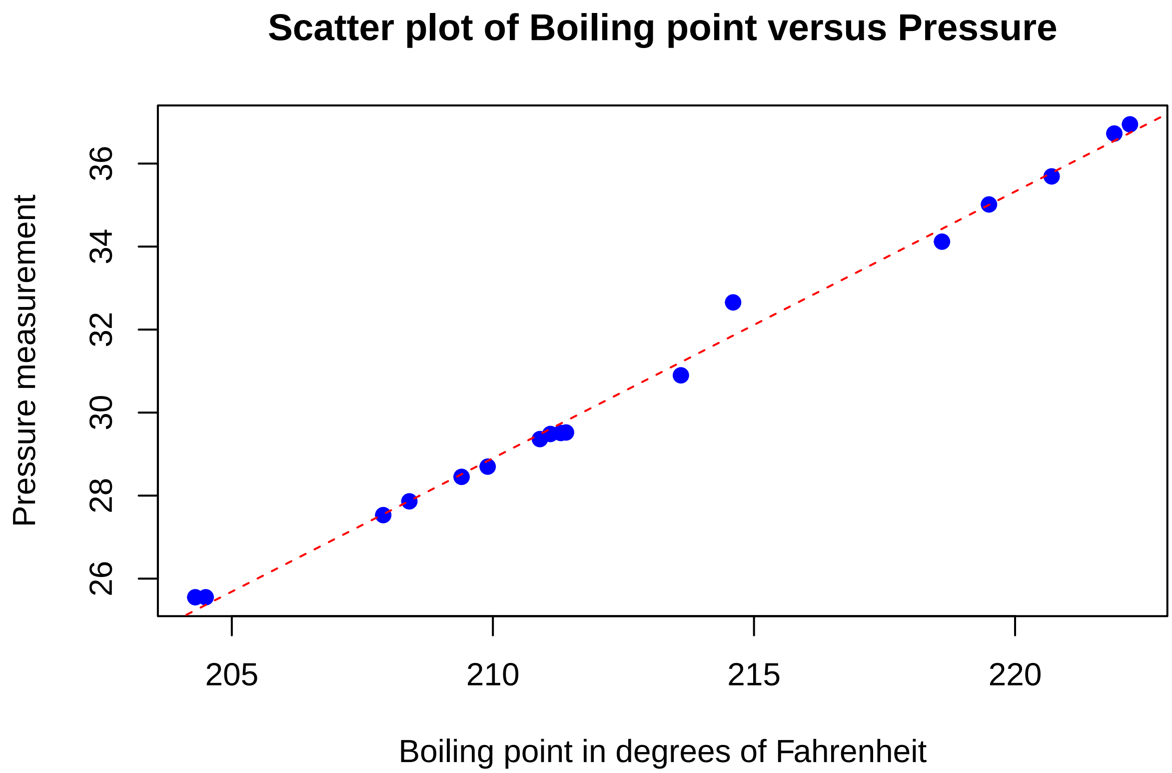
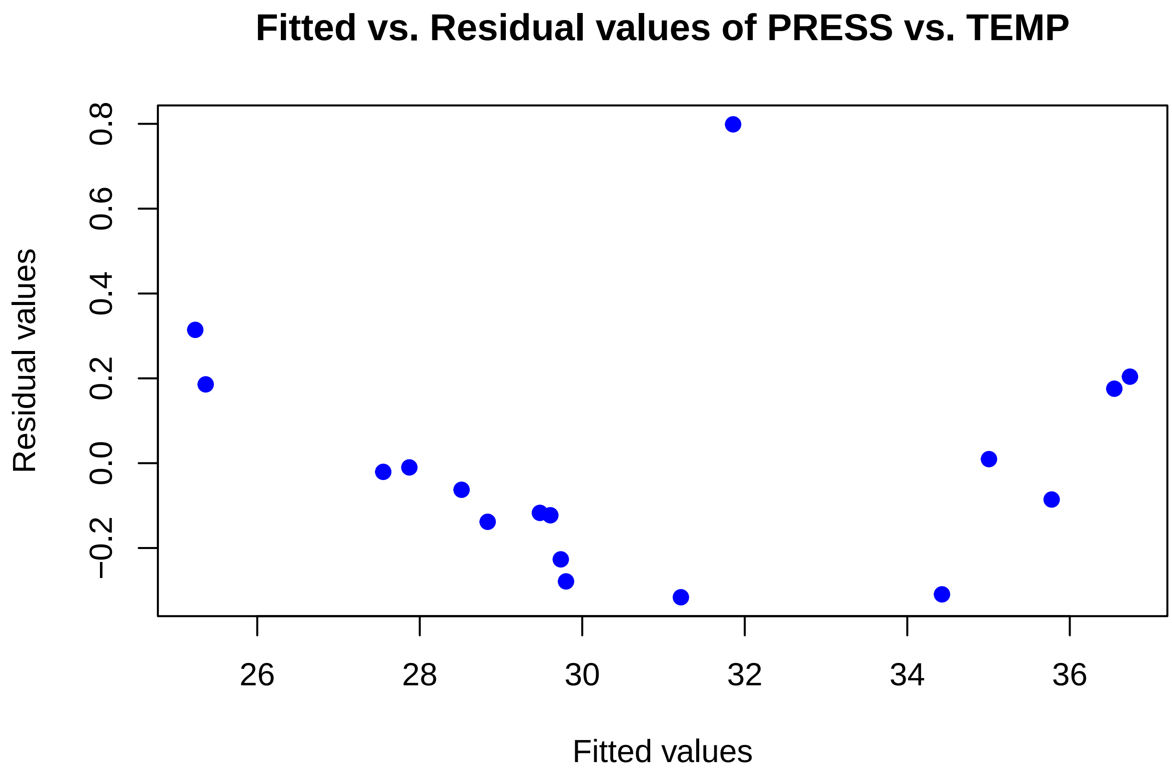
```r
#knitr::opts_chunk$set(echo = TRUE, tidy.opts=list(width.cutoff=80), tidy=TRUE)
knitr::opts_chunk$set(echo = TRUE)

forbes<-read.table("Forbes.txt",header=T)

#forbes # displays datatable

attach(forbes)
```

```
## The following objects are masked from forbes (pos = 4):
##
##     PRESS, TEMP

## The following objects are masked from forbes (pos = 28):
##
##     PRESS, TEMP
```

```r
n<-length(TEMP)

plot(TEMP,PRESS,pch=19,col="blue",xlab="Boiling point in degrees of
↪  Fahrenheit",ylab="Pressure measurement",main="Scatter plot of Boiling point versus
↪  Pressure")

fit<-lm(PRESS~TEMP) # least squares estimate

#summary(fit)
#fit$coef

#abline(a=fit$coef[1],b=fit$coef[2],lty=2,col="red")
abline(fit,lty=2,col="red")
```

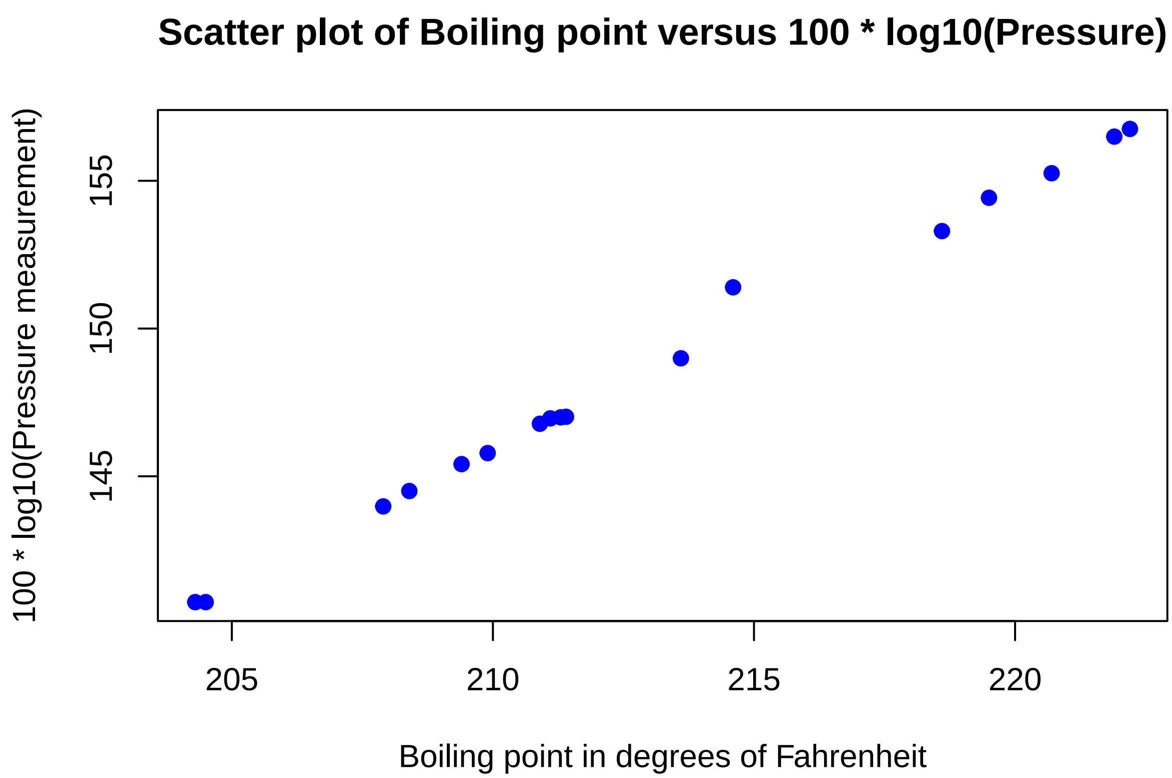



```
#fit$fit
#fit$res

plot(fit$fit,fit$res,pch=19,col="blue",xlab="Fitted values",ylab="Residual
↪  values",main="Fitted vs. Residual values of PRESS vs. TEMP")
```




```
# My observations for the residuals verus fitted values graph is that the data from
↪  Forbes.txt is fitted very well to a linear regression. This is because the residual
↪  values all seem to be less than +-0.5. This would mean that all the pressure
↪  measurements are within a range of +-0.5 to their fitted value, which suggests that a
↪  linear regression is a very good fit for the data. There is only 1 outlier of the 17
↪  data points which has a residual of +-1.0.
```

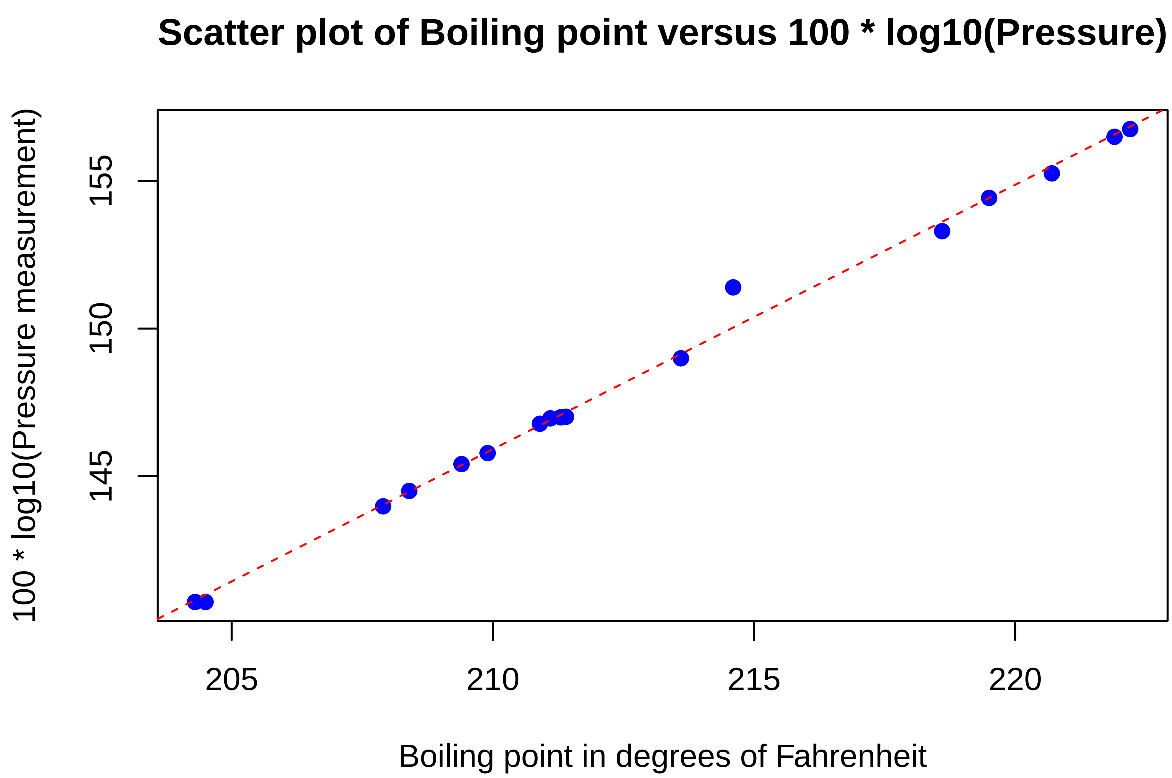
```
#PRESS2<-data.frame(PRESS)
PRESS_B=100*log10(PRESS)

#PRESS_B

plot(TEMP,PRESS_B,pch=19,col="blue",xlab="Boiling point in degrees of
↪  Fahrenheit",ylab="100 * log10(Pressure measurement)",main="Scatter plot of Boiling
↪  point versus 100 * log10(Pressure)")
```




```r
fit_C<-lm(PRESS_B~TEMP)

# Computed estimates of parameters:
fit_C$coef
```

```
## (Intercept)        TEMP
## -42.1641169   0.8956176
```

```r
# Fitted line:
# Y_^i = 0.8956176*X_i - 42.1641
plot(TEMP,PRESS_B,pch=19,col="blue",xlab="Boiling point in degrees of
    Fahrenheit",ylab="100 * log10(Pressure measurement)",main="Scatter plot of Boiling
    point versus 100 * log10(Pressure)") # need to plot before adding fitted line
abline(fit_C,lty=2,col="red")
```




```r
# ANOVA Table:
anova(fit_C)
```

```
## Analysis of Variance Table
##
## Response: PRESS_B
##           Df Sum Sq Mean Sq F value    Pr(>F)
## TEMP       1 425.76  425.76  2961.6 < 2.2e-16 ***
## Residuals 15   2.16    0.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# R^2 = 1 - (RSS/TSS)
#RSS - residual sum of squares
RSS_C = sum(fit_C$res^2)
#TSS - total sum of squares
TSS_C = sum((PRESS_B - (sum(PRESS_B)/n))^2)

R_2 = 1 - (RSS_C/TSS_C)
# R^2:
R_2
```

```
## [1] 0.9949606
```

```r
# b1 and b0 are estimators and have ^ hats

# fit_C$coef[2] = b1
# fit_C$coef[1] = b0

# CI = b +- t_(1-alpha/2,df=n-2) * standard_error(b)

# 95% Confidence interval for b1 and b0 of 100×log(PRESS) on TEMP

# 95% CI for b1
b1_95CI = c(fit_C$coef[2]-qt(1-0.05/2,n-2)*summary(fit_C)$coef[4],+
↪  fit_C$coef[2]+qt(1-0.05/2,n-2)*summary(fit_C)$coef[4])
b1_95CI
```

```
##      TEMP      TEMP
## 0.8605393 0.9306958
```

```r
# 95% CI for b0

b0_95CI = c(fit_C$coef[1]-qt(1-0.05/2,n-2)*summary(fit_C)$coef[3],+
↪  fit_C$coef[1]+qt(1-0.05/2,n-2)*summary(fit_C)$coef[3])
b0_95CI
```

```
## (Intercept) (Intercept)
##   -49.63670   -34.69153
```

```r
# Fitted value at 200F for 100×log(PRESS)
x_E<-200
fit_C$coef[1]+fit_C$coef[2]*x_E
```

```
## (Intercept)
##    136.9594
```

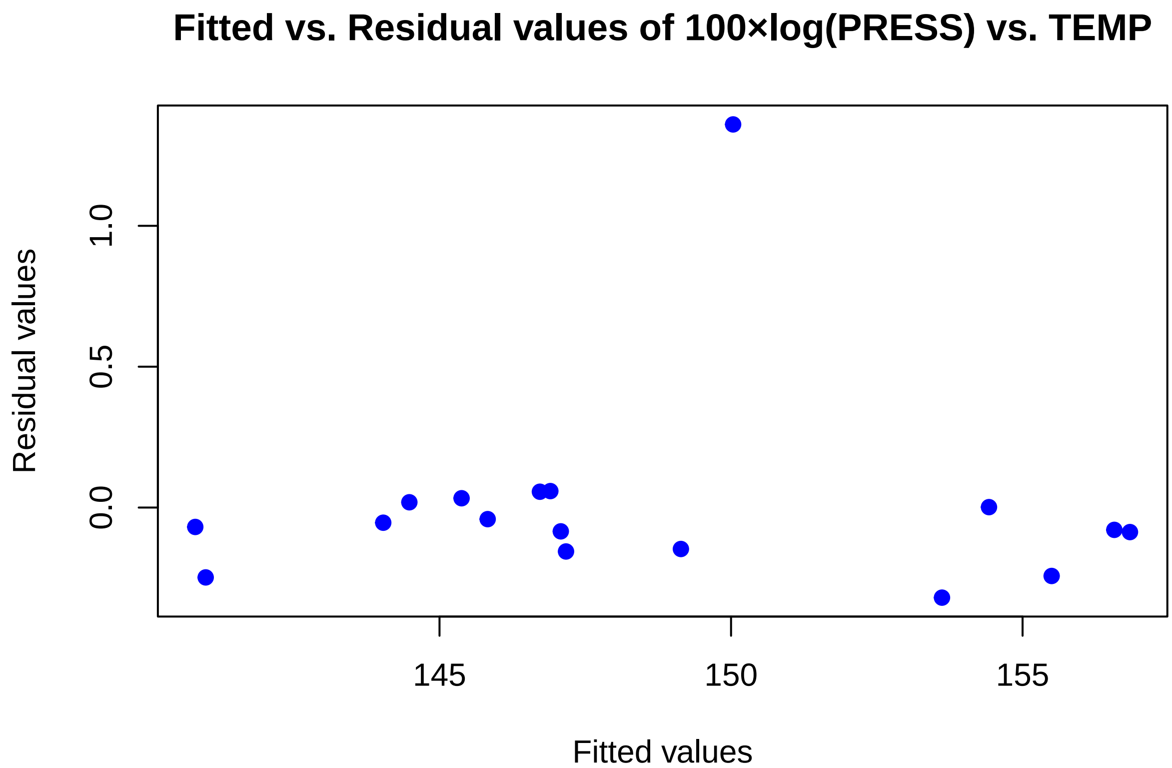
```r
# Prediction at 200F for 100×log(PRESS) at 99% prediction interval
newdata_E = data.frame(TEMP=200)
predict(fit_C, newdata_E, interval="predict", level = 0.99)
```

```
##        fit      lwr      upr
## 1 136.9594 135.6493 138.2695
```

```r
# 99% CI for 100×log(PRESS) for the fitted value at 200F
predict(fit_C, newdata_E, interval="confidence", level = 0.99)
```

```
##        fit      lwr      upr
## 1 136.9594 136.2753 137.6435
```

```r
plot(fit_C$fit,fit_C$res,pch=19,col="blue",xlab="Fitted values",ylab="Residual
↪  values",main="Fitted vs. Residual values of 100×log(PRESS) vs. TEMP")
```




```
# My conclusion is that the residuals versus fitted values is fitted extremely well; the
  residual plot of 100×log(PRESS) vs. TEMP is a better fit than the residual plot of
  PRESS vs. TEMP. All the residuals are mostly +-0.25, which is extremely well fitted.
  The linear regression line will fit the data from 100×log(PRESS) vs. TEMP extremely
  well. The only thing to note is the 1 outlier which has a residual of almost +-1.5.
  This outlier is even more of an outlier than that from PRESS vs. TEMP, but overall,
  the data from 100×log(PRESS) vs. TEMP is more fitted to the linear regression
  (because there are 16 other points that were almost 2x more better fitted and the 1
  outlier that is about 1.5x worse fitted).
```