**Experiment 9**

**Student Name: Husanpreet Kaur UID: 20BCS9512**

**Branch: BE-CSE Section/Group: 20BCS-DM-714/A**

**Semester: 6 Subject Code: 20CSP-376**

**Subject Name: Data Mining Lab Date of Performance: 09-05-2023**

1. **Aim/Overview of the practical:** Study of Regression Analysis using R Programming.
2. **Tools used:** RStudio
3. **Code:**

library(caTools)

library(ggplot2)

setwd("C:\\Users\\hp\\Documents\\DATA MINING CODES\\EXPERIMENT 9")

getwd()

data("Orange")

split = sample.split(Orange$age, SplitRatio = 0.9)

trainset = subset(Orange, split == TRUE)

testset = subset(Orange, split == FALSE)

lm.r= lm(formula = age ~ circumference, data = trainset)

coef(lm.r)

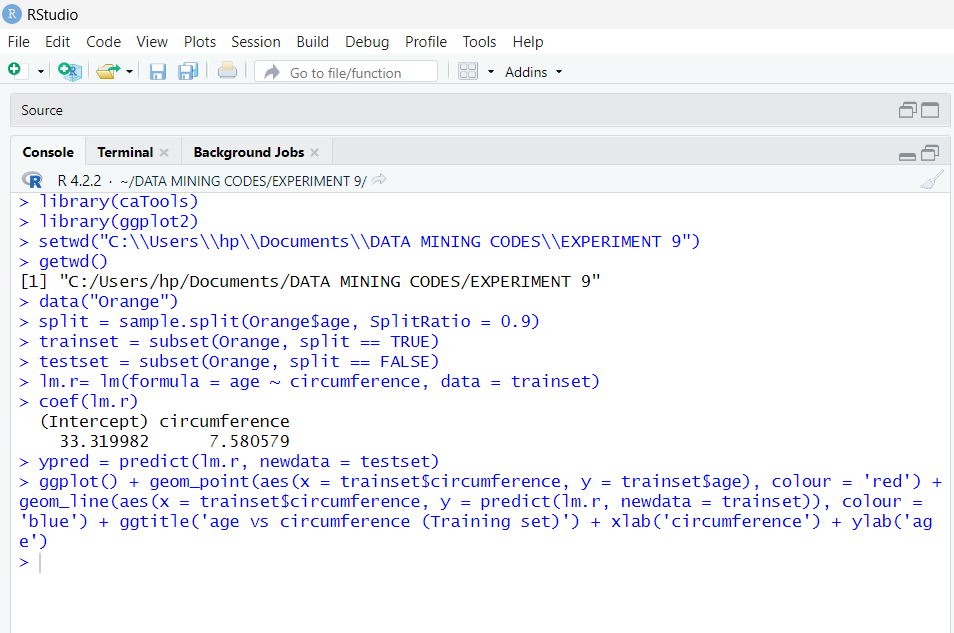
ypred = predict(lm.r, newdata = testset)

ggplot() + geom\_point(aes(x = trainset$circumference, y = trainset$age), colour = 'red') + geom\_line(aes(x = trainset$circumference, y = predict(lm.r, newdata = trainset)), colour = 'blue') + ggtitle('age vs circumference (Training set)') + xlab('circumference') + ylab('age')

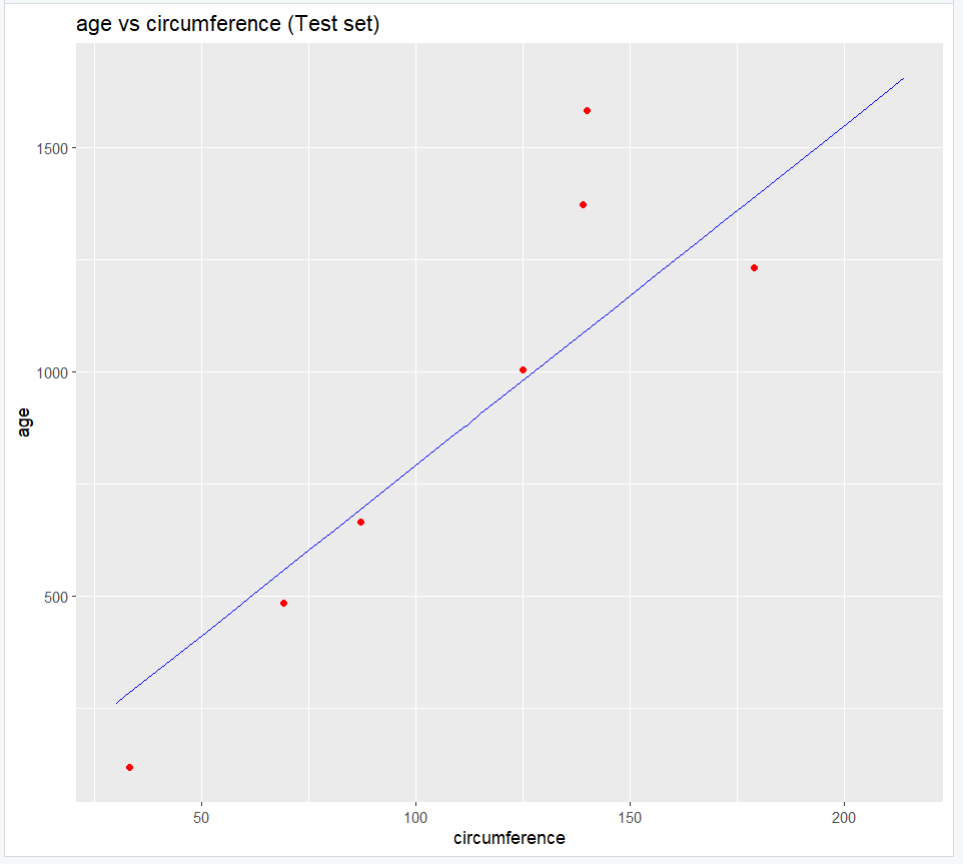
ggplot() + geom\_point(aes(x = testset$circumference, y = testset$age), colour = 'red') + geom\_line(aes(x = trainset$circumference, y = predict(lm.r, newdata = trainset)), colour = 'blue') + ggtitle('age vs circumference (Test set)') + xlab('circumference') + ylab('age')

1. **Output:**

RStudio:







1. **Observation:**

* Learnt how to use R and create a file in Rstudio.
* Learnt how to install packages in Rstudio.
* Learnt how to divide dataset in training and testing data.
* Learnt how to load dataset Orange in Rstudio.
* Learnt the use of caTools and ggplot2 libraries.
* Learnt how to plot data.