd=2, legend.label="Model without Outlier")
legend("topright", legend=c("Original Model", "Model without Outlier"), col=c("red", "blue"), lwd=2)
```