4) Salary\_hike -> Build a prediction model for Salary\_hike

**Ans :**

**R Code :**

## Simple Linear Regression

########## Salary Consumed Data Set #########

Sal <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Simple Linear Regression\\Salary\_Data.csv')

Exp <- Sal$YearsExperience

SH <- Sal$Salary

boxplot(Exp, col="dodgerblue4")

boxplot(SH,col="dodgerblue4")

plot(Exp,SH,main="Scatter Plot",

col="Dodgerblue4",

col.main="Dodgerblue4",

col.lab="Dodgerblue4",

xlab="Years of Experience",

ylab="Salary", pch=20)

reg.model<-lm(SH~Exp, data=Sal)

summary(reg.model)

plot(Exp,SH,main="Line of Best Fit",col="Dodgerblue4", col.main="Dodgerblue4")

abline(reg.model, col="red")

res <- signif(residuals(reg.model))

pre <- predict(reg.model)

segments(Exp,SH,Exp,predict(reg.model))

predict(reg.model,newdata= data.frame(Exp=c(10,11)))

**Results :**

> reg.model<-lm(SH~Exp, data=Sal)

> summary(reg.model)

Call:

lm(formula = SH ~ Exp, data = Sal)

Residuals:

Min 1Q Median 3Q Max

-7958.0 -4088.5 -459.9 3372.6 11448.0

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 25792.2 2273.1 11.35 5.51e-12 \*\*\*

Exp 9450.0 378.8 24.95 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5788 on 28 degrees of freedom

Multiple R-squared: 0.957, Adjusted R-squared: 0.9554

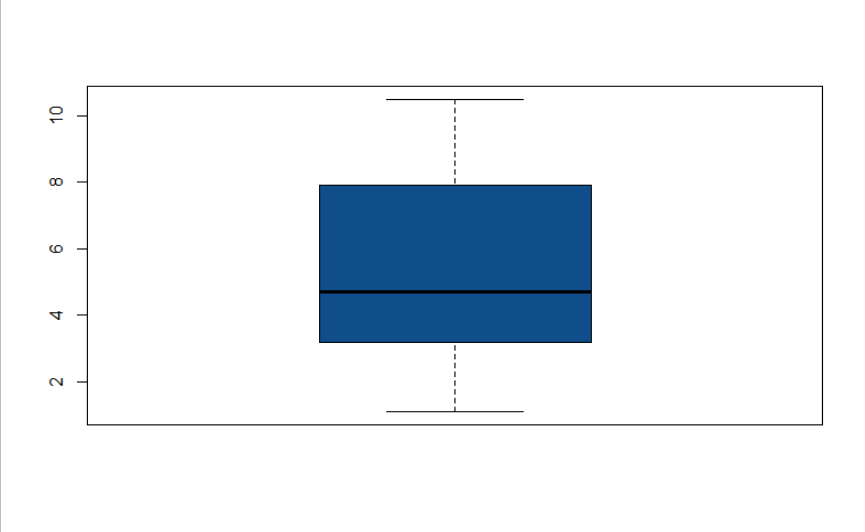
F-statistic: 622.5 on 1 and 28 DF, p-value: < 2.2e-16

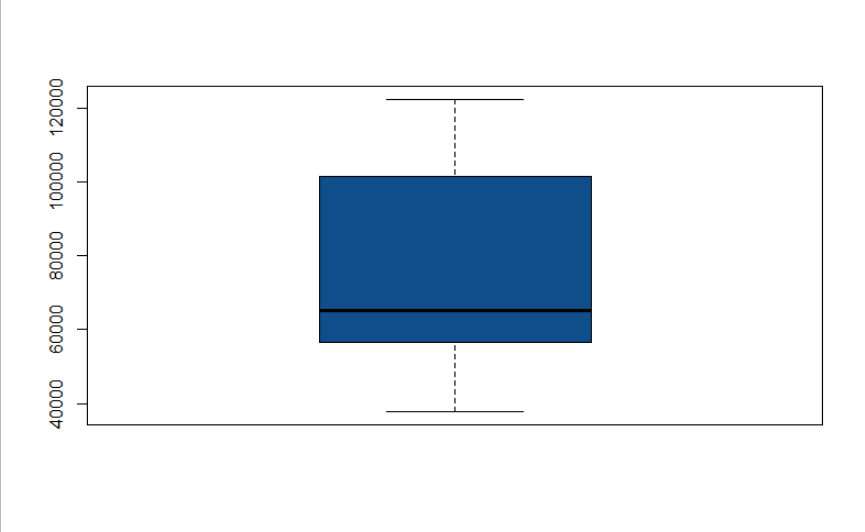
> predict(reg.model,newdata= data.frame(Exp=c(10,11)))

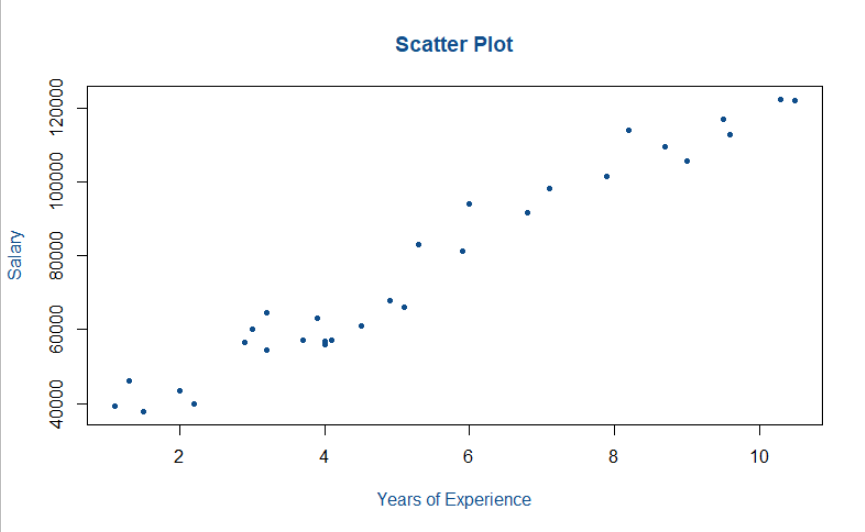
1 2

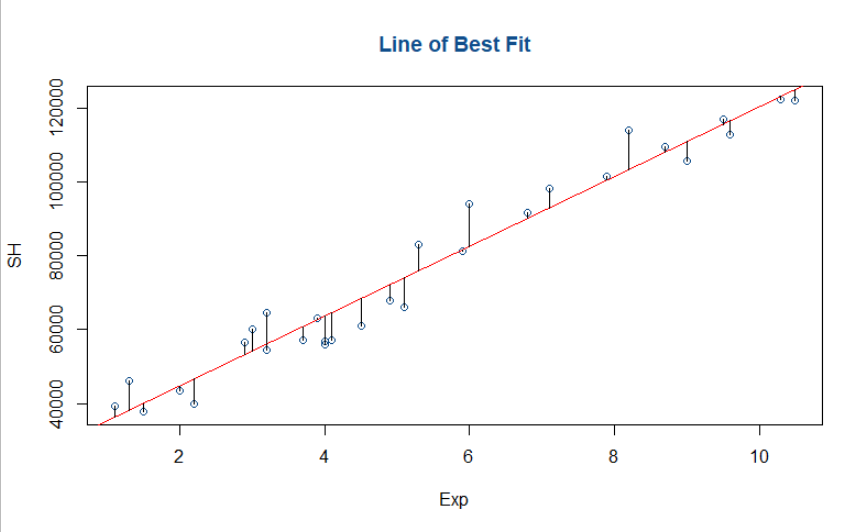
120291.8 129741.8

**Plots :**









**Inference :**

Getting good R squared value of 0.957.