R Notebook

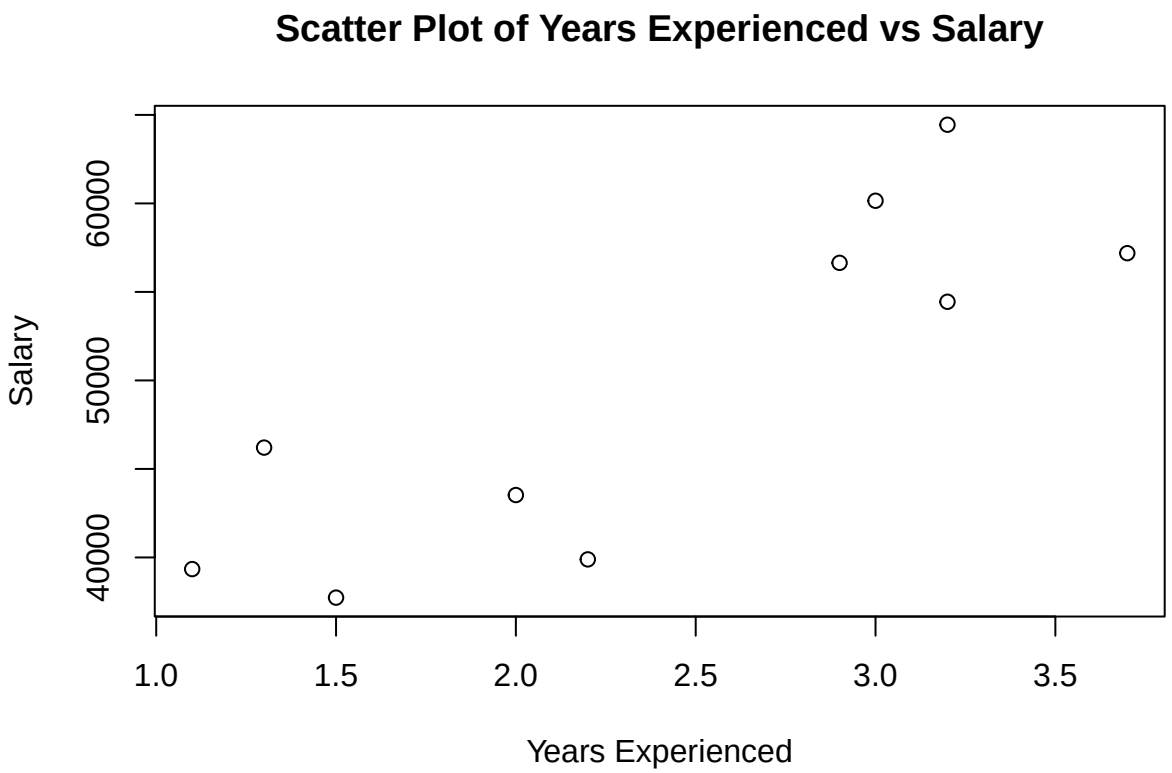
```
install.packages('caTools')
```

```
## Installing package into '/opt/R-3.5.3'
## (as 'lib' is unspecified)
```

```
## Warning in install.packages("caTools"): installation of package 'caTools' had
## non-zero exit status
```

```
# Create the data frame
data <- data.frame(
  Years_Exp = c(1.1, 1.3, 1.5, 2.0, 2.2, 2.9, 3.0, 3.2, 3.2, 3.7),
  Salary = c(39343.00, 46205.00, 37731.00, 43525.00,
             39891.00, 56642.00, 60150.00, 54445.00, 64445.00, 57189.00)
)
```

```
# Create the scatter plot
plot(data$Years_Exp, data$Salary,
     xlab = "Years Experienced",
     ylab = "Salary",
     main = "Scatter Plot of Years Experienced vs Salary")
```




```
library(caTools)
split = sample.split(data$Salary, SplitRatio = 0.7)
trainingset = subset(data, split == TRUE)
```

```
testset = subset(data, split == FALSE)

# Fitting Simple Linear Regression to the Training set
lm.r= lm(formula = Salary ~ Years_Exp,
         data = trainingset)
coef(lm.r)
```

```
## (Intercept)   Years_Exp
##   29078.818    8233.446
```

```
# Predicting the Test set results
ypred = predict(lm.r, newdata = testset)
```

```
library(ggplot2)

# Visualising the Training set results
ggplot() + geom_point(aes(x = trainingset$Years_Ex,
                          y = trainingset$Salary), colour = 'red') +
  geom_line(aes(x = trainingset$Years_Ex,
                y = predict(lm.r, newdata = trainingset)), colour = 'blue') +

  ggtitle('Salary vs Experience (Training set)') +
  xlab('Years of experience') +
  ylab('Salary')
```


```
# Visualising the Test set results
ggplot() +
  geom_point(aes(x = testset$Years_Exp, y = testset$Salary),
```

```
            colour = 'red') +
geom_line(aes(x = trainingset$Years_Exp,
              y = predict(lm.r, newdata = trainingset)),
          colour = 'blue') +
ggtitle('Salary vs Experience (Test set)') +
xlab('Years of experience') +
ylab('Salary')
```