Assignment

Vidipt Vashist MA22M025

2023-02-12

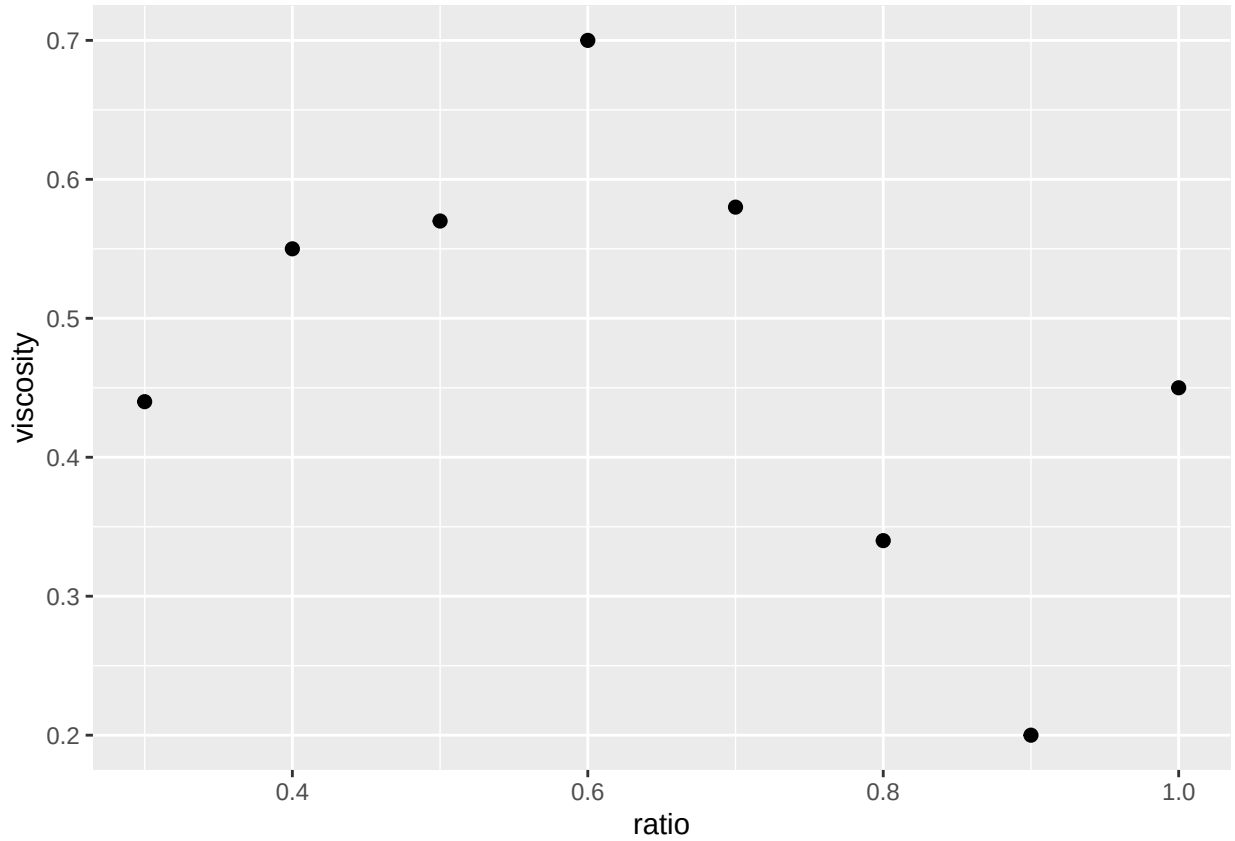
Question 2.14

Input data

```
ratio <- c(1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3)
viscosity <- c(0.45, 0.20, 0.34, 0.58, 0.70, 0.57, 0.55, 0.44)
df_14<- data.frame(ratio , viscosity)
```

a. Scatter plot

```
library(ggplot2) #importing library
ggplot(data =df_14, mapping = aes(x=ratio, y=viscosity)) + geom_point(size = 2) # size of point =3
```



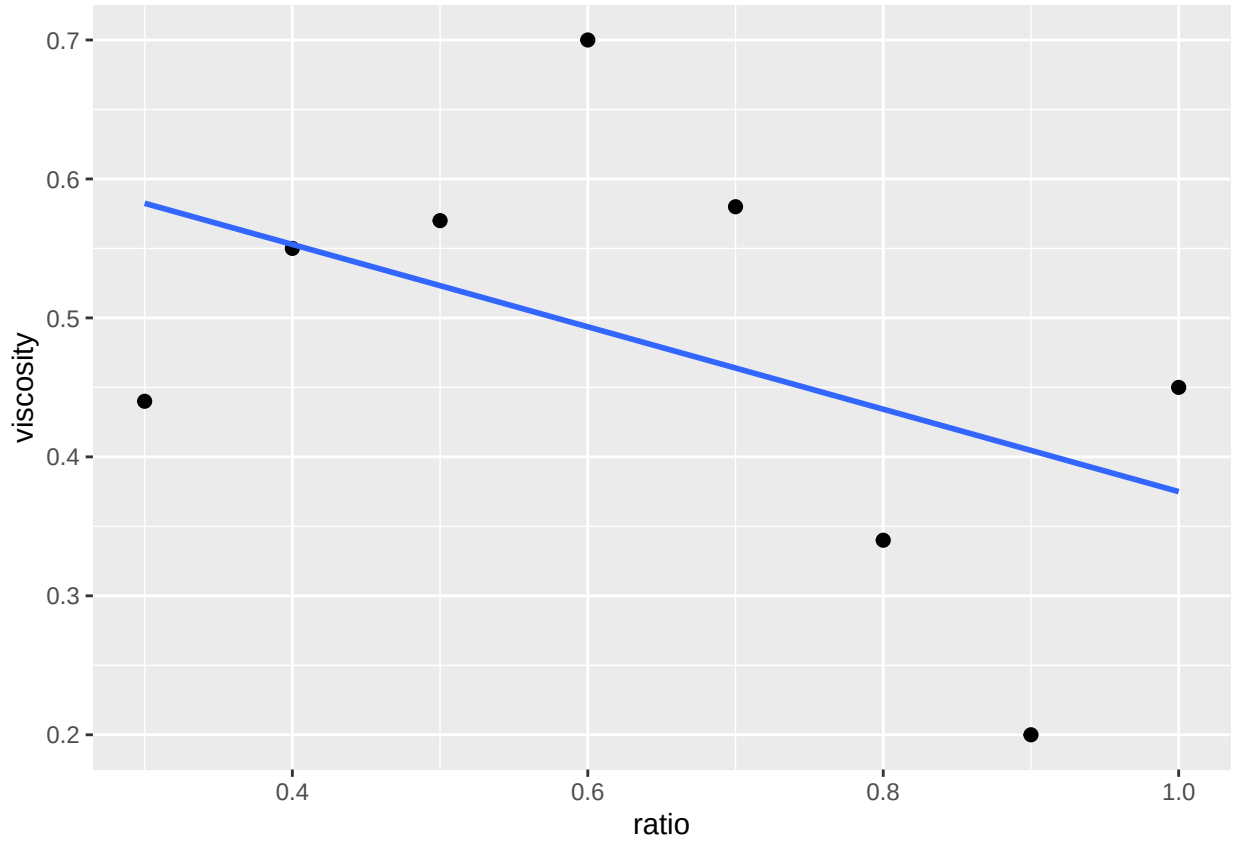
b. Estimate the prediction equation

```
model_14 = lm(viscosity~ratio)
coefficients(model_14)
```

```
## (Intercept)       ratio
##   0.6714286  -0.2964286
```

```
ggplot(data = df_14, mapping= aes(x=ratio, y=viscosity)) + geom_point(size = 2) + geom_smooth(method=lm
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



c. Analysis

```
summary(model_14)
```

Summary of model

```
##
## Call:
## lm(formula = viscosity ~ ratio)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.20464 -0.10634  0.02196  0.08527  0.20643
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.6714     0.1595   4.209  0.00563 **
## ratio        -0.2964     0.2314  -1.281  0.24754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.15 on 6 degrees of freedom
## Multiple R-squared:  0.2147, Adjusted R-squared:  0.08382
## F-statistic:  1.64 on 1 and 6 DF,  p-value: 0.2475
```

```
summary.aov(model_14)
```

Anova table

```
##             Df  Sum Sq Mean Sq F value Pr(>F)
## ratio        1 0.03691 0.03691    1.64  0.248
## Residuals    6 0.13498 0.02250
```

```
df(0.01 , 1 , 6)
```

```
## [1] 3.805085
```

thus accept null hypothesis Implies There no linear relation.

R2 value

```
summary(model_14)$r.squared
```

```
## [1] 0.2147065
```

model explain only 21% so linear relation is not present.

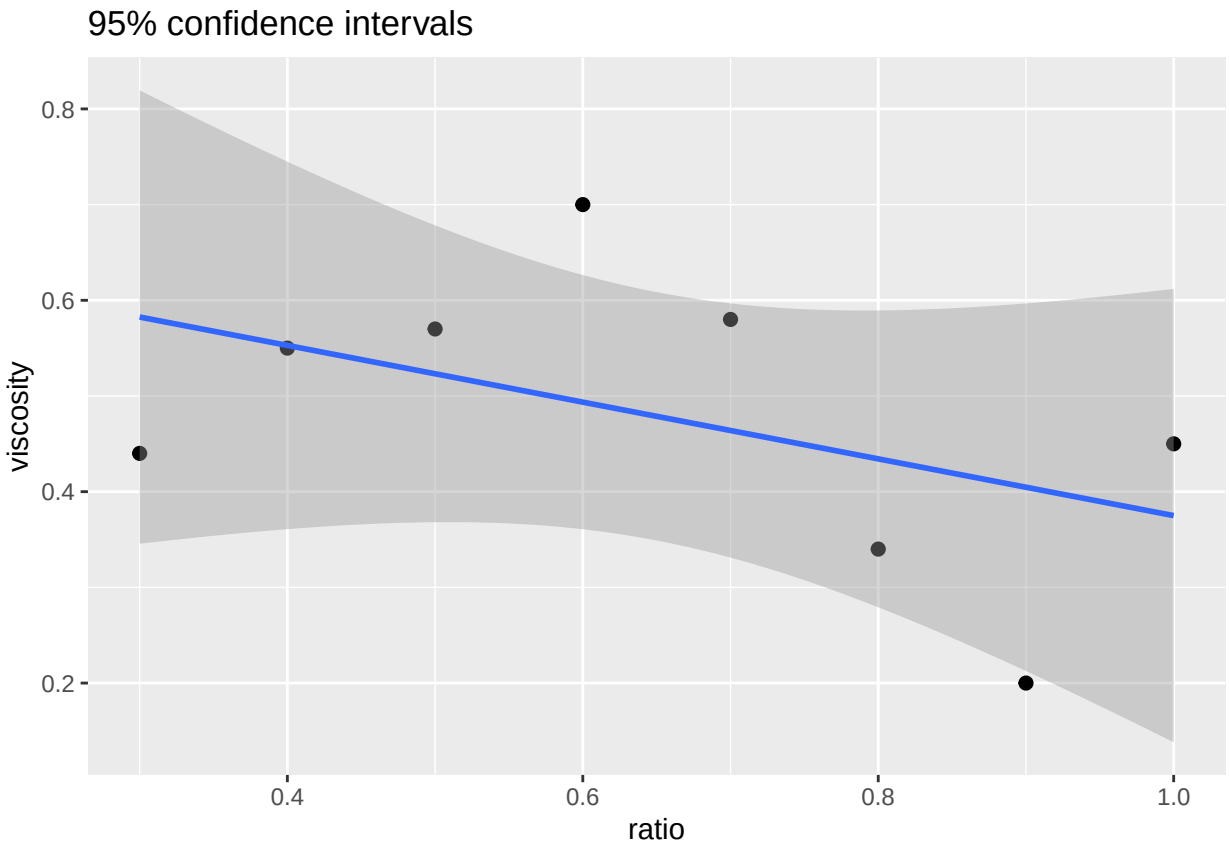
d. Calculate and plot 95% confidence and prediction band

```
cls<- predict(model_14 , interval ="confidence" , level = 0.95)
cls
```

```
##         fit       lwr       upr
## 1 0.3750000 0.1380944 0.6119056
## 2 0.4046429 0.2125969 0.5966888
## 3 0.4342857 0.2791946 0.5893768
## 4 0.4639286 0.3311165 0.5967406
## 5 0.4935714 0.3607594 0.6263835
## 6 0.5232143 0.3681232 0.6783054
## 7 0.5528571 0.3608112 0.7449031
## 8 0.5825000 0.3455944 0.8194056
```

```
ggplot(df_14, aes(x=ratio, y=viscosity)) + geom_point(size = 2) + geom_smooth(method=lm , se=TRUE) +
labs(title="95% confidence intervals",
       x=" ratio ", y = "viscosity")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



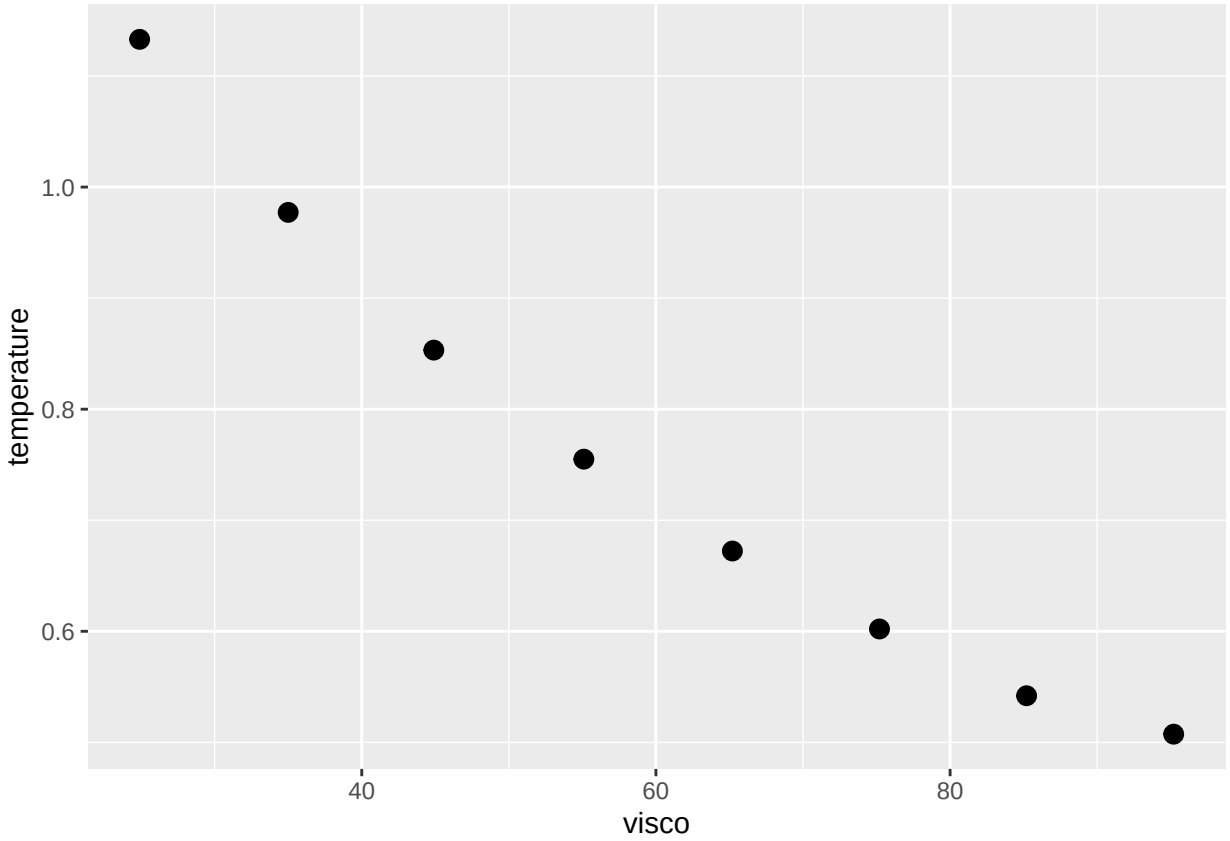
Question 2.15

Input data

```
visco <- c(24.9, 35, 44.9, 55.1, 65.2, 75.2, 85.2, 95.2)
temperature <- c(1.1330, 0.9772, 0.8532, 0.7550, 0.6723, 0.6021, 0.5420, 0.5074)
df <- data.frame(temperature , visco)
df
```

```
##   temperature visco
## 1      1.1330  24.9
## 2      0.9772  35.0
## 3      0.8532  44.9
## 4      0.7550  55.1
## 5      0.6723  65.2
## 6      0.6021  75.2
## 7      0.5420  85.2
## 8      0.5074  95.2
```

```
# Scatter plot
library(ggplot2)
ggplot(df, aes(x=visco, y=temperature)) + geom_point(size = 3)
```



a. Estimate the prediction equation

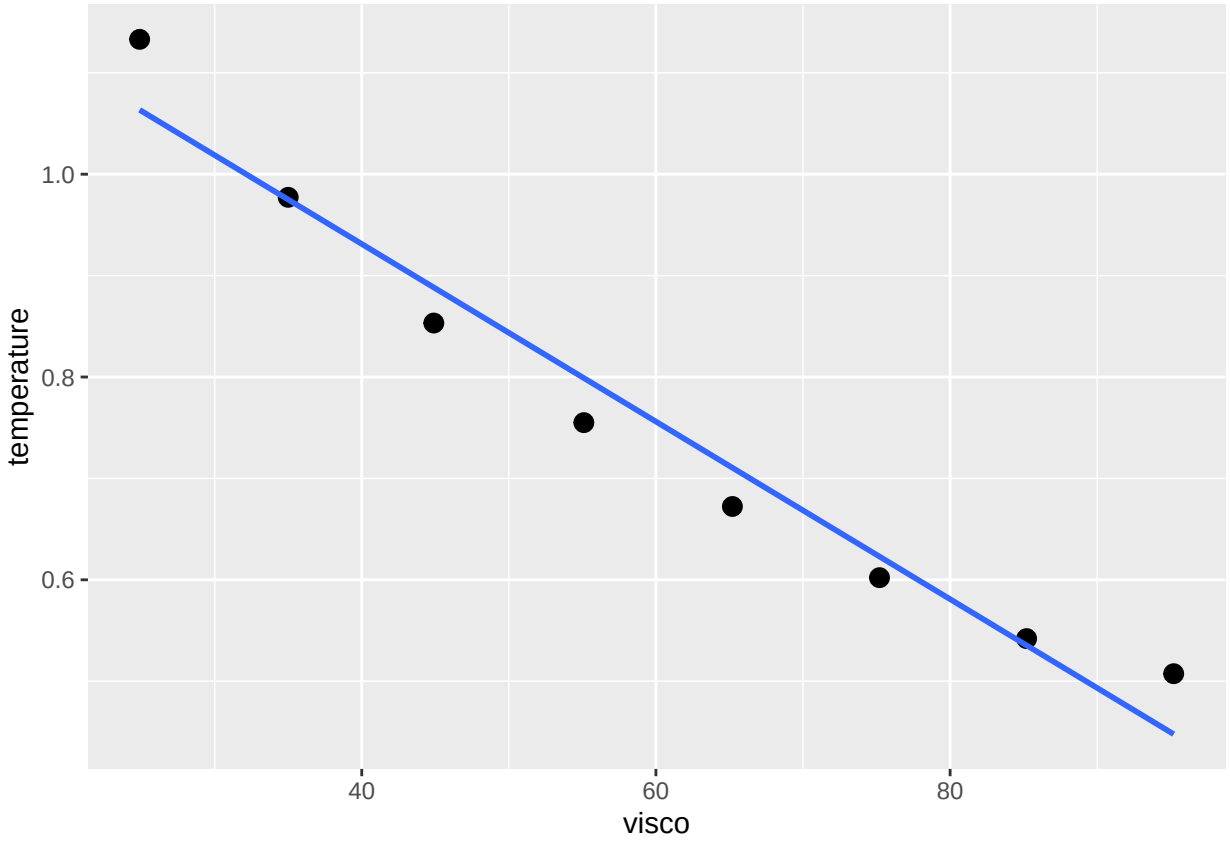
```
model_15 <- lm(temperature~visco)
coefficients(model_15)
```

```
##  (Intercept)        visco
##  1.281510655 -0.008757822
```

Regression line

```
ggplot(df, aes(x=visco, y=temperature)) + geom_point(size = 3) + geom_smooth(method=lm , se=FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



b. Analysis of model

```
summary(model_15)
```

```
##
## Call:
## lm(formula = temperature ~ visco)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.043955 -0.035863 -0.009305  0.019900  0.069559
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.2815107  0.0468683   27.34 1.58e-07 ***
## visco       -0.0087578  0.0007284  -12.02 2.01e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04743 on 6 degrees of freedom
## Multiple R-squared:  0.9602, Adjusted R-squared:  0.9535
## F-statistic: 144.6 on 1 and 6 DF,  p-value: 2.007e-05
```

```
summary.aov(model_15)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
```

```
## visco        1 0.3253  0.3253   144.6 2.01e-05 ***
## Residuals    6 0.0135  0.0022
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
df(0.001 , 1, 6)
```

```
## [1] 12.09602
```

As our F value = 144.6 > f(0.001, 1 , 6)

we reject null hypothesis and accept alternate hypothesis which implies there is linear relation

```
summary(model_15)$r.squared
```

```
## [1] 0.9601535
```

Model explain 96% of variation.

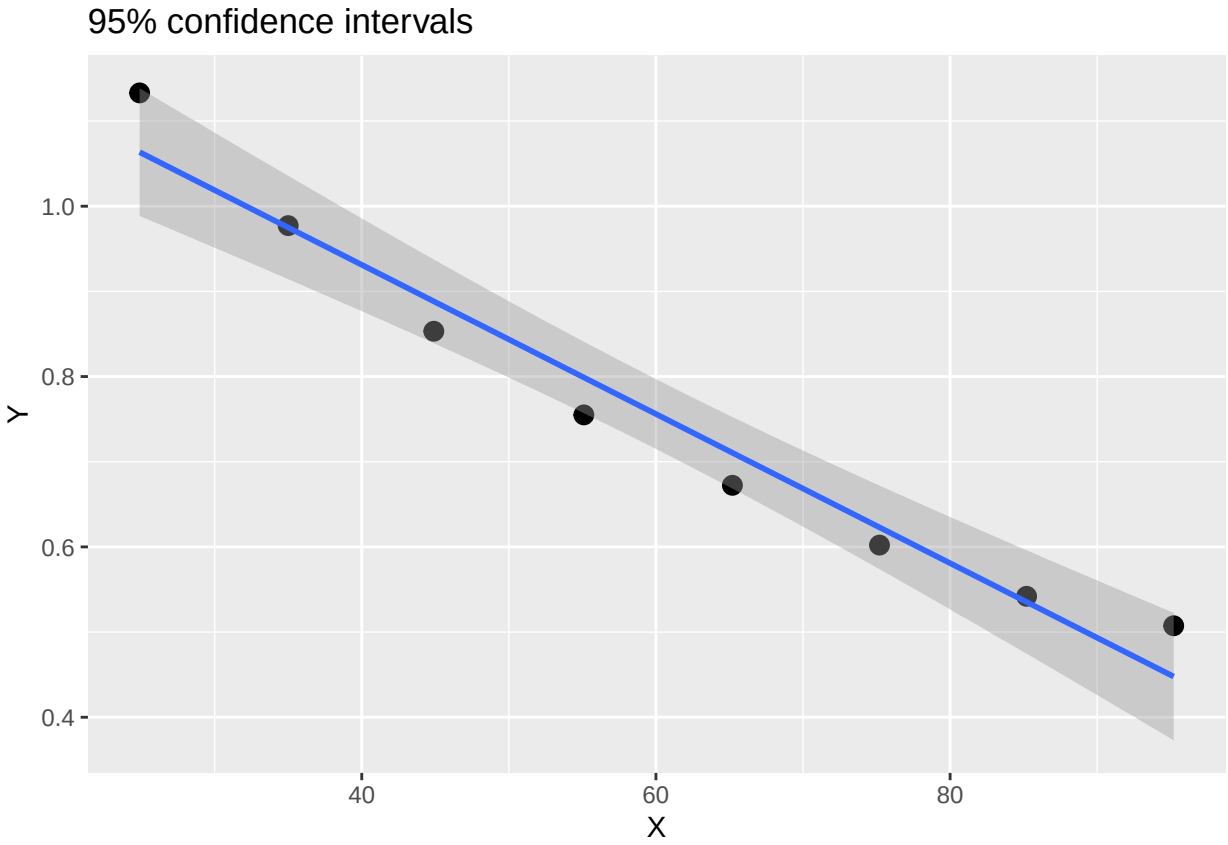
c. Calculate and plot 95% confidence intervals

```
cls_15<-predict(model_15 , interval="confidence" , level =0.95)
cls_15<-data.frame(cls_15) # Prediction bands in DataFrame
cls_15$temperature<-temperature
cls_15
```

```
##         fit       lwr       upr temperature
## 1 1.0634409 0.9884960 1.1383857      1.1330
## 2 0.9749869 0.9142988 1.0356750      0.9772
## 3 0.8882844 0.8391259 0.9374430      0.8532
## 4 0.7989546 0.7569675 0.8409418      0.7550
## 5 0.7105006 0.6684657 0.7525355      0.6723
## 6 0.6229224 0.5738373 0.6720075      0.6021
## 7 0.5353442 0.4746233 0.5960651      0.5420
## 8 0.4477660 0.3729329 0.5225990      0.5074
```

```
ggplot(df, aes(x=visco, y=temperature)) + geom_point(size = 3) +
geom_smooth(method=lm, se=TRUE, fullrange=FALSE, level=0.95) +
labs(title="95% confidence intervals",
       x=" X ", y = "Y")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Question 2.16

Input Data

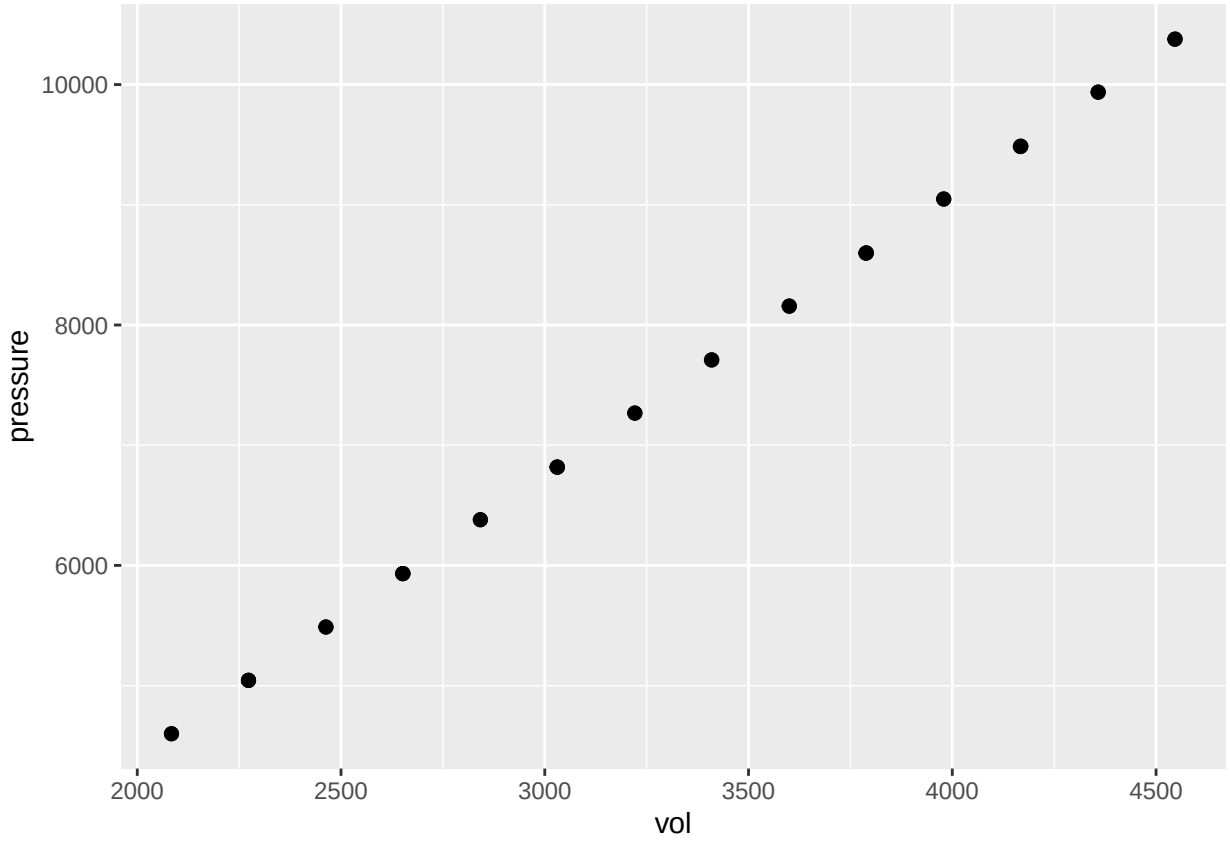
```
vol<- c(2084,2084 , 2273, 2273,2273, 2463 ,2463 ,2651 , 2652,2652,2842, 2842,3030,3031,3031,3221,3221,3
)
pressure <-c(4599,4600,5044,5043,5044,5488,5487,5931,5932,5932,6380,6380,6818,6817,6818,7266,7268,7709,
)
df_16 <- data.frame(vol , pressure)
df_16
```

```
##     vol pressure
## 1  2084     4599
## 2  2084     4600
## 3  2273     5044
## 4  2273     5043
## 5  2273     5044
## 6  2463     5488
## 7  2463     5487
## 8  2651     5931
## 9  2652     5932
## 10 2652     5932
## 11 2842     6380
## 12 2842     6380
## 13 3030     6818
```

```
## 14 3031     6817
## 15 3031     6818
## 16 3221     7266
## 17 3221     7268
## 18 3409     7709
## 19 3410     7710
## 20 3600     8156
## 21 3600     8158
## 22 3788     8597
## 23 3789     8599
## 24 3789     8600
## 25 3979     9048
## 26 3979     9048
## 27 4167     9484
## 28 4168     9487
## 29 4168     9487
## 30 4358     9936
## 31 4358     9938
## 32 4546    10377
## 33 4547    10379
```

Scatter Plot

```
ggplot(df_16 , aes(x =vol , y = pressure))+ geom_point(size = 2)
```



Correlation

```
cor(df_16$pressure , df_16$vol)
```

```
## [1] 0.9999995
```

Very high correlation

Fitting of model

```
model_16 <- lm(df_16$vol~df_16$pressure)
summary(model_16)
```

```
##
## Call:
## lm(formula = df_16$vol ~ df_16$pressure)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2596 -0.5411 -0.0325  0.3937  1.8441
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.239e+02  5.568e-01   222.5   <2e-16 ***
## df_16$pressure 4.262e-01  7.278e-05  5855.4   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7421 on 31 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 3.429e+07 on 1 and 31 DF,  p-value: < 2.2e-16
```

Anova Table

```
summary.aov(model_16)
```

```
##                Df   Sum Sq  Mean Sq  F value Pr(>F)
## df_16$pressure  1 18879734 18879734 34286009 <2e-16 ***
## Residuals      31       17        1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
df(0.001 , 1 ,31)
```

```
## [1] 12.50789
```

We reject null hypothesis implies linear relation.

R2 value

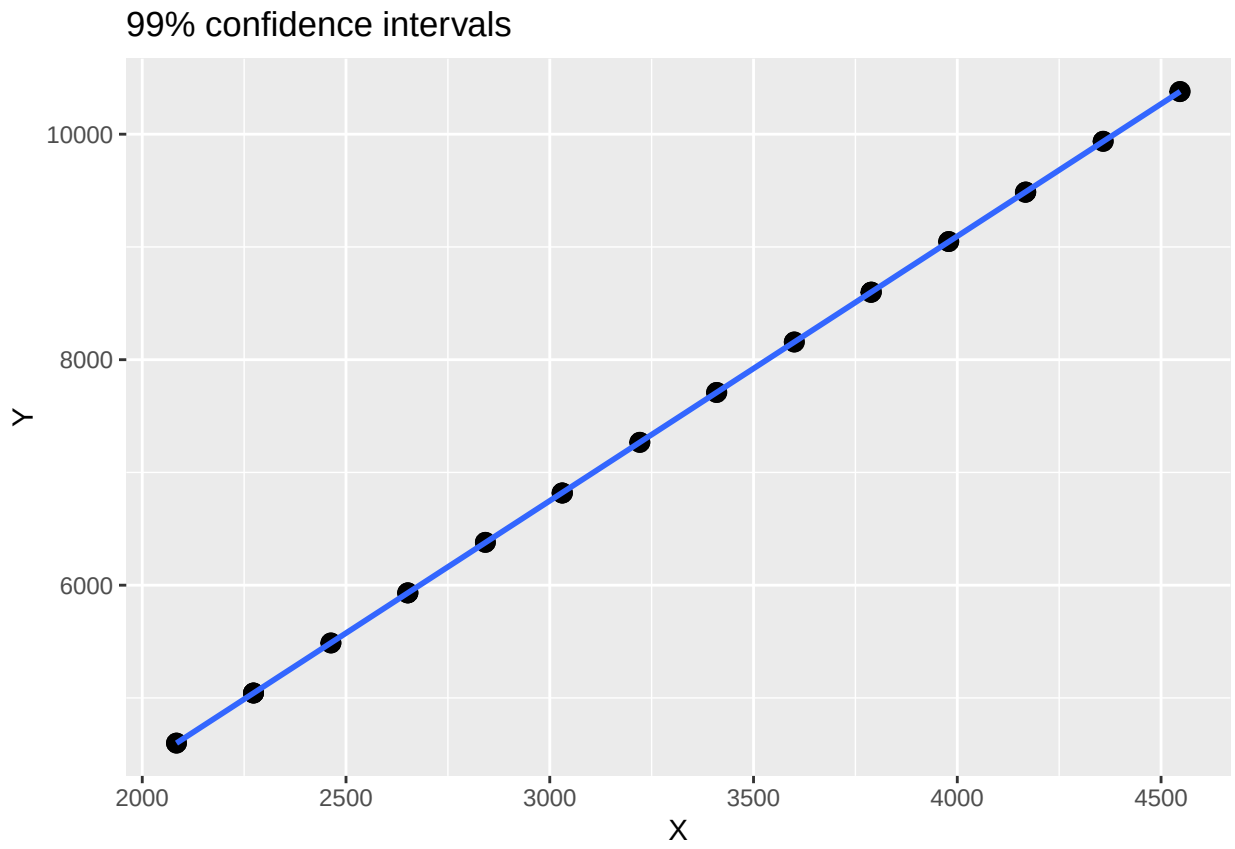
```
summary(model_16)$r.squared
```

```
## [1] 0.9999991
```

model explain 99% variation

```
ggplot(df_16, aes(x=vol ,y=pressure)) + geom_point(size = 3) + geom_smooth(method=lm, se=TRUE, fullrang
labs(title="99% confidence intervals",
       x=" X ", y = "Y")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Question 2.17

Input Data

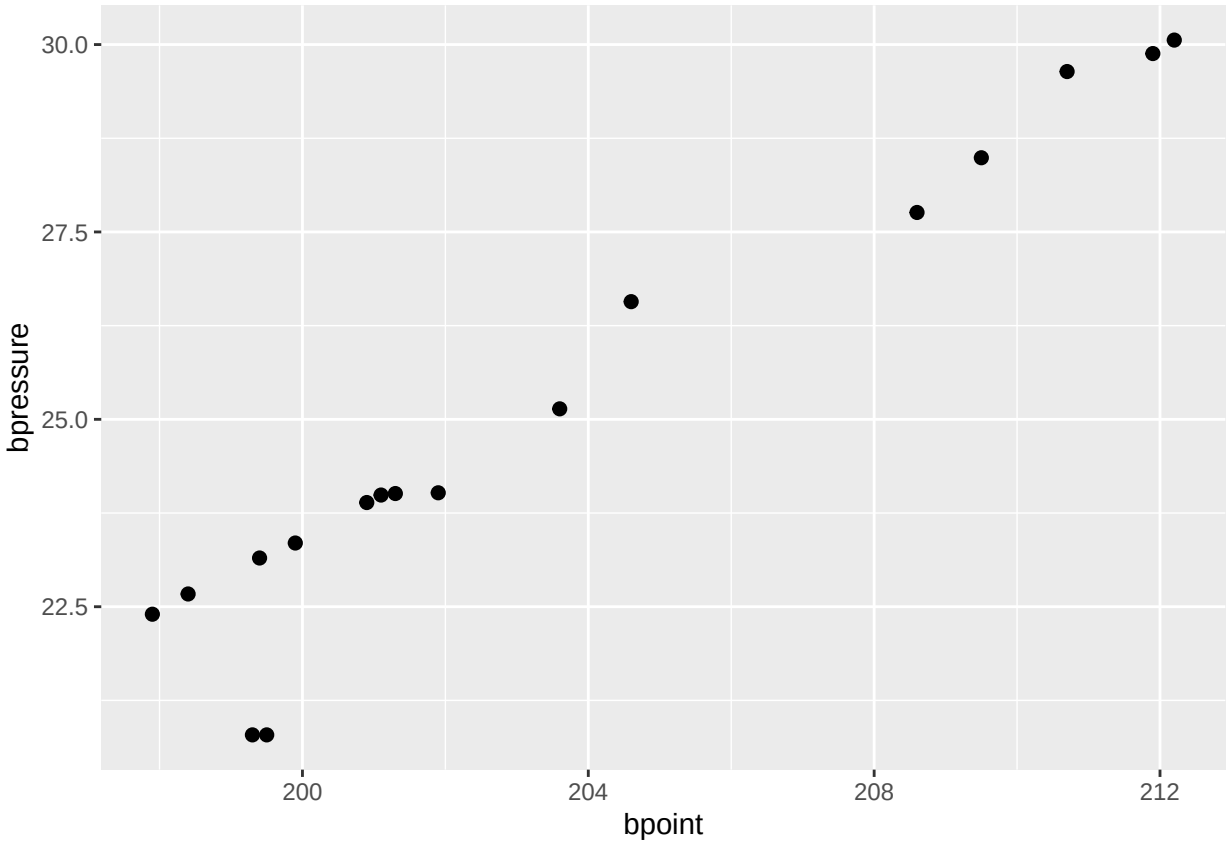
```
bpoint <- c(199.5,199.3,197.9,198.4,199.4,199.9,200.9,201.1,201.9,201.3,203.6,204.6,209.5,208.6,210.7,2

)
bpressure <- c(20.79,20.79,22.40,22.67,23.15,23.35,23.89,23.99,24.02,24.01,25.14,26.57,28.49,27.76,29.6
)
df_17 <- data.frame(bpoint , bpressure)
df_17
```

```
##    bpoint bpressure
## 1   199.5     20.79
## 2   199.3     20.79
## 3   197.9     22.40
## 4   198.4     22.67
## 5   199.4     23.15
## 6   199.9     23.35
```

```
## 7   200.9     23.89
## 8   201.1     23.99
## 9   201.9     24.02
## 10  201.3     24.01
## 11  203.6     25.14
## 12  204.6     26.57
## 13  209.5     28.49
## 14  208.6     27.76
## 15  210.7     29.64
## 16  211.9     29.88
## 17  212.2     30.06
```

Scatter Plot

```
ggplot(df_17 , aes(x =bpoint , y = bpressure))+ geom_point(size = 2)
```



Corelation value

```
cor(bpoint,bpressure)
```

```
## [1] 0.9683853
```

Fitting model

```r
model_17 <- lm(bpressure ~ bpoint)
summary(model_17)
```

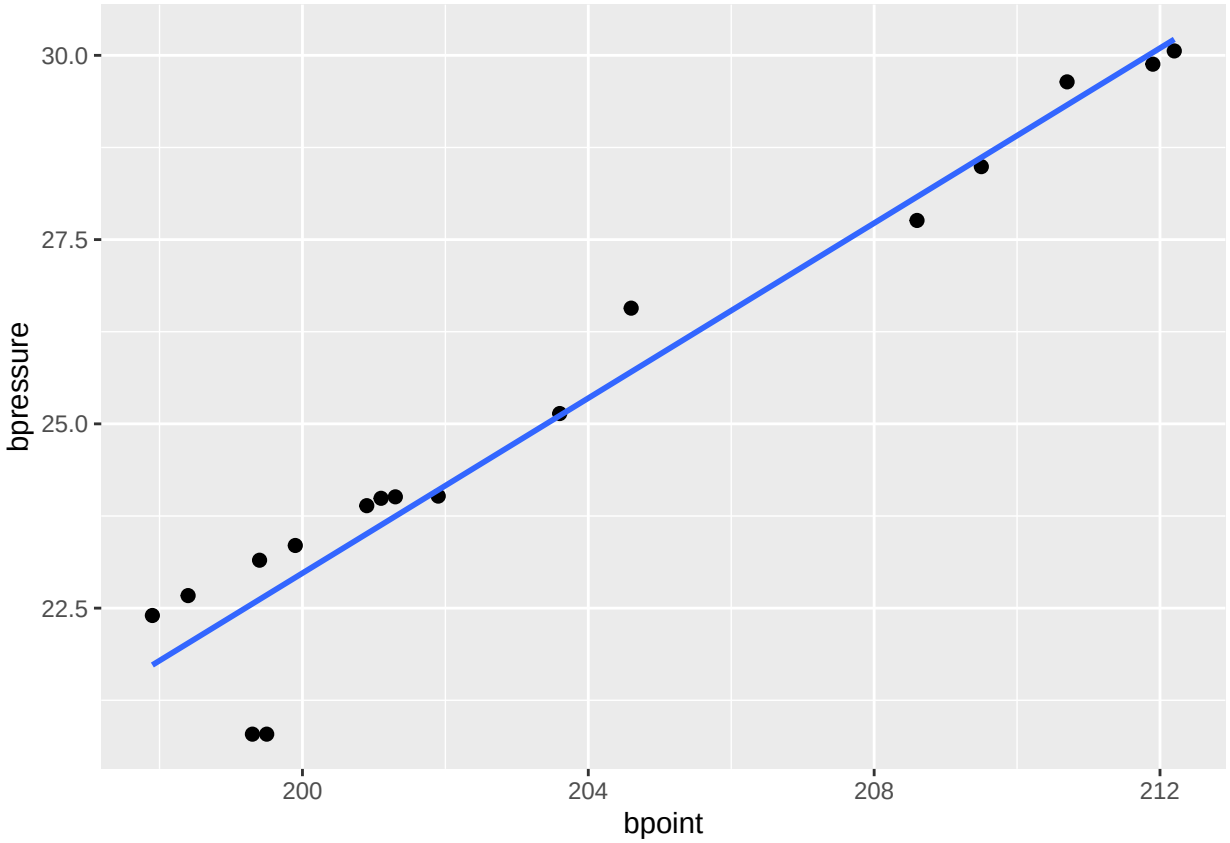
```
##
## Call:
## lm(formula = bpressure ~ bpoint)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8876 -0.1570  0.2638  0.4350  0.8648
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -95.75717    8.04046  -11.91 4.79e-09 ***
## bpoint        0.59366    0.03949   15.04 1.88e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7916 on 15 degrees of freedom
## Multiple R-squared:  0.9378, Adjusted R-squared:  0.9336
## F-statistic:   226 on 1 and 15 DF,  p-value: 1.879e-10
```

```r
coefficients(model_17)
```

```
## (Intercept)      bpoint
## -95.7571683   0.5936579
```

```r
# Regression line
ggplot(df_17 , aes(x =bpoint , y = bpressure))+ geom_point(size = 2) + geom_smooth(method=lm , se=FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Anova Table

```
summary.aov(model_17)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## bpoint       1  141.7  141.65     226 1.88e-10 ***
## Residuals   15    9.4    0.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
df(0.001 , 1 , 15)
```

```
## [1] 12.40068
```

Thus we reject null hypothesis.

R2 value

```
summary(model_17)$r.squared
```

```
## [1] 0.9377701
```

Model explain 93% of variation

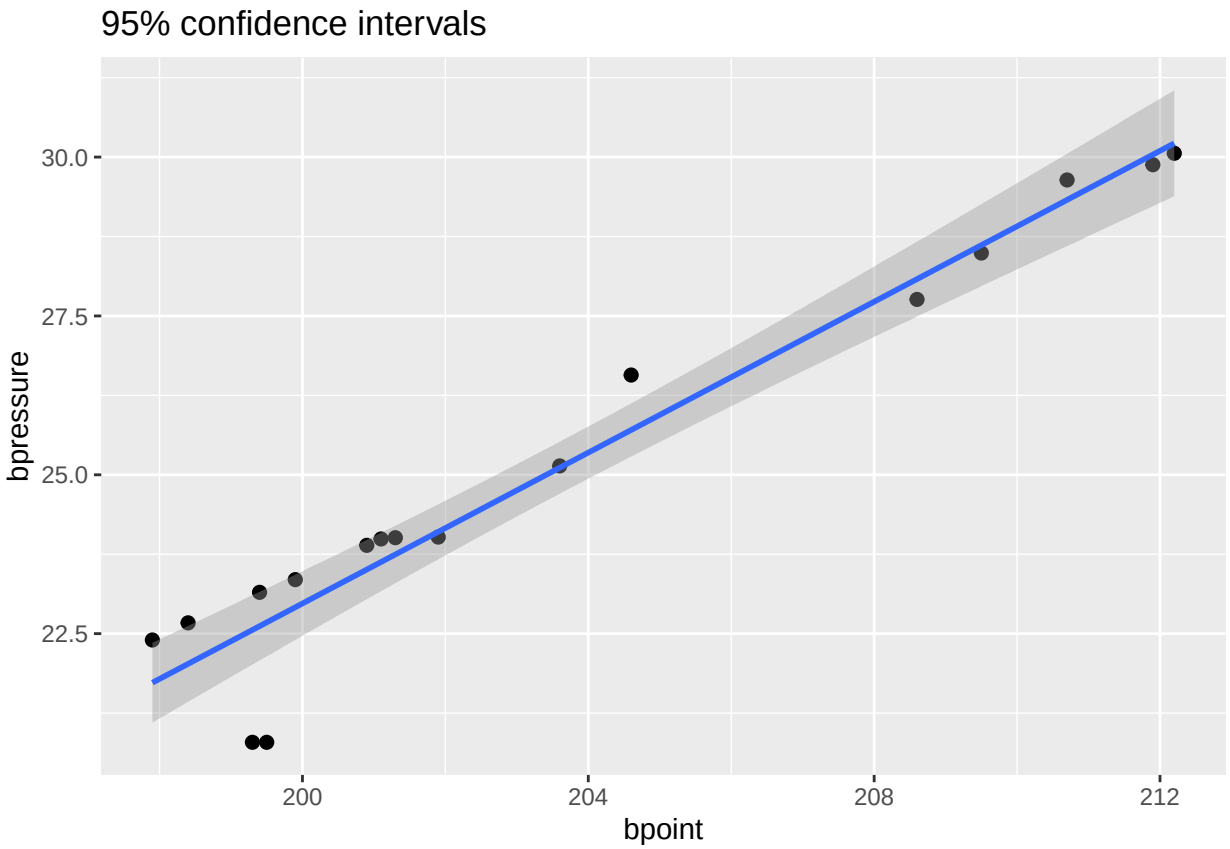
95% Interval Estimation

```
cls_16<-predict(model_16 , interval="confidence" , level =0.95)
cls_16<-data.frame(cls_16) # Prediction bands in DataFrame
cls_16
```

```
##         fit      lwr      upr
## 1  2083.892 2083.394 2084.389
## 2  2084.318 2083.821 2084.815
## 3  2273.541 2273.098 2273.984
## 4  2273.115 2272.672 2273.558
## 5  2273.541 2273.098 2273.984
## 6  2462.764 2462.373 2463.156
## 7  2462.338 2461.946 2462.730
## 8  2651.562 2651.216 2651.908
## 9  2651.988 2651.642 2652.334
## 10 2651.988 2651.642 2652.334
## 11 2842.916 2842.609 2843.223
## 12 2842.916 2842.609 2843.223
## 13 3029.582 3029.303 3029.861
## 14 3029.156 3028.877 3029.435
## 15 3029.582 3029.303 3029.861
## 16 3220.510 3220.245 3220.775
## 17 3221.363 3221.098 3221.627
## 18 3409.307 3409.041 3409.574
## 19 3409.733 3409.467 3410.000
## 20 3599.809 3599.525 3600.093
## 21 3600.662 3600.377 3600.946
## 22 3787.754 3787.440 3788.068
## 23 3788.606 3788.292 3788.921
## 24 3789.032 3788.718 3789.347
## 25 3979.961 3979.605 3980.316
## 26 3979.961 3979.605 3980.316
## 27 4165.774 4165.373 4166.176
## 28 4167.053 4166.651 4167.455
## 29 4167.053 4166.651 4167.455
## 30 4358.407 4357.953 4358.862
## 31 4359.260 4358.805 4359.714
## 32 4546.352 4545.843 4546.861
## 33 4547.204 4546.695 4547.714
```

```
ggplot(df_17 , aes(x =bpoint , y = bpressure))+ geom_point(size = 2) + geom_smooth(method=lm , se=TRUE)
labs(title="95% confidence intervals", x=" bpoint ", y = "bpressure")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



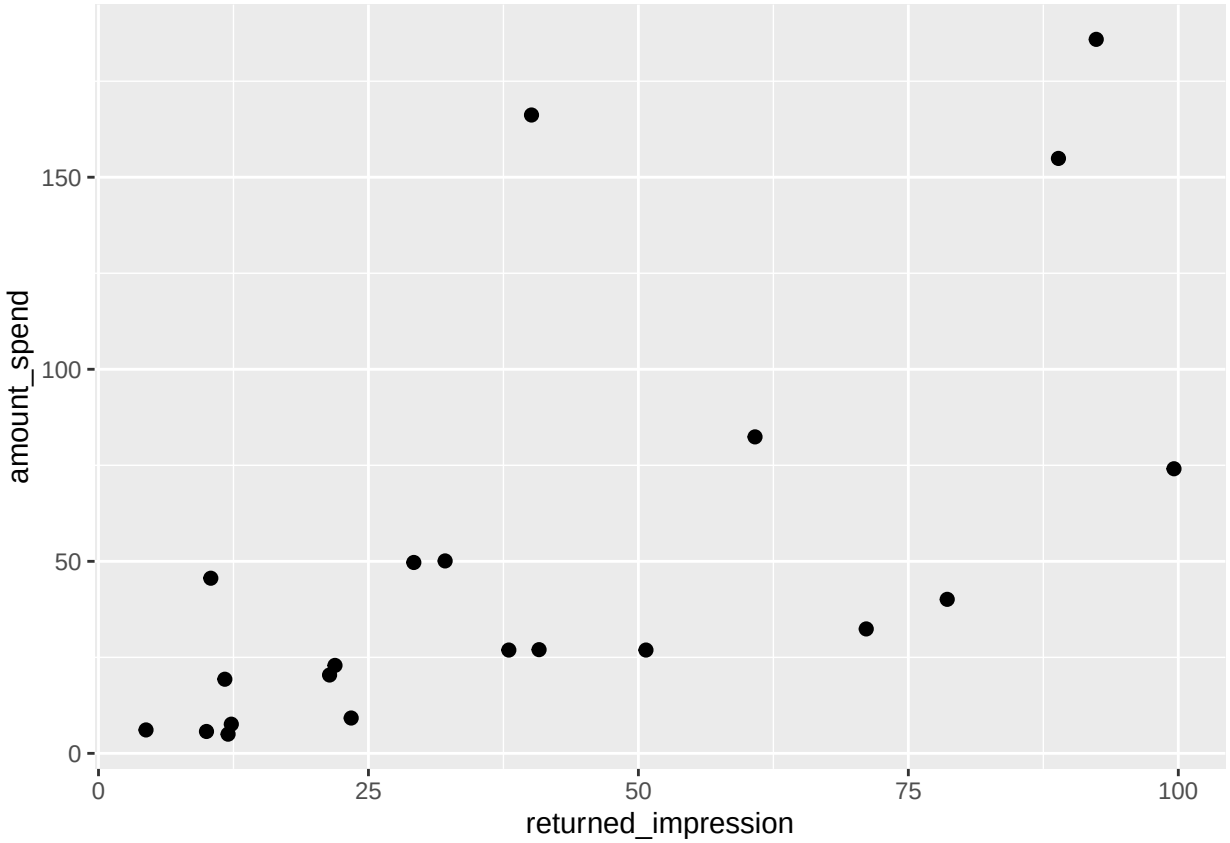
Question 2.18

Input Data

```
amount_spend <- c(50.1, 74.1, 19.3,22.9,82.4,40.1,185.9,26.9,20.4,166.2,27,45.6,154.9,5,49.7,26.9,5.7,7
 )
returned_impression <-c(32.1, 99.6, 11.7,21.9,60.8,78.6,92.4,50.7,21.4,40.1,40.8,10.4,88.9,12,29.2,38,1
firms <-c("Miller Lite","Pepsi","Stroh ","Federal Express","Burger King","Coca-Cola","McDonalds","MCI",
)
df_18 <- data.frame(amount_spend , returned_impression)
```

Scatter plot

```
ggplot(df_18 , aes(x =returned_impression , y = amount_spend))+ geom_point(size = 2)
```



a. Fit simple linear regression model

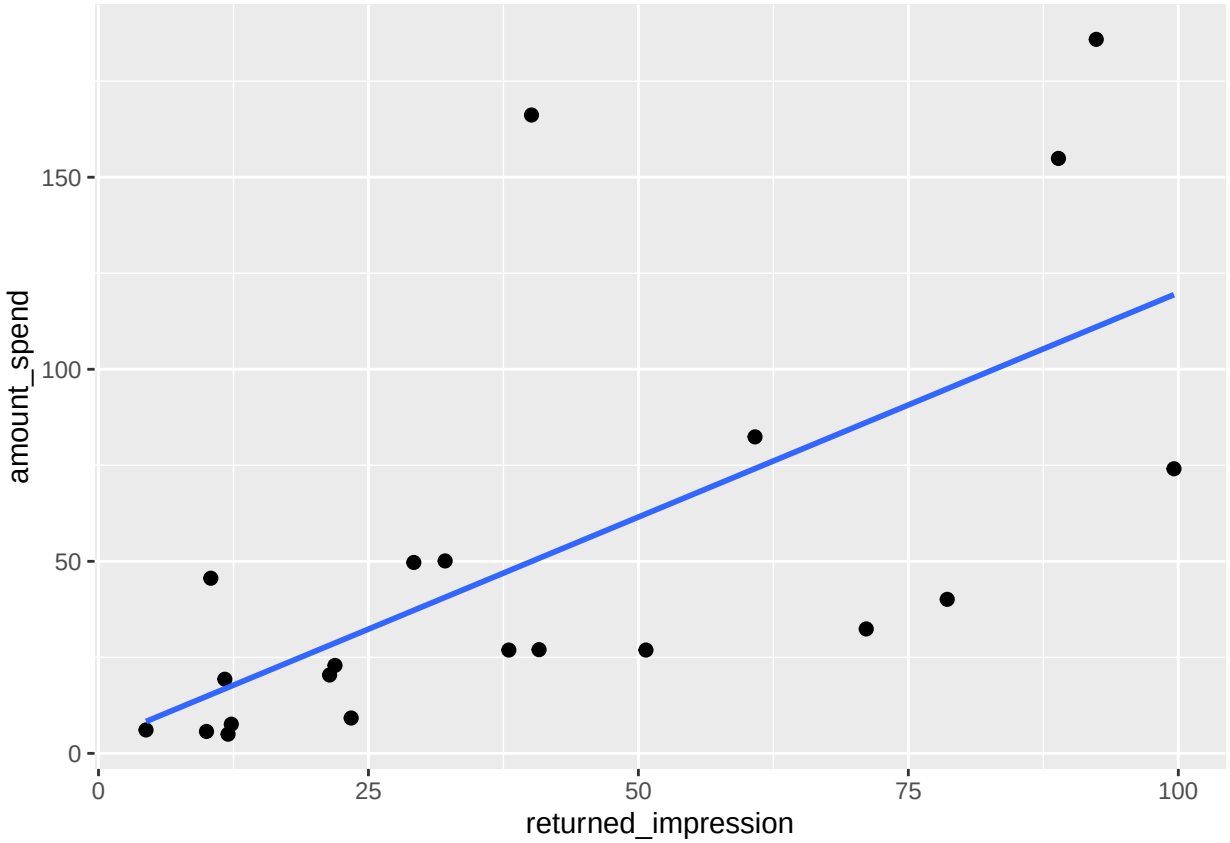
```
model_18 <- lm(amount_spend ~ returned_impression)
coefficients(model_18)
```

```
##         (Intercept) returned_impression
##            3.161273            1.167349
```

Plot Regression line

```
ggplot(df_18 , aes(x =returned_impression , y = amount_spend))+ geom_point(size = 2)  + geom_smooth(meth
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



b. Check is their significant relation

H0: B1 = 0

H1: B1 !=0

```
summary.aov(model_18)
```

```
##                     Df Sum Sq Mean Sq F value  Pr(>F)
## returned_impression  1  24825   24825   13.98 0.00139 **
## Residuals           19  33731    1775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
df(.001 , 1, 19)
```

```
## [1] 12.44428
```

So there is significant relation. thus reject null hypothesis.

R2 value

```
summary(model_18)$r.squared
```

```
## [1] 0.4239509
```

Model explain 42.3% of variation. so there is still a lot of unexplained variation in this model.

c. Construct 95% confidence interval

```
cls_18 <- predict(model_18 , interval = "confidence" , level =0.95)
cls_18<-data.frame(cls_18)
cls_18$amount_spend <- amount_spend
cls_18$firms <- firms
```

```
ggplot(df_18 , aes(x =returned_impression , y = amount_spend))+ geom_point(size = 2)  + geom_smooth(metl
labs(title="95% confidence intervals", x=" returned_impression ", y = "amount_spend")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```