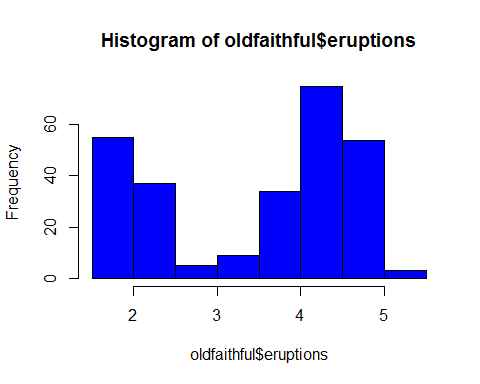
Oldfaithful Gyser

**Downloading data** Oldfaithful is one of famouse gyser in Yellowstone national park.

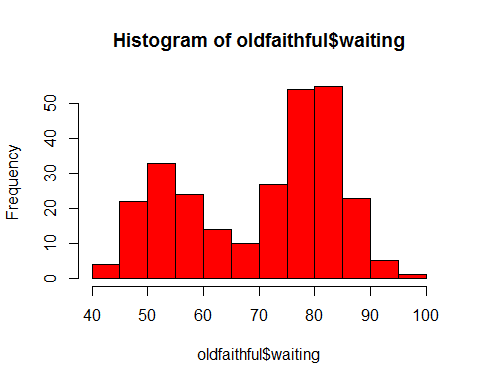
oldfaithful=read.table("C:/Users/chink/Documents/oldfaithful.txt",na.strings = T,header = F)

#putting colun names   
  
colnames(oldfaithful)=c("sn.no","eruptions","waiting")

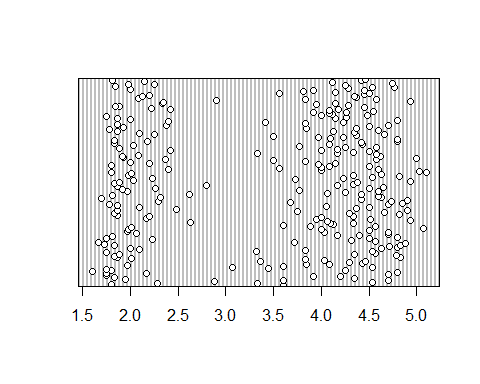
#Histogram of x=Eruption  
hist(oldfaithful$eruptions,col="blue")

 From the histogram of the eruptions, can conclude that it is bimodal distribution.

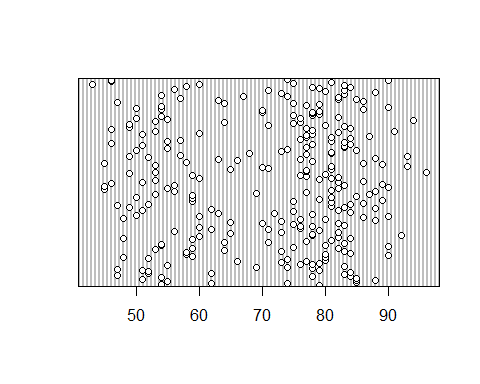
#histogram of x=waiting  
hist(oldfaithful$waiting,col="red")

 It is bimodal histogram.

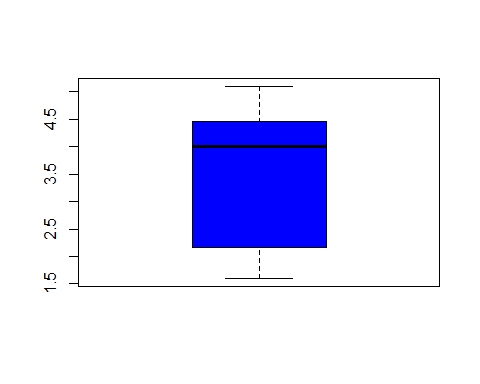
#Dot plot of x=eruptions  
dotchart(oldfaithful$eruptions)



#Dotplot of x=waiting  
dotchart(oldfaithful$ waiting)



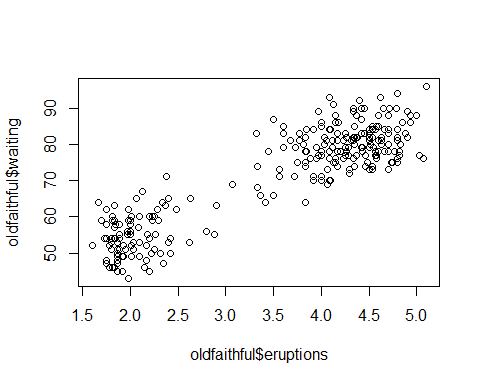
#Boxplot of x=eruptions  
boxplot(oldfaithful$eruptions,col="blue")



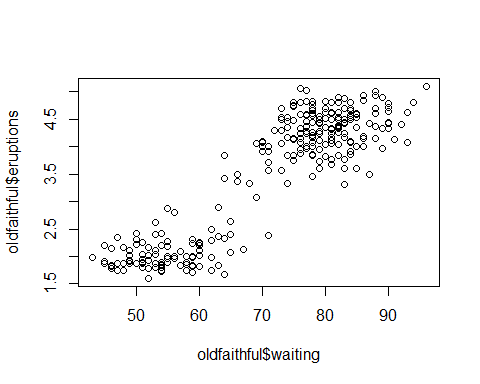
library(ggplot2)

#ggplot(data=oldfaithful,aes(x="",y=eruptions))+geom\_boxplot(color="blue")+scale\_x\_discrete(names="predictor")+scale\_y\_continuous(name="Eruption time in mint")+geom\_jitter(width = 0.2,color="Red")

#Scatterplot for x and y  
plot(oldfaithful$waiting~oldfaithful$eruptions)



# Scatterplot of y and x  
plot(oldfaithful$eruptions~oldfaithful$waiting)



#Correlation test  
cor(oldfaithful$waiting,oldfaithful$eruptions)

## [1] 0.9008112

cor.test(oldfaithful$waiting,oldfaithful$eruptions)

##   
## Pearson's product-moment correlation  
##   
## data: oldfaithful$waiting and oldfaithful$eruptions  
## t = 34.089, df = 270, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.8756964 0.9210652  
## sample estimates:  
## cor   
## 0.9008112

waiting time and eruption are 90% correlated.

#Correlation test  
cor(oldfaithful$eruptions,oldfaithful$waiting)

## [1] 0.9008112

# chossing x and y and running regression model  
# x=eruptions y=waiting   
oldfaithful\_lm=lm(oldfaithful$waiting~oldfaithful$eruptions)  
summary(oldfaithful\_lm)

##   
## Call:  
## lm(formula = oldfaithful$waiting ~ oldfaithful$eruptions)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.0796 -4.4831 0.2122 3.9246 15.9719   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 33.4744 1.1549 28.98 <2e-16 \*\*\*  
## oldfaithful$eruptions 10.7296 0.3148 34.09 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.914 on 270 degrees of freedom  
## Multiple R-squared: 0.8115, Adjusted R-squared: 0.8108   
## F-statistic: 1162 on 1 and 270 DF, p-value: < 2.2e-16

#Confidence interval  
confint(oldfaithful\_lm,level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) 31.20069 35.74810  
## oldfaithful$eruptions 10.10996 11.34932

confidence interval of intercept (31.20 ,35.74) and confidence interval of slope (10.10 ,11.34)

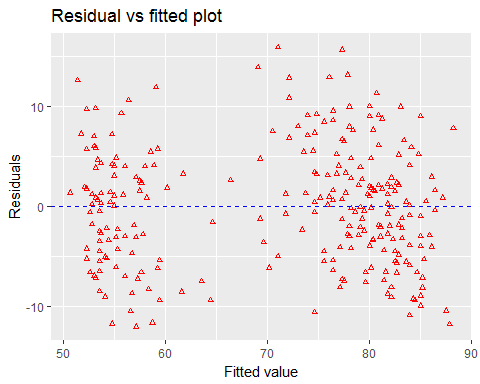
# ANOVA model  
anova(oldfaithful\_lm)

## Analysis of Variance Table  
##   
## Response: oldfaithful$waiting  
## Df Sum Sq Mean Sq F value Pr(>F)   
## oldfaithful$eruptions 1 40644 40644 1162.1 < 2.2e-16 \*\*\*  
## Residuals 270 9443 35   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#chossing x and y and running regression model  
# y=eruptions x=waiting   
oldfaithful\_lm1=lm(oldfaithful$eruptions~oldfaithful$waiting)  
summary(oldfaithful\_lm1)

##   
## Call:  
## lm(formula = oldfaithful$eruptions ~ oldfaithful$waiting)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.29917 -0.37689 0.03508 0.34909 1.19329   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.874016 0.160143 -11.70 <2e-16 \*\*\*  
## oldfaithful$waiting 0.075628 0.002219 34.09 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4965 on 270 degrees of freedom  
## Multiple R-squared: 0.8115, Adjusted R-squared: 0.8108   
## F-statistic: 1162 on 1 and 270 DF, p-value: < 2.2e-16

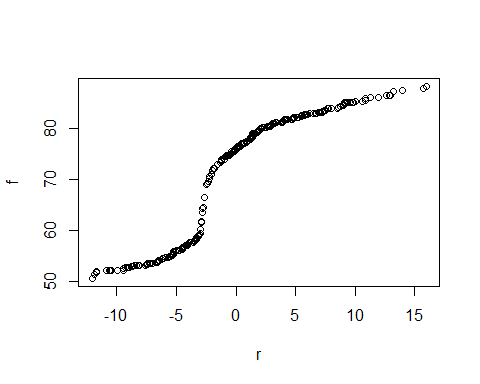
#Plot of residuals and fitted value   
ggplot(oldfaithful\_lm,aes(x=.fitted,y=.resid))+geom\_point(size=1,shape=2,col="red")+geom\_hline(yintercept = 0,col="blue",linetype="dashed")+xlab("Fitted value")+ylab("Residuals")+ggtitle("Residual vs fitted plot")



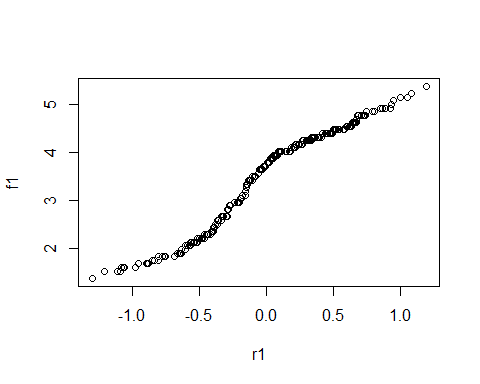
#Finding residuals   
r=resid(oldfaithful\_lm)  
r1=resid(oldfaithful\_lm1)

#Finding fitte value  
f=fitted.values(oldfaithful\_lm)  
f1=fitted.values(oldfaithful\_lm1)

#QQ plot for residuals   
qqplot(r,f)



#QQ plot  
qqplot(r1,f1)



#Shapiro test for normality  
shapiro.test(r)

##   
## Shapiro-Wilk normality test  
##   
## data: r  
## W = 0.98851, p-value = 0.02947

As p-value is 0.02 reject null hypothesis and conclude residuals are not normal.

#shapiro test for normality  
shapiro.test(r1)

##   
## Shapiro-Wilk normality test  
##   
## data: r1  
## W = 0.99278, p-value = 0.2106

Residuals are normal.

#Brown forsytha test for equal variance  
oldfaithful\_lm$groups=ifelse(oldfaithful$eruptions<3,c("1"),c("2"))  
oldfaithful\_lm$groups

## [1] "2" "1" "2" "1" "2" "1" "2" "2" "1" "2" "1" "2" "2" "1" "2" "1" "1"  
## [18] "2" "1" "2" "1" "1" "2" "2" "2" "2" "1" "2" "2" "2" "2" "2" "2" "2"  
## [35] "2" "1" "1" "2" "1" "2" "2" "1" "2" "1" "2" "2" "2" "1" "2" "1" "2"  
## [52] "2" "1" "2" "1" "2" "2" "1" "2" "2" "1" "2" "1" "2" "1" "2" "2" "2"  
## [69] "1" "2" "2" "1" "2" "2" "1" "2" "1" "2" "2" "2" "2" "2" "2" "1" "2"  
## [86] "2" "2" "2" "1" "2" "1" "2" "1" "2" "1" "2" "2" "2" "1" "2" "1" "2"  
## [103] "1" "2" "2" "1" "2" "1" "2" "2" "2" "1" "2" "2" "1" "2" "1" "2" "1"  
## [120] "2" "1" "2" "2" "1" "2" "2" "1" "2" "1" "2" "1" "2" "1" "2" "1" "2"  
## [137] "1" "2" "1" "2" "2" "1" "2" "2" "2" "1" "2" "1" "2" "1" "2" "2" "1"  
## [154] "2" "2" "2" "2" "2" "1" "2" "1" "2" "1" "2" "2" "2" "1" "2" "1" "2"  
## [171] "1" "1" "2" "2" "2" "2" "2" "1" "2" "2" "1" "2" "2" "2" "1" "2" "2"  
## [188] "1" "2" "1" "2" "1" "2" "2" "2" "2" "2" "2" "1" "2" "1" "2" "2" "1"  
## [205] "2" "1" "2" "2" "1" "2" "1" "2" "1" "2" "2" "2" "1" "2" "1" "2" "1"  
## [222] "2" "1" "2" "2" "2" "2" "2" "2" "2" "2" "1" "2" "1" "2" "1" "1" "2"  
## [239] "2" "1" "2" "1" "2" "1" "2" "2" "1" "2" "1" "2" "1" "2" "2" "2" "2"  
## [256] "2" "2" "2" "1" "2" "2" "2" "1" "2" "1" "1" "2" "2" "1" "2" "1" "2"

# install.packages("lawstat") if required  
library(lawstat)

#Equal variance test  
levene.test(r,oldfaithful\_lm$groups)

##   
## modified robust Brown-Forsythe Levene-type test based on the  
## absolute deviations from the median  
##   
## data: r  
## Test Statistic = 0.013577, p-value = 0.9073

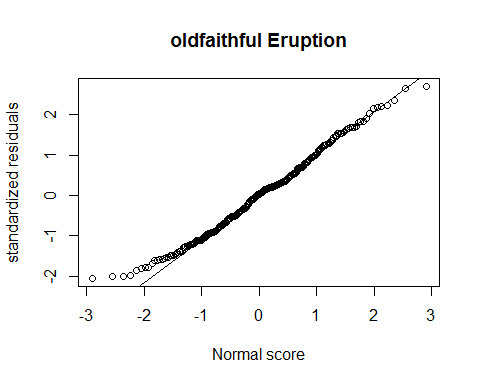
#install.packages("lmtest")  
library(lmtest)

# Test for constant variance, as p>0.05 then it has constant variance   
library(car)

ncvTest(oldfaithful\_lm)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 0.003226255 Df = 1 p = 0.9547044

# Standarized Residual plot   
oldfaithful\_s=rstandard(oldfaithful\_lm)  
qqnorm(oldfaithful\_s,ylab="standardized residuals",xlab="Normal score",main="oldfaithful Eruption")  
qqline(oldfaithful\_s)

 F test for goodness of fit.

#code for reduce model   
reduce=lm(oldfaithful$waiting~oldfaithful$eruptions)  
full=lm(oldfaithful$waiting~0+as.factor(oldfaithful$eruptions))  
anova(reduce,full)

## Analysis of Variance Table  
##   
## Model 1: oldfaithful$waiting ~ oldfaithful$eruptions  
## Model 2: oldfaithful$waiting ~ 0 + as.factor(oldfaithful$eruptions)  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 270 9443.4   
## 2 146 4052.9 124 5390.5 1.566 0.004635 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Box Cox in R

boxCox(oldfaithful$waiting~oldfaithful$eruptions,lambda = seq(-0.5,2,length=5))

