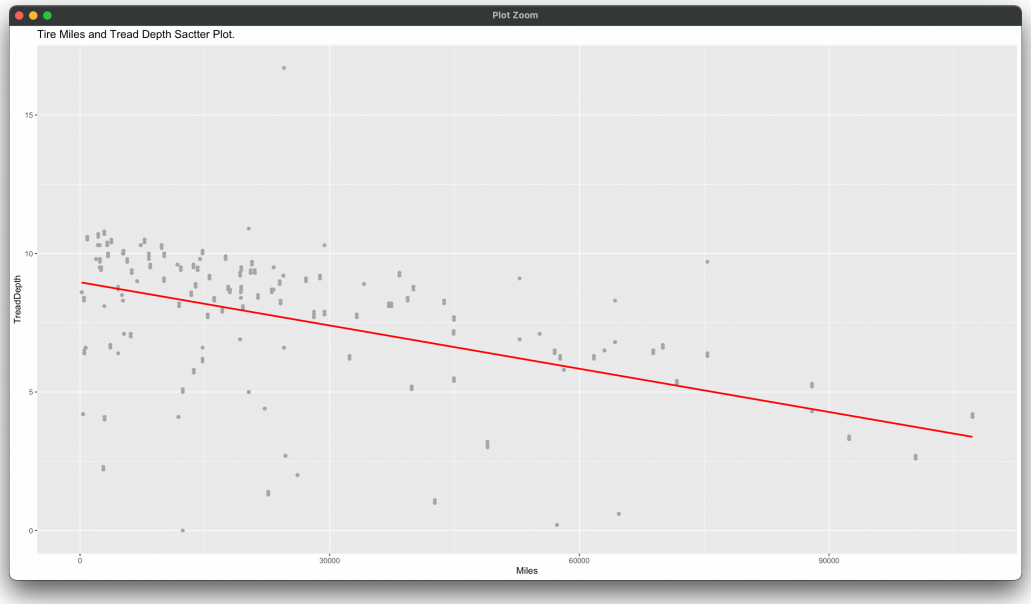
1. All code from your R script

```
# YUEH-TING WU
# MIS 545 Section 02
# Lab04WuY.R
# In this R programming, importing a csv file and imputing missing data, and
# then normalization, discretization and dummy coding.
# Then, make it visualization.

# Install the tidyverse and dummies packages
# install.packages("tidyverse")
# install.packages("dummies")

# Load the tidyverse and dummies libraries
library(tidyverse)
library(dummies)

# Set the working directory to my Lab04 folder
setwd("~/MIS 545/Lab04")

# Read the TireTread.csv into a tibble called tireTread1
tireTread1 <- read_csv(file = "TireTread.csv",
                       col_types = "cfnni",
                       col_names = TRUE)

# Display tireTread1 in the console
print(tireTread1)

# Display the structure of tireTread1 in the console
str(tireTread1)

# Display the summary of tireTread1 in the console
summary(tireTread1)

# Impute missing data for UsageMonths with the mean and store the result into
# a new tibble called tireTread2

tireTread2 <- tireTread1 %>%
  mutate(UsageMonths = ifelse(is.na(UsageMonths),
                              mean(UsageMonths, na.rm = TRUE),UsageMonths))

# Run a summary on tireTread2 and view it in the data viewer to ensure that the
# missing values have been replaced with the mean
summary(tireTread2)

# Determine outliers in the TreadDepth feature.
# Calculate outlier min and max and store into variables called outlierMin and
```

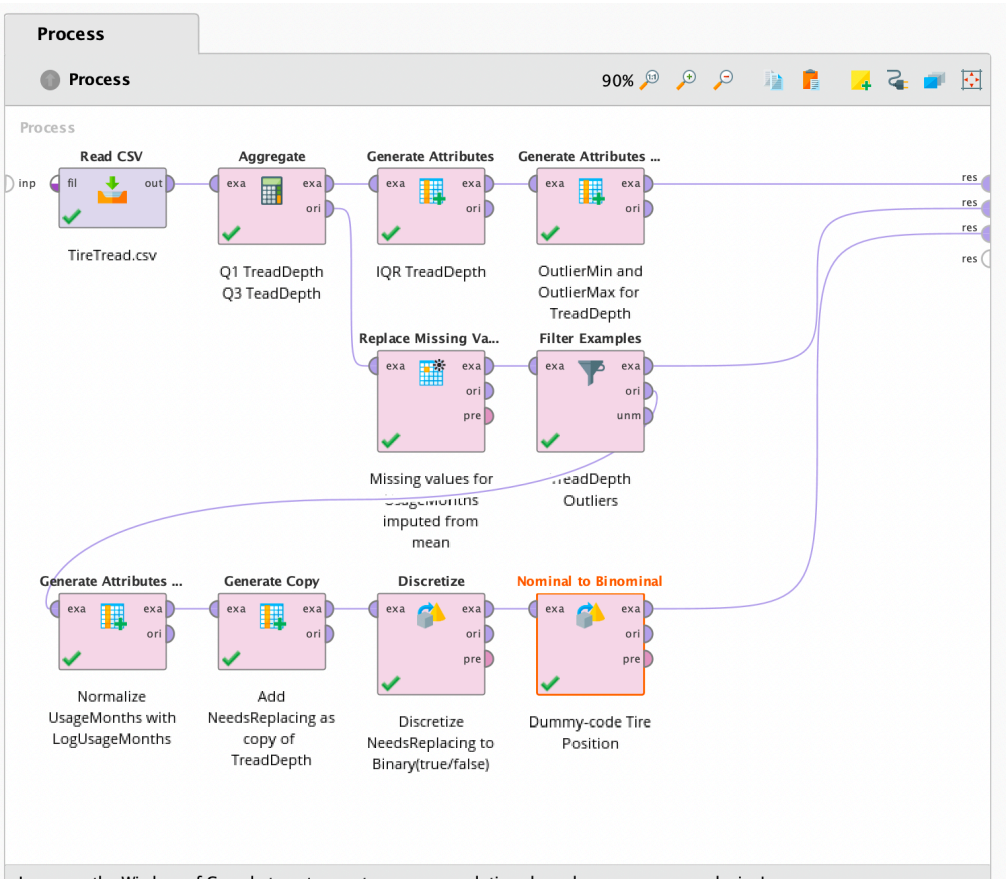
```
# outlierMax
outlierMin <- quantile(tireTread2$TreadDepth, .25) -
  (IQR(tireTread2$TreadDepth * 1.5))
outlierMax <- quantile(tireTread2$TreadDepth, .75) +
  (IQR(tireTread2$TreadDepth * 1.5))

# Keep the outliers in the dataset, but add the outliers to their own tibble
# treadDepthOutliers

treadDepthOutliers <- tireTread2 %>%
  filter(tireTread2$TreadDepth < outlierMin |
           tireTread2$TreadDepth > outlierMax)

# Normalize the UsageMonths feature by taking the log of UsageMonths into a new
# feature called LogUsageMonths and store the additional column in a tibble
# called tireTread3
tireTread3 <- tireTread2 %>%
  mutate(LogUsageMonths = log(UsageMonths))

# Discretize TreadDepth into NeedsReplacing (tires with tread depth of less
# than or equal to 1.6mm need replacing) and store into new tireTread4 tibble
tireTread4 <- tireTread3 %>%
  mutate(NeedsReplacing = TreadDepth <= 1.6)

# Dummy code the Position (LF, RF, LR, RR) feature
# Start by converting tireTread4 into a data frame
tireTread4DataFrame <- data.frame(tireTread4)
# Dummy code the Position using dummy.data.frame(), convert it back into a
# tibble, and store the result into a new tireTread5 tibble
tireTread5 <- as_tibble(dummy.data.frame(data = tireTread4DataFrame,
                                         name = "Position"))

# Use ggpolt() to build a scatter plot of Miles (x) with TreadDepth (y).
TireMilesandTreadDepthScatterPlot <- ggplot(data = tireTread5,
                                            aes(x = Miles,
                                                y = TreadDepth))
# Use the default point size, but change the point color to dark gray.
TireMilesandTreadDepthScatterPlot + geom_point()
TireMilesandTreadDepthScatterPlot + geom_point(color = "dark gray")
# Add a liner best fit line to the plot and color it red.
TireMilesandTreadDepthScatterPlot + geom_point(color = "dark gray") +
  geom_smooth(method = lm,
              level = 0,
              color = "red")
# Add a title to the scatter, "Tire Miles and Tread Depth Scatter Plot."
TireMilesandTreadDepthScatterPlot + geom_point(color = "dark gray") +
```

```
geom_smooth(method = lm,
            level = 0,
            color = "red") +
labs(title = "Tire Miles and Tread Depth Sactter Plot.")
```

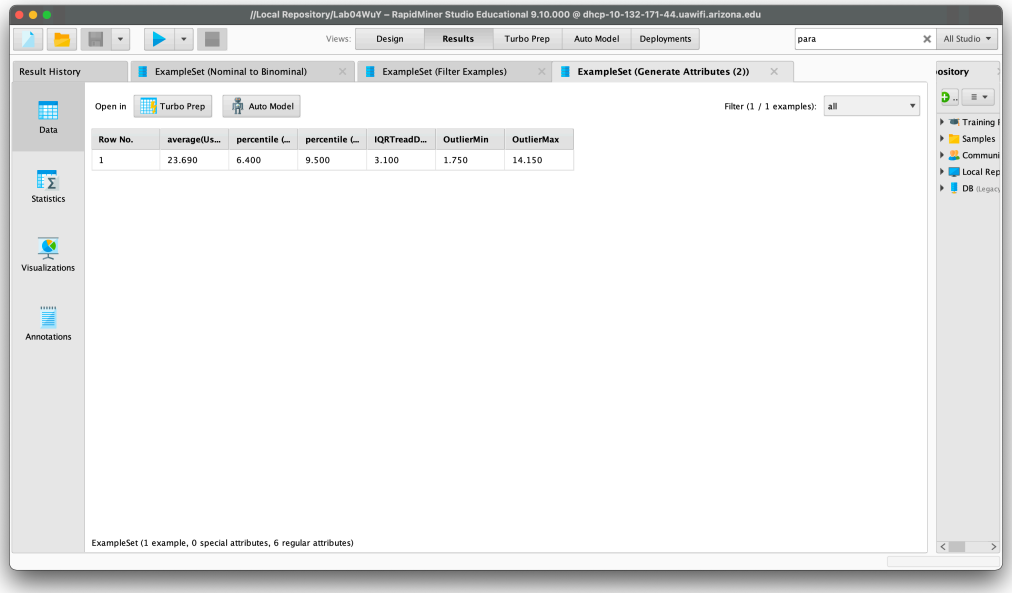
2. Scatter plot copy/pasted from R



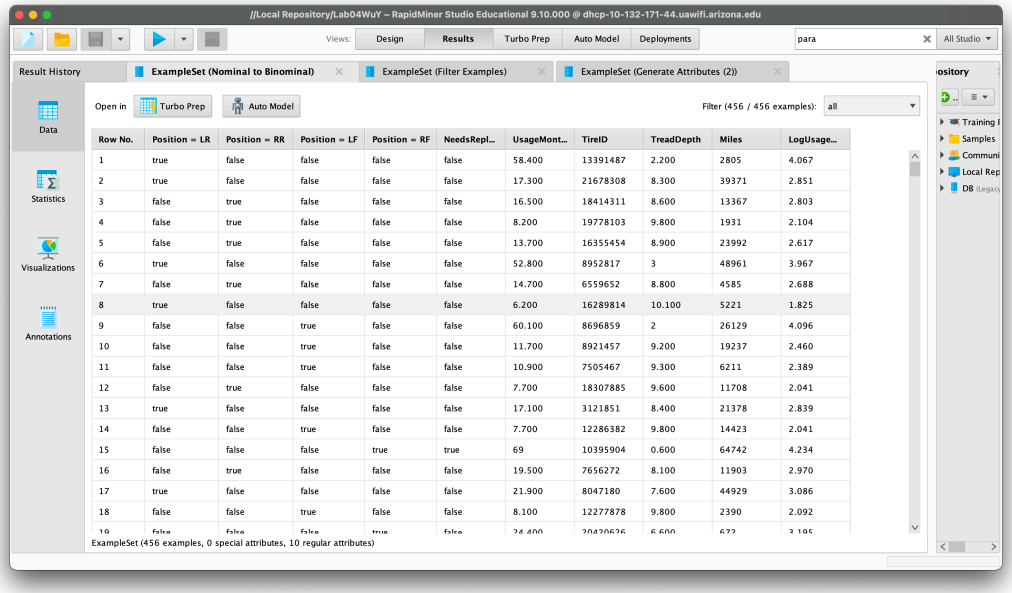
3. A screenshot of your RapidMiner process



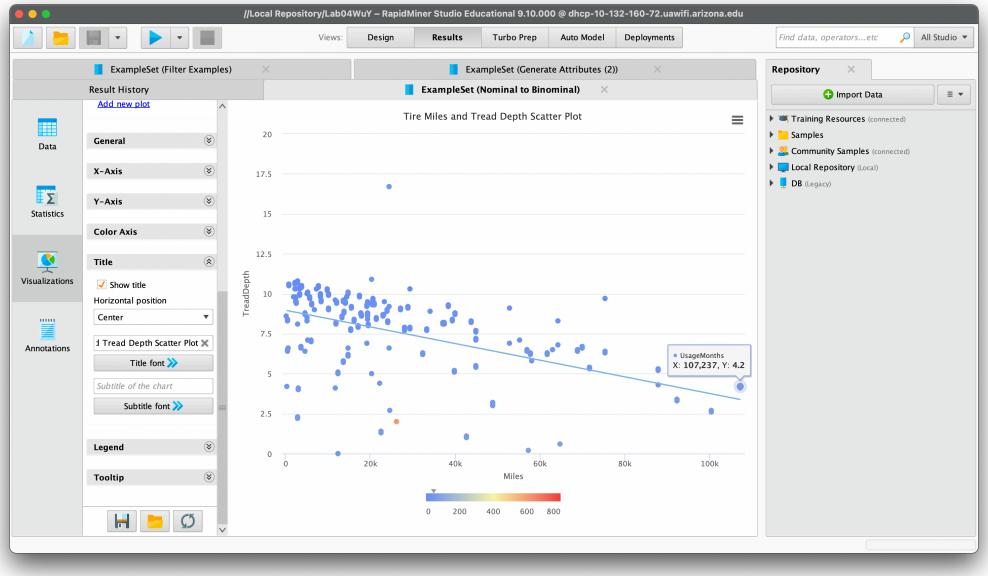
4. A screenshot of your identified outliers in RapidMiner result



5. A screenshot of your final RapidMiner results after performing all of the normalization, discretization, and dummy-coding steps



6. Scatter plot copy/paste from RapidMiner



7. Answer the following question in a brief paragraph: Based on a visual inspection of the tire miles and tread depth scatterplot, does a correlation exist between the two features? Explain why or why not.

Based on the scatter plot, I think there is a correlation between tire miles and tread depth: the tire miles are less, the tread depth is more higher.