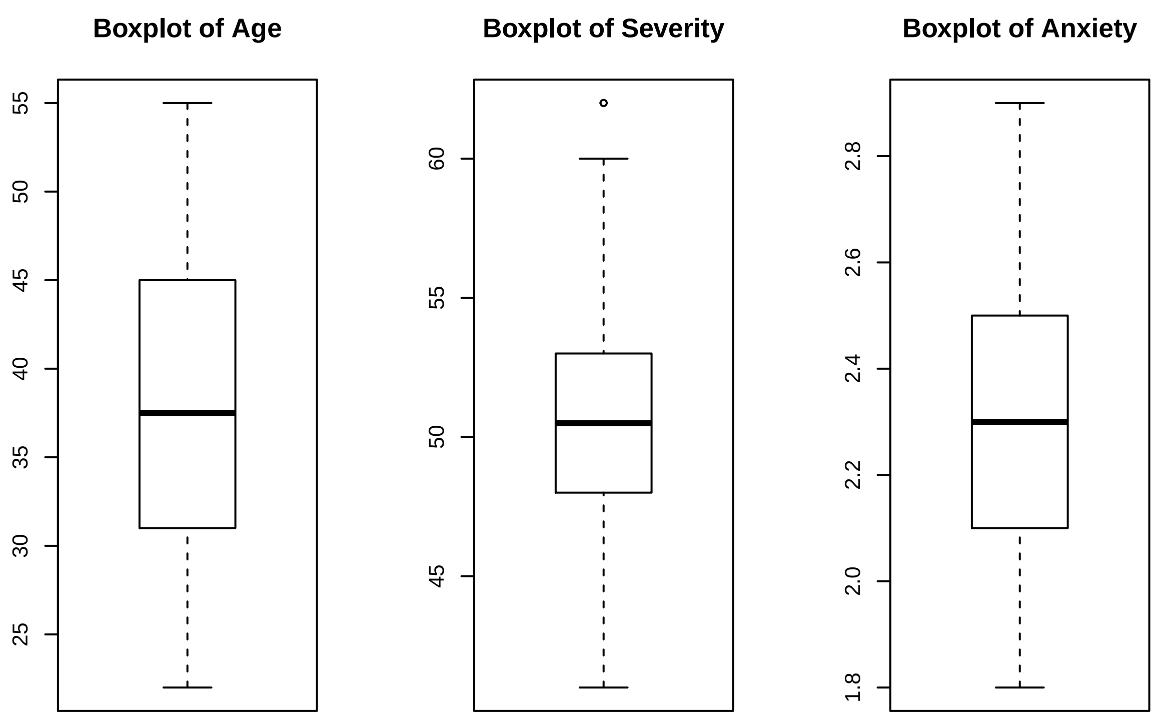
STAT GR5205 – Section 005 HW 5

Bo Rong br2498

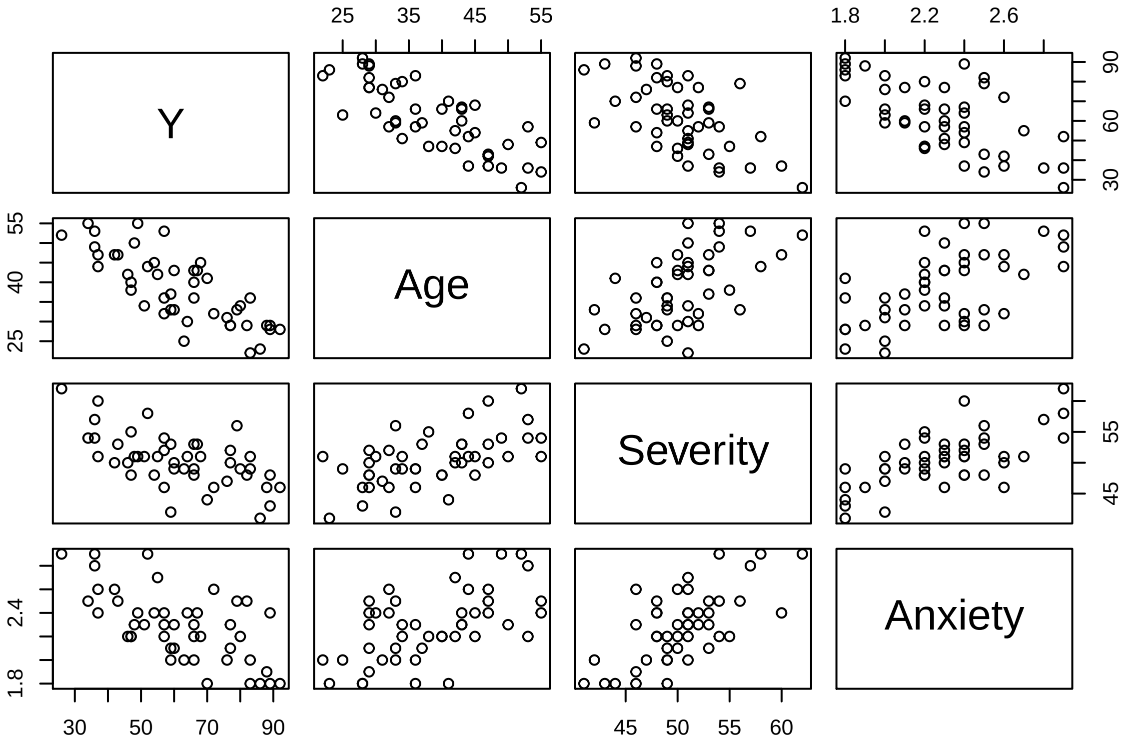
Nov. 6th, 2016

```
#1.
#(a)
filename <- "~/Downloads/patient_satisfaction.txt"
ps<- read.table(file=filename, header=T)
par(mfrow=c(1,3))
boxplot(ps$Age, main="Boxplot of Age")
boxplot(ps$Severity, main="Boxplot of Severity")
boxplot(ps$Anxiety, main="Boxplot of Anxiety")
```




```
#There is a outlier for severity.
```

```
#(b)
pairs(ps)
```



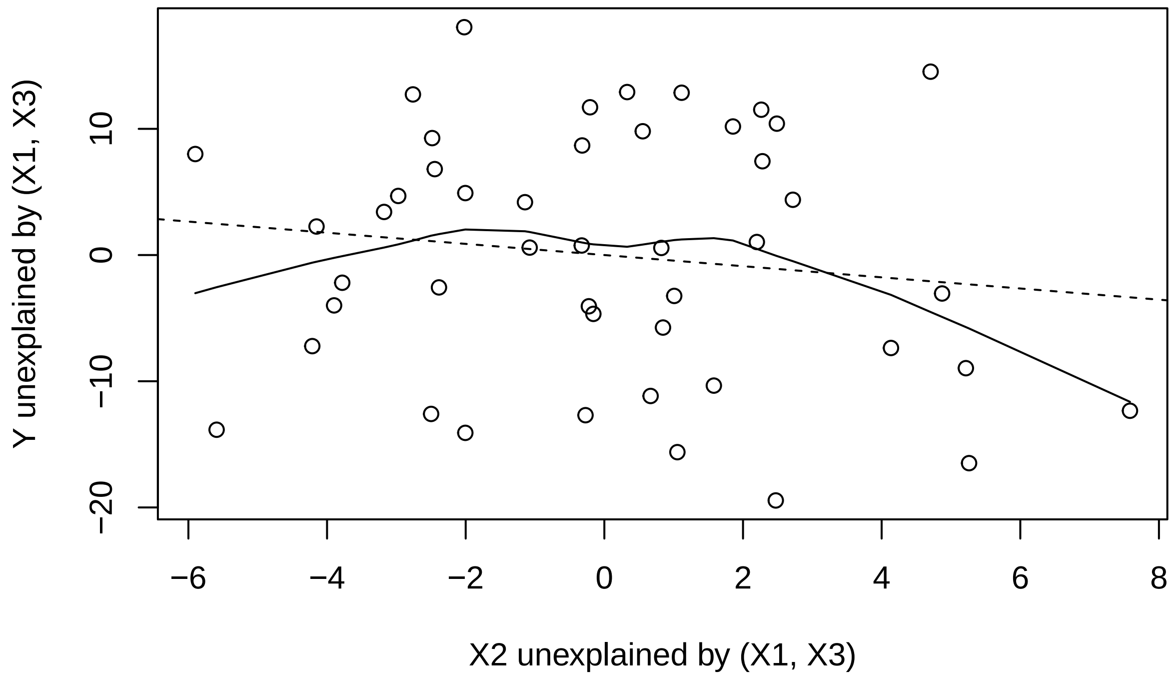

```
#Each of Age,Severity and Anxiety is negatively related to Y . The three predictor variables are all po
```

```
#(C)
fit<-lm(Y~Age+Severity+Anxiety,data=ps)
fit
```

```
##
## Call:
## lm(formula = Y ~ Age + Severity + Anxiety, data = ps)
##
## Coefficients:
## (Intercept)          Age     Severity      Anxiety
##     158.491       -1.142       -0.442      -13.470
```

```
#Yhat = 158.491 -1.142*Age - 0.442*Severity - 13.470*Anxiety, In this model, b2 = -0.442, which indicat
```

```
#(d)
x1 <-ps$Age
x2<-ps$Severity
x3<-ps$Anxiety
Y<-ps$Y
fitY.13 <- lm(Y ~ x1+x3, data=ps)
fit2.13 <- lm(x2 ~ x1+x3, data=ps)
plot(resid(fitY.13) ~ resid(fit2.13),
ylab="Y unexplained by (X1, X3)",
xlab="X2 unexplained by (X1, X3)")
lines(lowess(resid(fitY.13) ~ resid(fit2.13)))
abline(lm(resid(fitY.13) ~ resid(fit2.13)), lty=2)
```



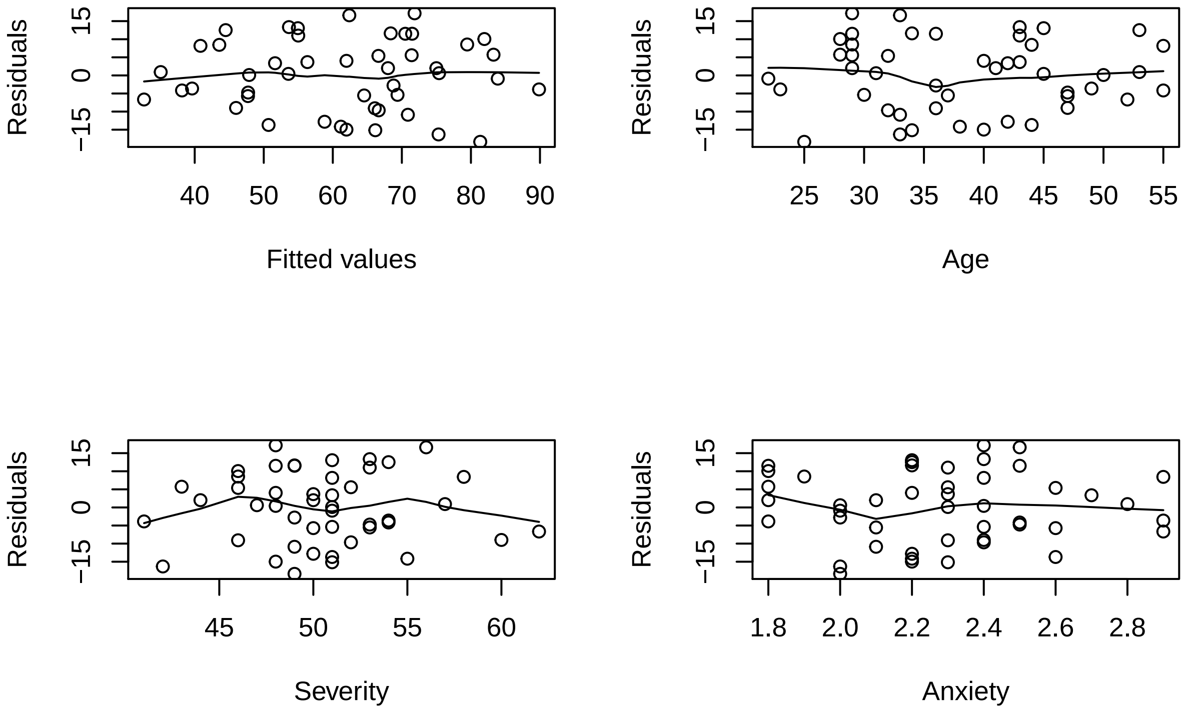

```r
#The slop of the relation between satisfaction unexplained by severity and severity unexplained by Age

coef(lm(resid(fitY.13) ~ resid(fit2.13)))
```

```
##    (Intercept) resid(fit2.13)
##   1.888373e-16  -4.420043e-01
```

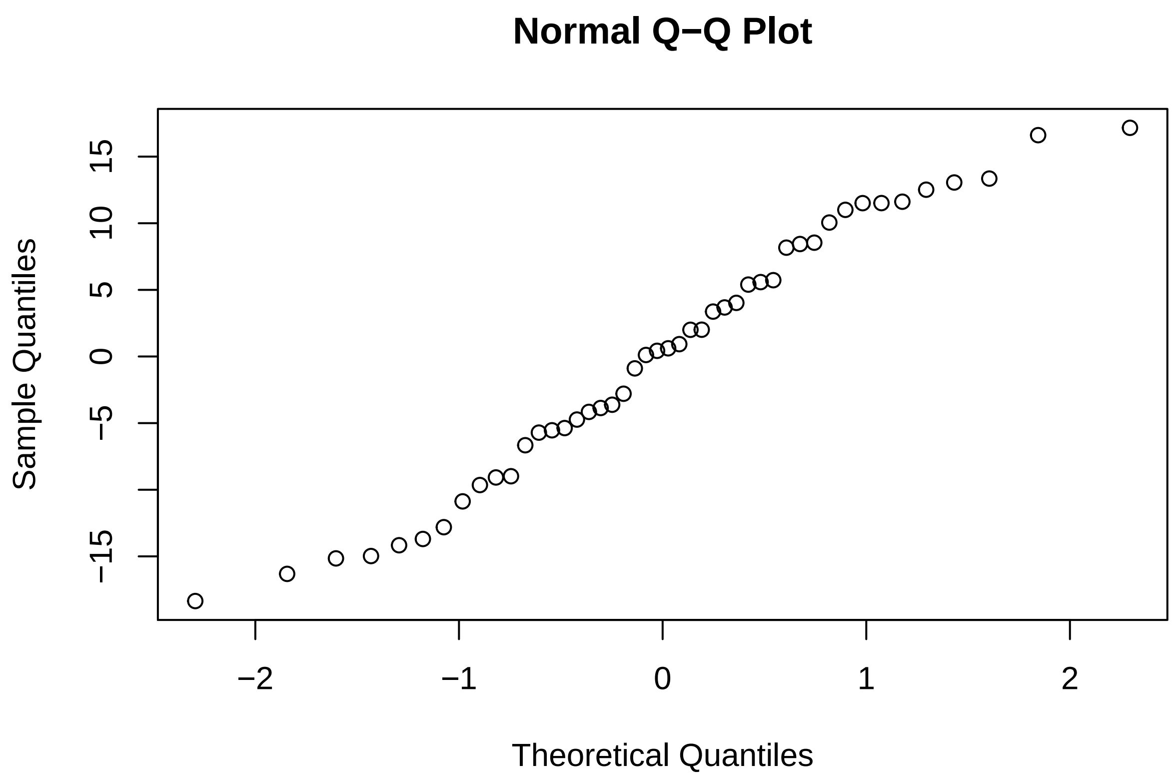
```r
#The slope is also -0.442, which is consistent with part(c).
```

```r
#(e)
par(mfrow=c(2,2))
plot(resid(fit)~fitted(fit),xlab="Fitted values",ylab="Residuals")
lines(lowess(resid(fit) ~ fitted(fit)))
plot(resid(fit) ~ x1, xlab="Age",ylab="Residuals")
lines(lowess(resid(fit) ~ x1))
plot(resid(fit) ~ x2,xlab="Severity", ylab="Residuals")
lines(lowess(resid(fit) ~ x2))
plot(resid(fit) ~ x3, xlab="Anxiety",ylab="Residuals")
lines(lowess(resid(fit) ~ x3))
```




```r
#We don't see violation against assumptions.
```

```r
#(f)
qqnorm(resid(fit))
```



```
#The plot is close to a straight line, so error terms appears to be reasonable.
```

```
#2.
#(a)
#H0:beta1=beta2=beta3=0
#Ha:not all betai=0
fit0<-lm(Y~1, data=ps)
fit123 <- lm(Y ~ x1 + x2 + x3)
anova(fit0, fit123)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ 1
## Model 2: Y ~ x1 + x2 + x3
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     45 13369.3
## 2     42  4248.8  3    9120.5 30.052 1.542e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#F=30.052,and the p-value is 1.542e-10, which is almost zero,So we reject H0.
```

```
#(b)
alpha <-0.10
g <- 3
1-alpha/g
```

```
## [1] 0.9666667
```

```
confint(fit123, level=1-alpha/g)[-1,]
```

```
##         1.67 %    98.33 %
## x1  -1.614248 -0.6689755
## x2  -1.524510  0.6405013
## x3 -29.092028  2.1517012
```

```
#The 90% confidence intervals: -1.614248 <= beta1 <= -0.6689755, -1.524510 <= beta2 <= 0.6405013
#and -29.092028 <= beta3 <= 2.1517012.
```

```
#(c)
summary(fit123)$r.squared
```

```
## [1] 0.6821943
```

```
#Thus 68% of all the Y's explained by x1, x2 and x3.
```

```
#(d)
x123 <-data.frame(x1=35, x2=45, x3=2.2)
predict(fit123, newdata=x123, interval="confidence")
```

```
##        fit      lwr      upr
## 1 69.01029 63.63288 74.38769
```

```r
#The 95% confidence interval:(63.63288,74.38769)
```

```r
#(e)
predict(fit123, newdata=x123, interval="prediction")
```

```
##        fit      lwr      upr
## 1 69.01029 48.01224 90.00833
```

```r
#The 95% confident interval:(48.01224,90.00833)
```

```r
#3.
#(a)
fit213<-lm(Y~x2 + x1 + x3)
anova(fit213)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x2         1 4860.3  4860.3 48.0439 1.822e-08 ***
## x1         1 3896.0  3896.0 38.5126 2.008e-07 ***
## x3         1  364.2   364.2  3.5997   0.06468 .
## Residuals 42 4248.8   101.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#(b)
#H0 :beta3 =0, Ha:beta3!=0
summary(fit213)
```

```
##
## Call:
## lm(formula = Y ~ x2 + x1 + x3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3524  -6.4230   0.5196   8.3715  17.1601
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
## x2           -0.4420     0.4920  -0.898   0.3741
## x1           -1.1416     0.2148  -5.315 3.81e-06 ***
## x3          -13.4702     7.0997  -1.897   0.0647 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.06 on 42 degrees of freedom
## Multiple R-squared:  0.6822, Adjusted R-squared:  0.6595
## F-statistic: 30.05 on 3 and 42 DF,  p-value: 1.542e-10
```

```
#t=-1.897,P-value=0.0647. And F=3.5997 ,P-value=0.0647. Thus F=3.5997= (-1.897)^2 = t^2.So these two te
```

```
#(c)
#H0:beta2 =beta3=0, Ha=not all beta2 or beta3 =0.
fit1 <- lm(Y ~ x1)
anova(fit1, fit123)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ x1
## Model 2: Y ~ x1 + x2 + x3
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     44 5093.9
## 2     42 4248.8  2    845.07 4.1768 0.02216 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#F=4.1768,P-value=0.02216,If x2 and x3 don't provide any other information rather than x1 contained,
#the probability would be .02216.
```

```
#4.
#(a)
fit12 <- lm(Y ~ x1+x2)
fit123 <- update(fit12, ~ . + x3)
summary(fit12)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 156.6718598 18.6396443  8.405303 1.273843e-10
## x1           -1.2676542  0.2103519 -6.026351 3.347580e-07
## x2           -0.9207881  0.4348935 -2.117273 4.005967e-02
```

```
#The corresponding coefficient of x2 in fit12 has greater absolute value than in fit.
#This indicates that without introducing x3, part of the reduction of variance that can be taken by x3
```

```
#(b)
fit31 <- lm(Y~x3+x1)
anova(fit1)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x1         1 8275.4  8275.4  71.481 9.058e-11 ***
## Residuals 44 5093.9   115.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(fit31)
```

```
## Analysis of Variance Table
```

```
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x3         1 5554.9  5554.9  55.158 3.117e-09 ***
## x1         1 3483.9  3483.9  34.593 5.434e-07 ***
## Residuals 43 4330.5   100.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#SSR(X1|X3) =3483.9<8275.4=SSR(X1)
fit2 <-lm(Y ~ x2)
fit32 <-lm(Y~x3+x2)
anova(fit2)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value   Pr(>F)
## x2         1 4860.3  4860.3  25.132 9.23e-06 ***
## Residuals 44 8509.0   193.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit32)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x3         1 5554.9  5554.9  33.612 7.197e-07 ***
## x2         1  708.0   708.0   4.284   0.04451 *
## Residuals 43 7106.4   165.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#SSR(X2|X3)=708.0<4860.3=SSR(X2)
```

```r
#(c)
cor(x1,x3)
```

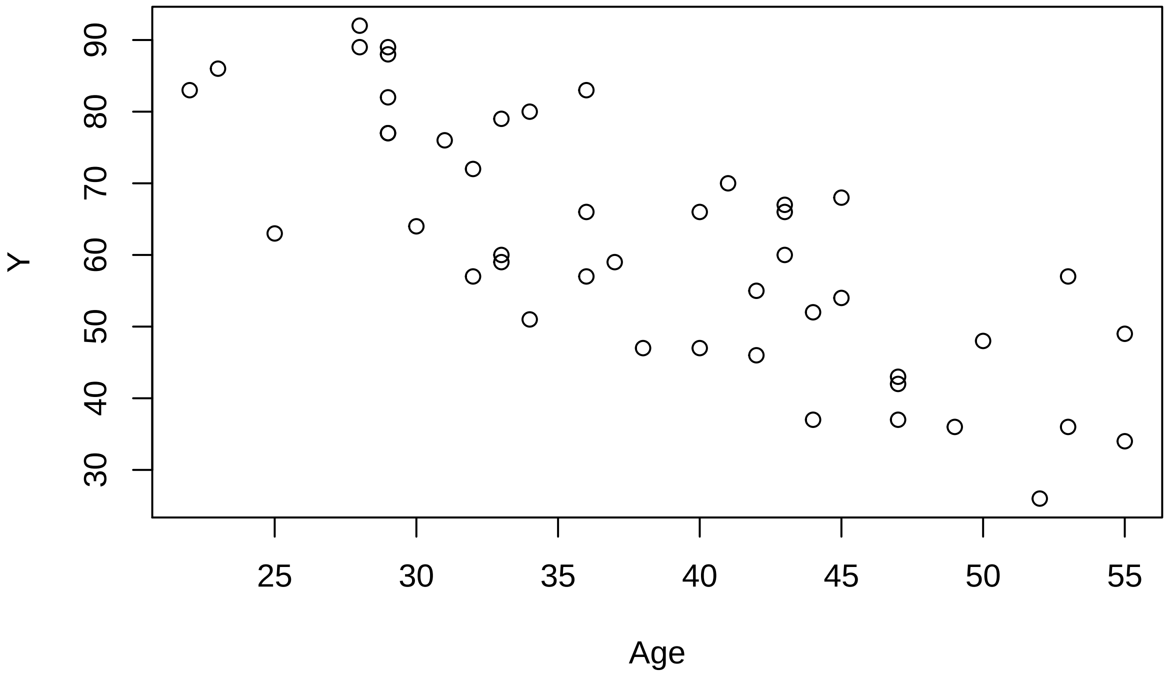
```
## [1] 0.5696775
```

```r
cor(x2,x3)
```

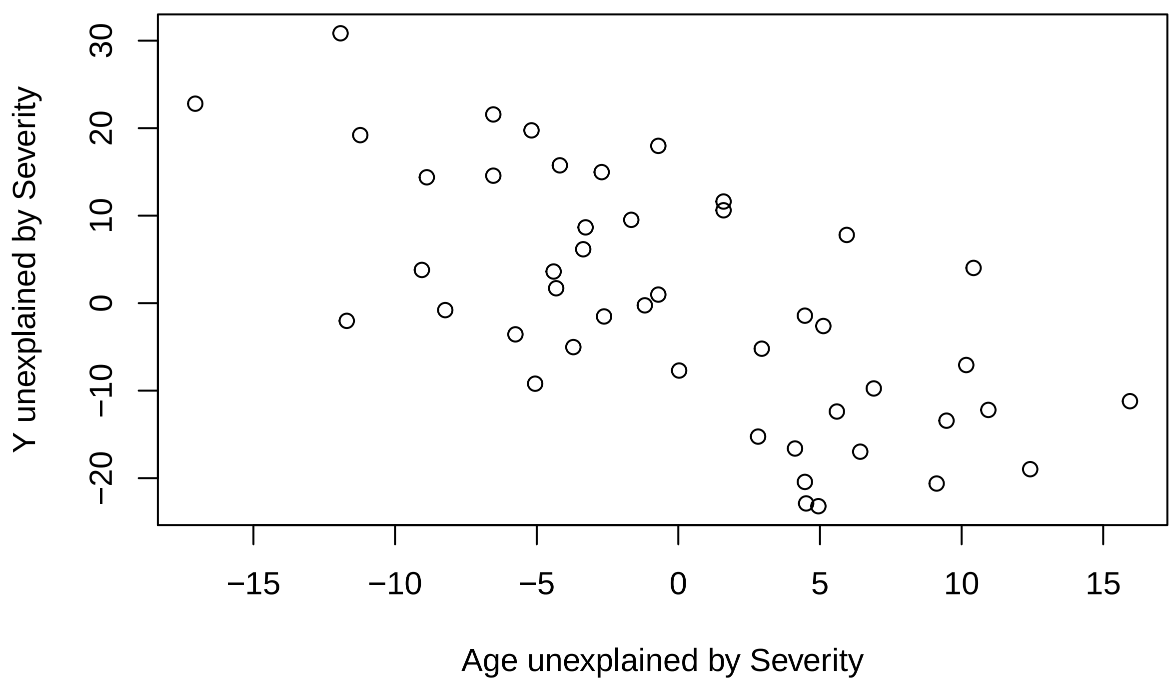
```
## [1] 0.6705287
```

```r
#We can see that Anxiety has positive relationship with both Age and Severity. Thus some of the predict
```

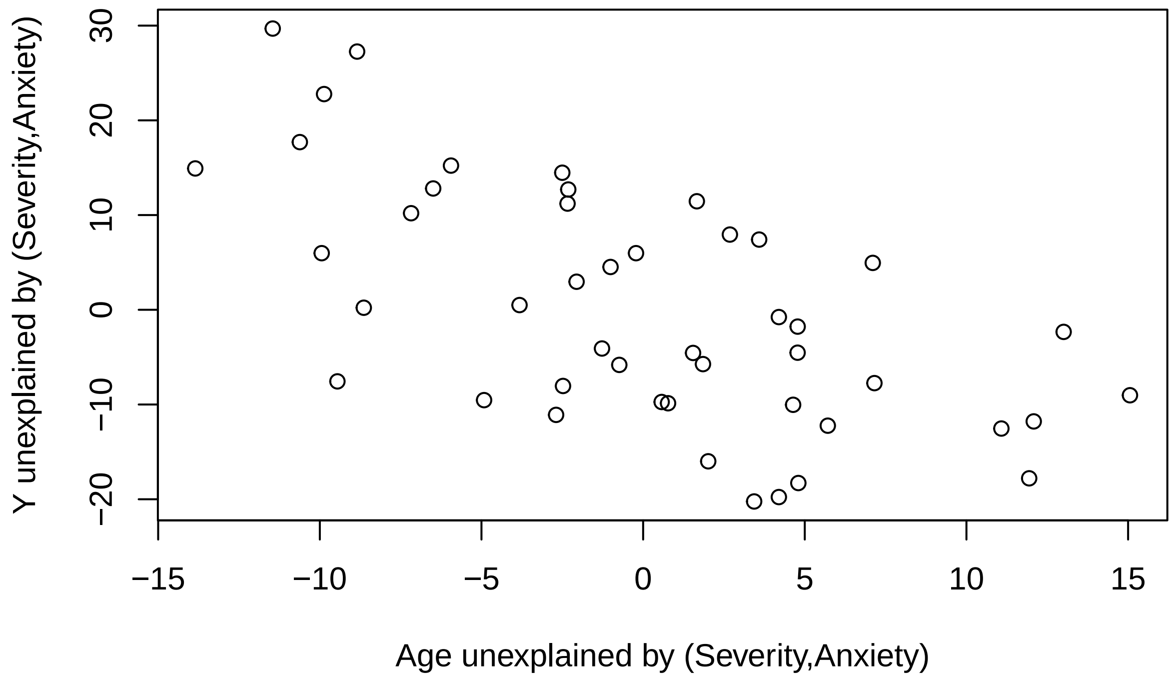
```r
#5.
#(a)i.
plot(Y ~ x1,xlab="Age")
```




```r
#ii.
fit2 <- lm(Y ~ x2)
fit1_2 <- lm(x1 ~ x2)
plot(resid(fit2) ~ resid(fit1_2),ylab="Y unexplained by Severity",xlab="Age unexplained by Severity")
```



```r
#iii.
fit23 <- lm(Y ~ x2 + x3)
fit1_23 <- lm(x1 ~ x2+x3)
plot(resid(fit23) ~ resid(fit1_23),ylab="Y unexplained by (Severity,Anxiety)",xlab="Age unexplained by
```




```r
#(b)
summary(lm(Y ~ x1))$r.squared
```

```
## [1] 0.6189843
```

```r
#R^2(Y1)=0.6189843
summary(lm(resid(fit2) ~ resid(fit1_2)))$r.squared
```

```
## [1] 0.4578709
```

```r
#R^2(Y1|2)=0.4578709
summary(lm(resid(fit23) ~ resid(fit1_23)))$r.squared
```

```
## [1] 0.4021102
```

```r
#R^2(Y1|23)=0.4021102
#So,R^2(Y1)=0.6189843 > R^2(Y1|2)=0.46 > R^2(Y1|23)=0.4021102
#The degree of marginal linear association between Y and X1 is reduced by adjusting for X2,
#and reduced by adjusting for X2 and X3. This was apparent in the scatterplots of part(a).
```