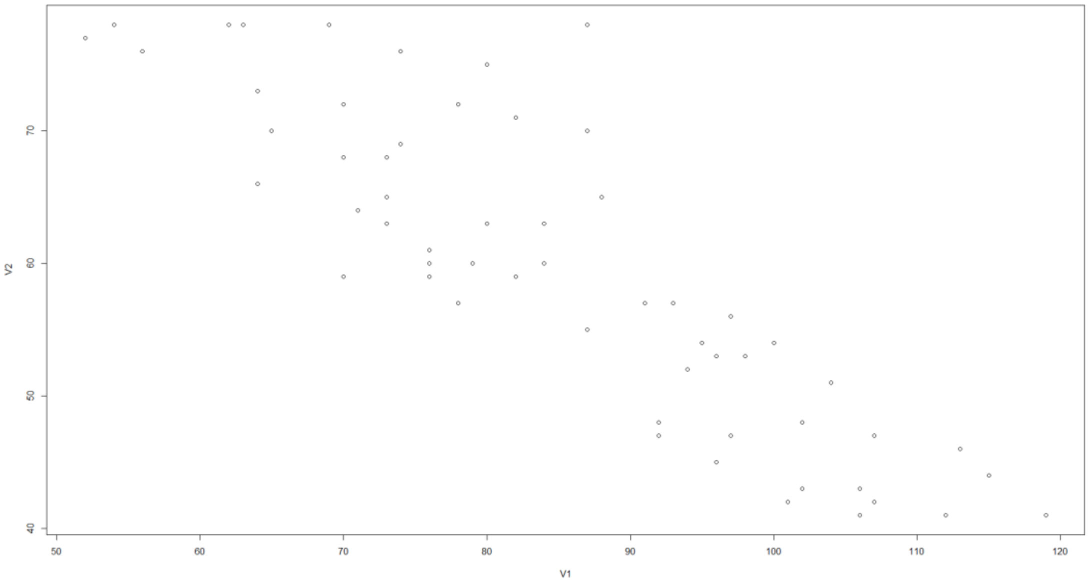
#########Example 2: mass and age data############################

```
## Read in data
ex.data <- read.table(file="W:/teaching/stat440540/data/CH1/CH01PR27.txt")
```

take a look at how data looks

```
> head(ex.data)
   V1 V2
1 106 43
2 106 41
3  97 47
4 113 46
5  96 45
6 119 41

plot(ex.data)
```



give variables V1 V2 names

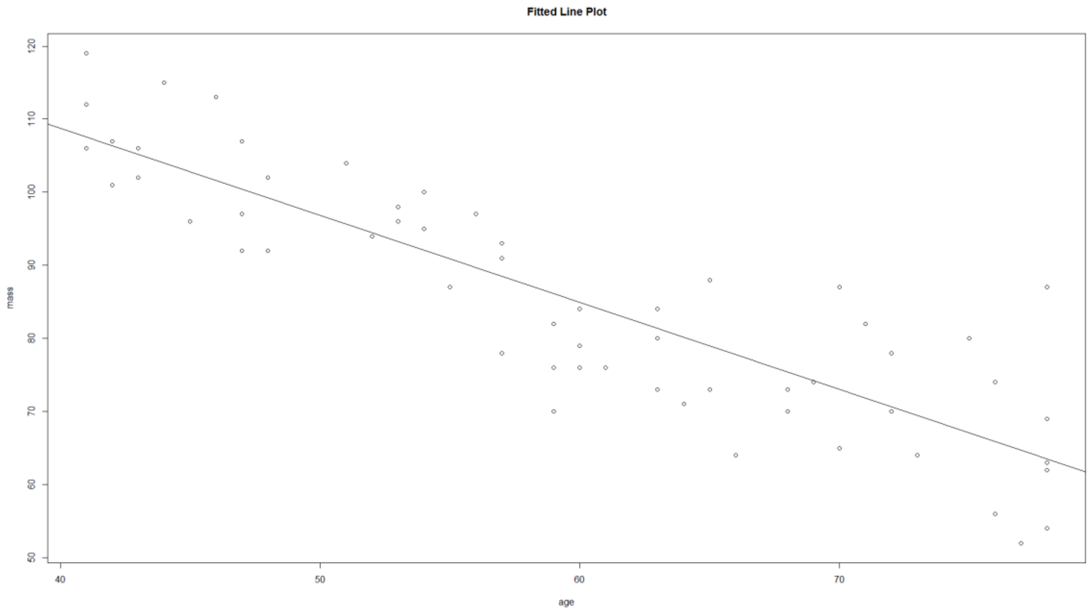
```
names(ex.data)[1]<-"mass"
names(ex.data)[2]<-"age"
```

```
> ## How many observations
> n <- nrow(ex.data)
> n
[1] 60

> ## Fit the estimated regression Line
> myfit <- lm(mass~age, data=ex.data)  ## Fit the model 'mass' = a + b*'age'
> myfit

Call:
lm(formula = mass ~ age, data = ex.data)

Coefficients:
```

```
(Intercept)         age
     156.35        -1.19
```

the estimated regression line is \hat y = 156.35 –1.19 x

```
> ## Assign the estimates to b_0 and b_1
> b_0 <- myfit$coef[1]
> b_1 <- myfit$coef[2]
>
>
> ## Plot the data with the fitted LS line
> plot(ex.data$age, ex.data$mass, xlab="age", ylab="mass", main="Fitted Line Plot")
> abline(b_0, b_1)
>
```




```
> ## 95% confidence intervals for beta_0 and beta_1
> confint(myfit, level=.95) #change level=0.9 to get the 90% CI
                  2.5 %          97.5 %
(Intercept) 145.312572    167.380556
age           -1.370545      -1.009446
>
```

```
>> ######################################################
> ## Hypothesis tests for
> ##   H_0: beta_0=0 vs. H_a: beta_0 != 0 and
> ##   H_0: beta_1=0 vs. H_a: beta_1 != 0
> summary(myfit, level=.95)

Call:
lm(formula = mass ~ age, data = ex.data)

Coefficients:
             Estimate   Std. Error   t value   Pr(>|t|)
(Intercept) 156.3466     5.5123      28.36     <2e-16 ***
age          -1.1900     0.0902     -13.19     <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.173 on 58 degrees of freedom
Multiple R-squared:  0.7501,    Adjusted R-squared:  0.7458
F-statistic: 174.1 on 1 and 58 DF,  p-value: < 2.2e-16
```

t-value –13.19 = -1.1900/0.0902, pvalue < <<0, reject null hypothesis that H_0: beta_1=0

F-statistic = 174.1 = (-13.19)^2, reject H_0: beta_1=0

R-squared 0.7501 = SSR/SSTO = 11627.5/(11627.5+3874.4)

```
> ## Make an ANOVA table and F-test
> anova(myfit)
Analysis of Variance Table

Response: mass
           Df    Sum Sq   Mean Sq   F value     Pr(>F)
age         1   11627.5   11627.5   174.06    < 2.2e-16 ***
Residuals 58     3874.4      66.8
```

F value 174.06 = MSR/MSE = 11627.5/66.8

```
> ## 95% CI for mean response when X=50
> newdata <- data.frame(age=50)
> predict(myfit, newdata, interval="confidence", level=.95)
       fit        lwr        upr
1  96.84679   94.0701   99.62348
> ## 95% PI for new observation when X=50
> newdata <- data.frame(age=50)
> predict(myfit, newdata, interval="prediction", level=.95)
       fit        lwr        upr
1 96.84679   80.25244   113.4411
```

Prediction interval is wider than the confidence interval, since we also need to account the variability of y for the new observation

```
> ######################General approach to testing in regression########
```

full model

```
> myfit <- lm(mass~age, data=ex.data)
```

reduced model

```
> myfitr <- lm(mass~1, data=ex.data)
> anova(myfit)
Analysis of Variance Table

Response: mass
          Df  Sum Sq  Mean Sq  F value    Pr(>F)
age        1 11627.5  11627.5   174.06 < 2.2e-16 ***
Residuals 58  3874.4     66.8

> anova(myfitr)
Analysis of Variance Table

Response: mass
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 59  15502  262.75

> anova(myfit,myfitr)
Analysis of Variance Table

Model 1: mass ~ age
Model 2: mass ~ 1
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     58  3874.4
2     59 15501.9 -1    -11628 174.06 < 2.2e-16 ***
```

SSE(R)-SSE(F)= 15502-3874.4=11627.6,
MSE(F)=66.8, F= 11627.6/66.8= 174, P-value=<<0,
reject the reduced model, conclude that the reduced model is not appropriate.