Stats 330 A4

Richard Choi

Due Date: 12pm Friday 25 October

Question 1

Create a data frame heart.df that contains the data from the file heart.data. Make sure each of the variables has been specified as the appropriate class. Include the output from the str(heart.df) and summary(heart.df) in your answer.

```
data <- read.table("heart.txt")
factor_columns = c("sex", "cp", "fbs", "restecg", "exang", "num")
heart.df <- as.data.frame(data)
heart.df[, factor_columns] <- lapply(heart.df[, factor_columns], as.factor)

str(heart.df)

## 'data.frame':    261 obs. of  11 variables:
##  $ age     : int  28 29 30 31 32 32 32 33 34 34 ...
##  $ sex     : Factor w/ 2 levels "0","1": 2 2 1 1 1 2 2 2 1 2 ...
##  $ cp      : Factor w/ 4 levels "1","2","3","4": 2 2 1 2 2 2 2 3 2 2 ...
##  $ trestbps: int  130 120 170 100 105 110 125 120 130 150 ...
##  $ chol    : int  132 243 237 219 198 225 254 298 161 214 ...
##  $ fbs     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ restecg : Factor w/ 3 levels "0","1","2": 3 1 2 2 1 1 1 1 1 2 ...
##  $ thalach : int  185 160 170 150 165 184 155 185 190 168 ...
##  $ exang   : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ oldpeak : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ num     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...

summary(heart.df)

##       age        sex     cp         trestbps          chol        fbs
##  Min.   :28.00   0: 69   1: 10   Min.   : 92.0   Min.   : 85.0   0:242
##  1st Qu.:42.00   1:192   2: 92   1st Qu.:120.0   1st Qu.:208.0   1: 19
##  Median :49.00           3: 46   Median :130.0   Median :242.0
##  Mean   :47.77           4:113   Mean   :132.6   Mean   :248.8
##  3rd Qu.:54.00                   3rd Qu.:140.0   3rd Qu.:280.0
##  Max.   :65.00                   Max.   :200.0   Max.   :603.0
##  restecg    thalach      exang      oldpeak       num
##  0:208   Min.   : 82.0   0:178   Min.   :0.0000   0:163
##  1: 47   1st Qu.:122.0   1: 83   1st Qu.:0.0000   1: 98
##  2:  6   Median :140.0           Median :0.0000  
##          Mean   :139.2           Mean   :0.6123  
##          3rd Qu.:155.0           3rd Qu.:1.0000  
##          Max.   :190.0           Max.   :5.0000  
```

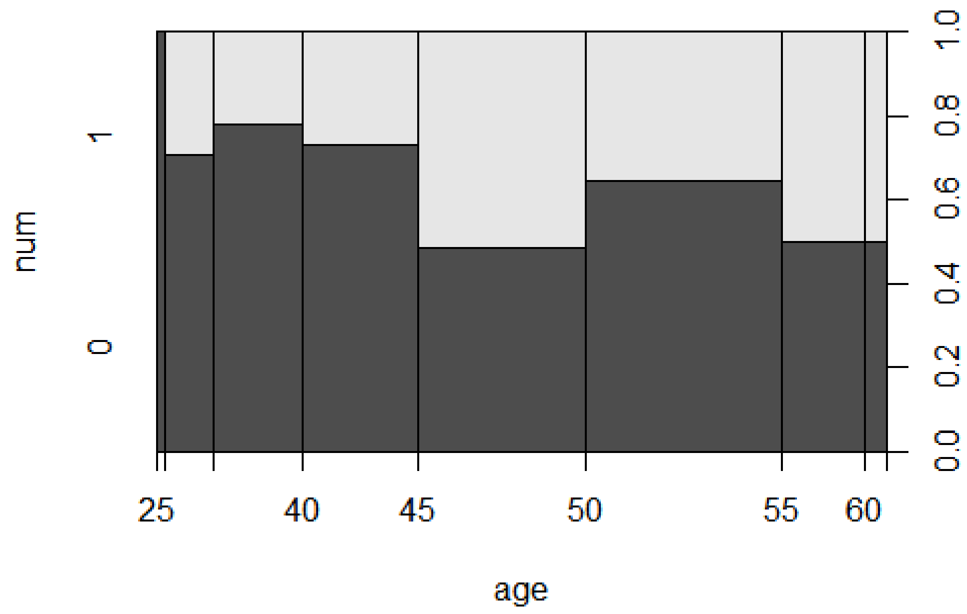
```
head(heart.df, 10)
```

```
##     age sex cp trestbps chol fbs restecg thalach exang oldpeak num
## 1    28   1  2      130  132   0       2     185     0       0   0
## 2    29   1  2      120  243   0       0     160     0       0   0
## 4    30   0  1      170  237   0       1     170     0       0   0
## 5    31   0  2      100  219   0       1     150     0       0   0
## 6    32   0  2      105  198   0       0     165     0       0   0
## 7    32   1  2      110  225   0       0     184     0       0   0
## 8    32   1  2      125  254   0       0     155     0       0   0
## 9    33   1  3      120  298   0       0     185     0       0   0
## 10   34   0  2      130  161   0       0     190     0       0   0
## 11   34   1  2      150  214   0       1     168     0       0   0
```

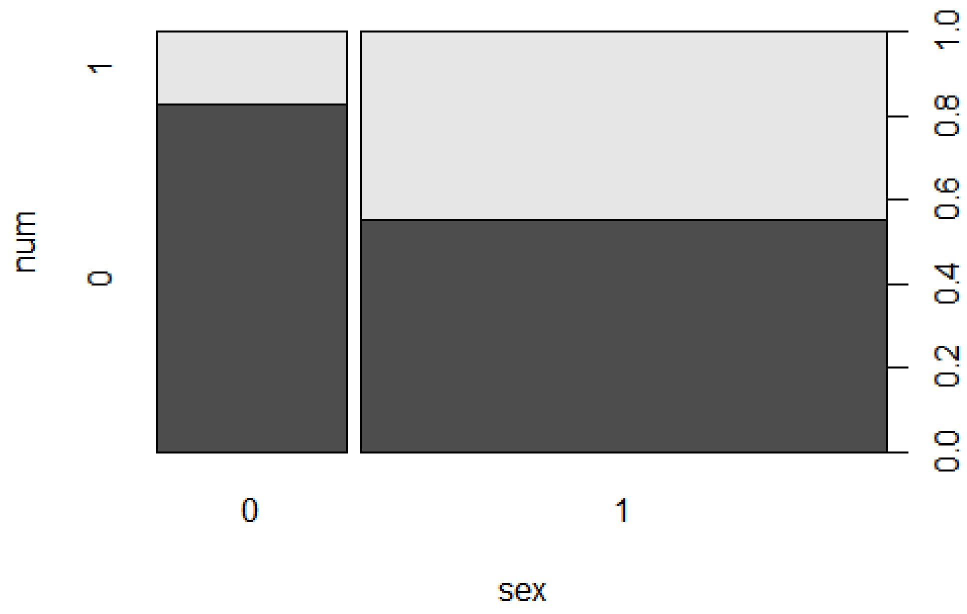
Question 2

For each explanatory variable, create a suitable plot that explores the relationship between that variable and the response. Briefly comment on these plots.

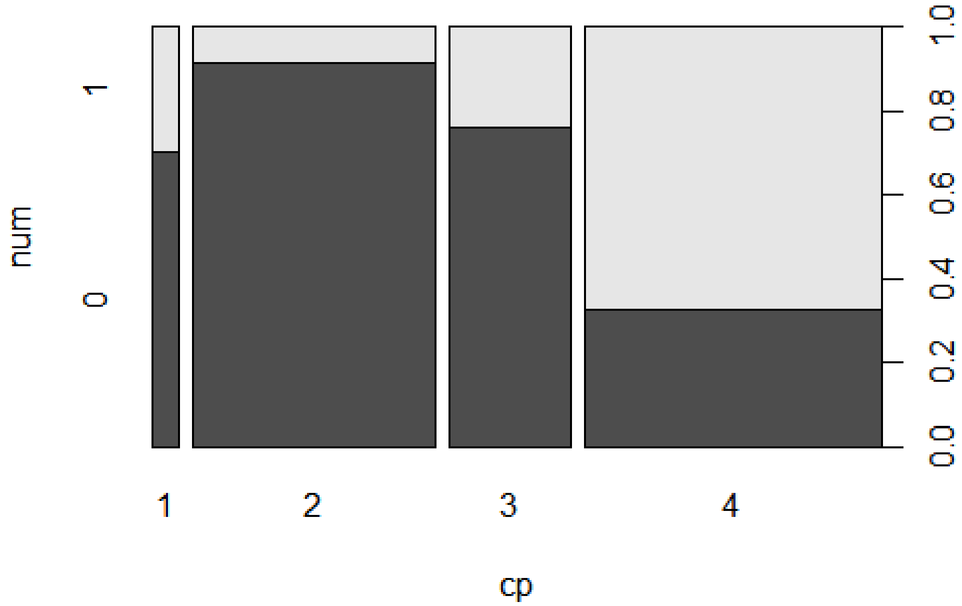
```
plot(num ~ age, data = heart.df)
```



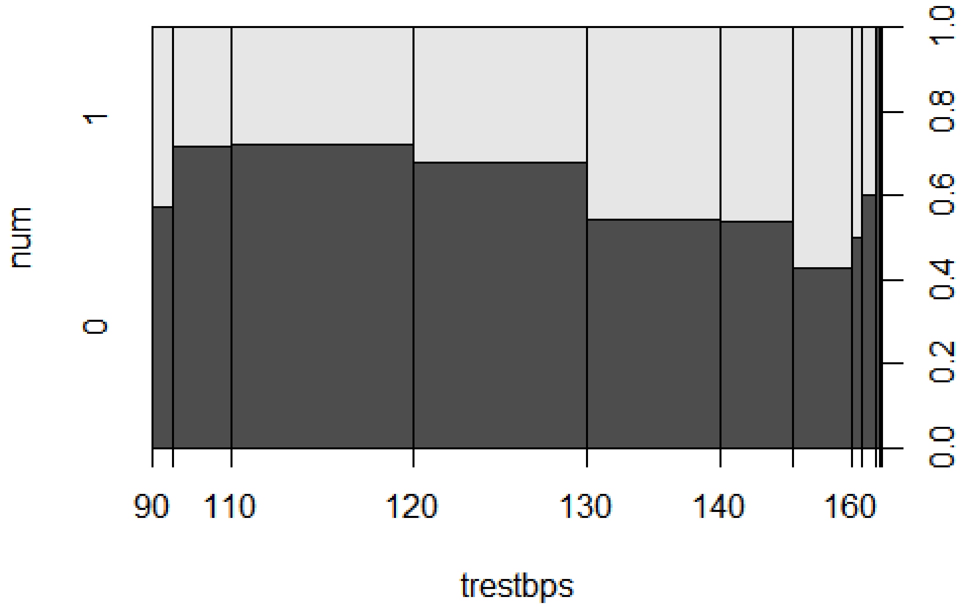
```r
plot(num ~ sex, data = heart.df)
```



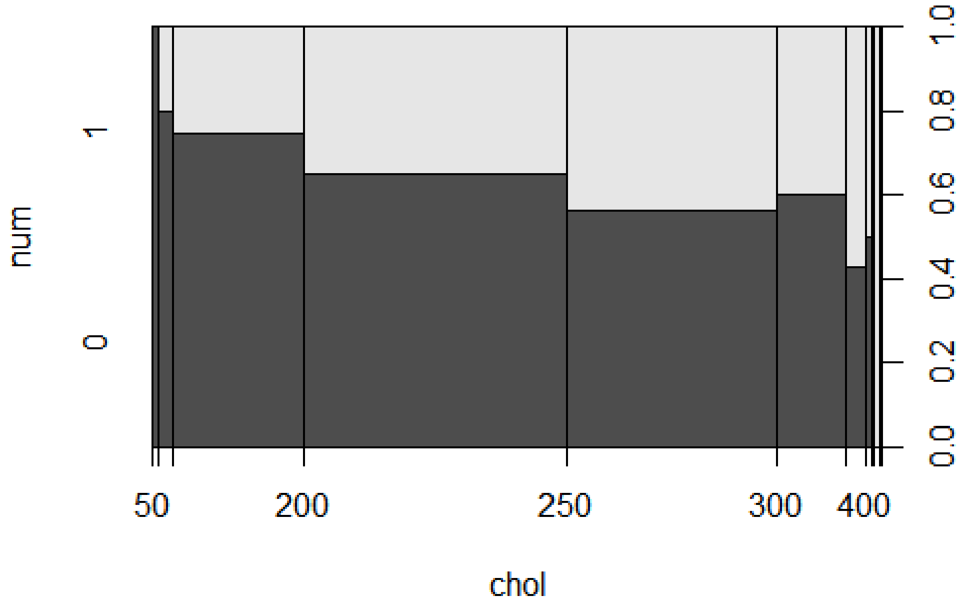
```r
plot(num ~ cp, data = heart.df)
```



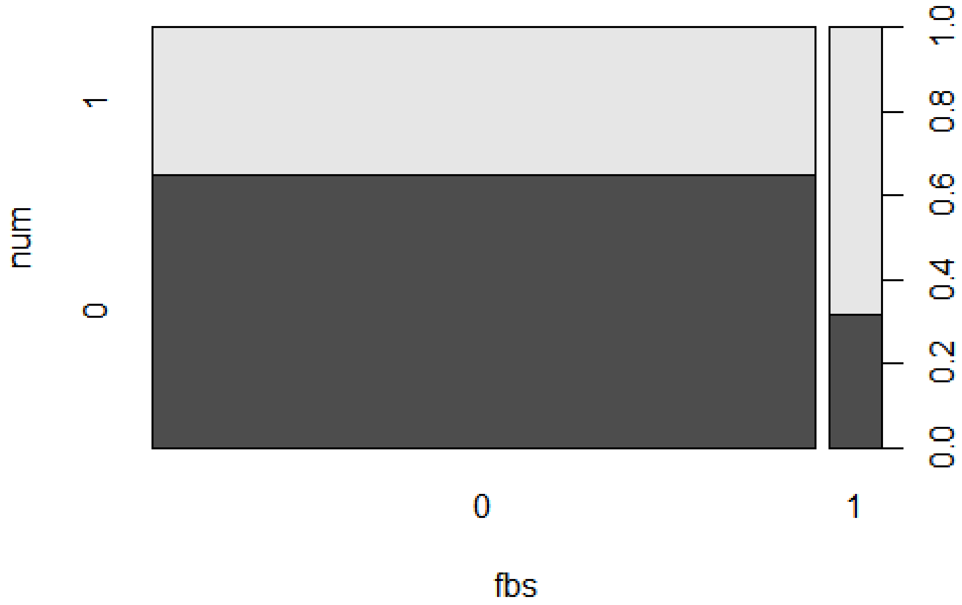
```
plot(num ~ trestbps, data = heart.df)
```



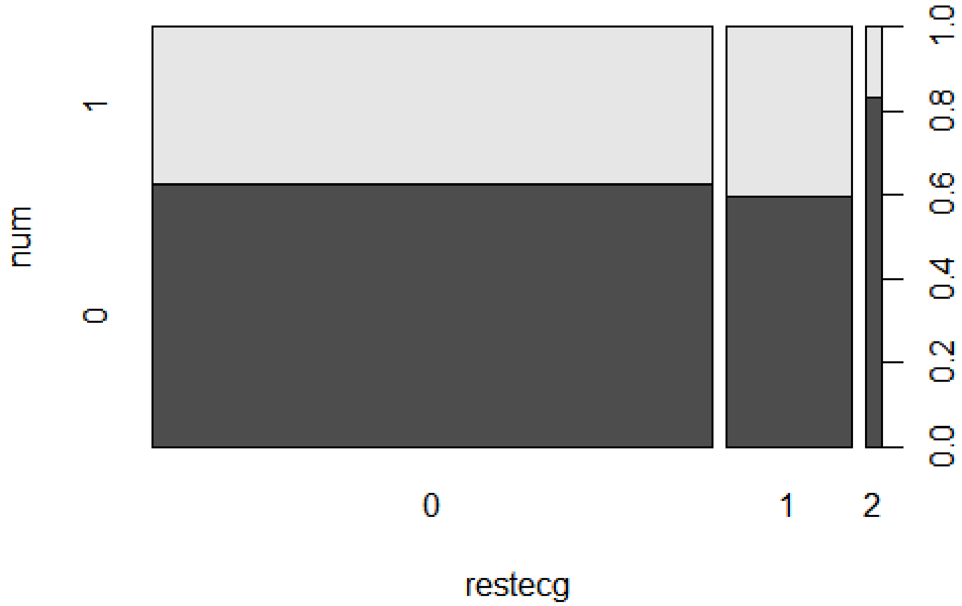
```
plot(num ~ chol, data= heart.df)
```



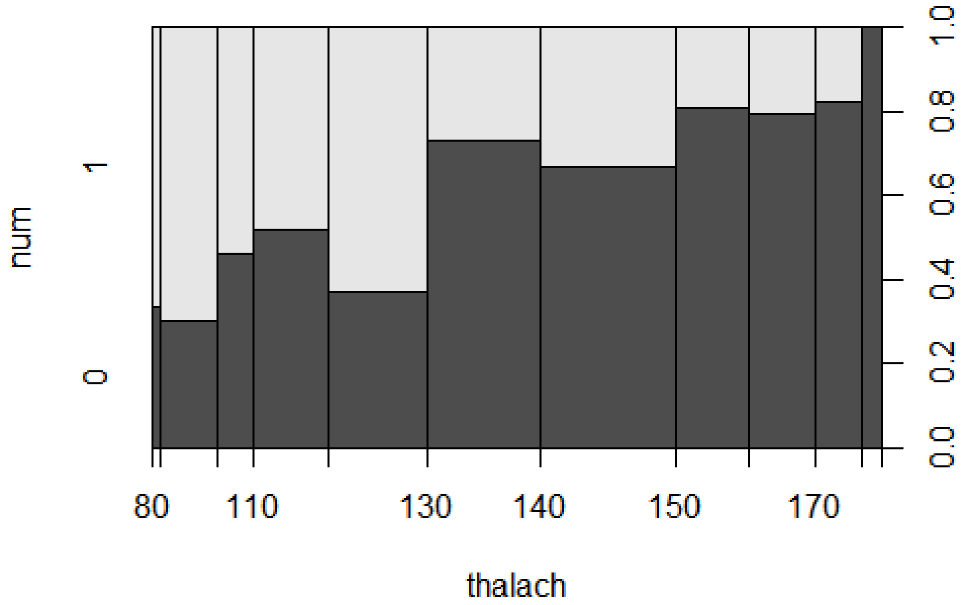
```r
plot(num ~ fbs, data = heart.df)
```



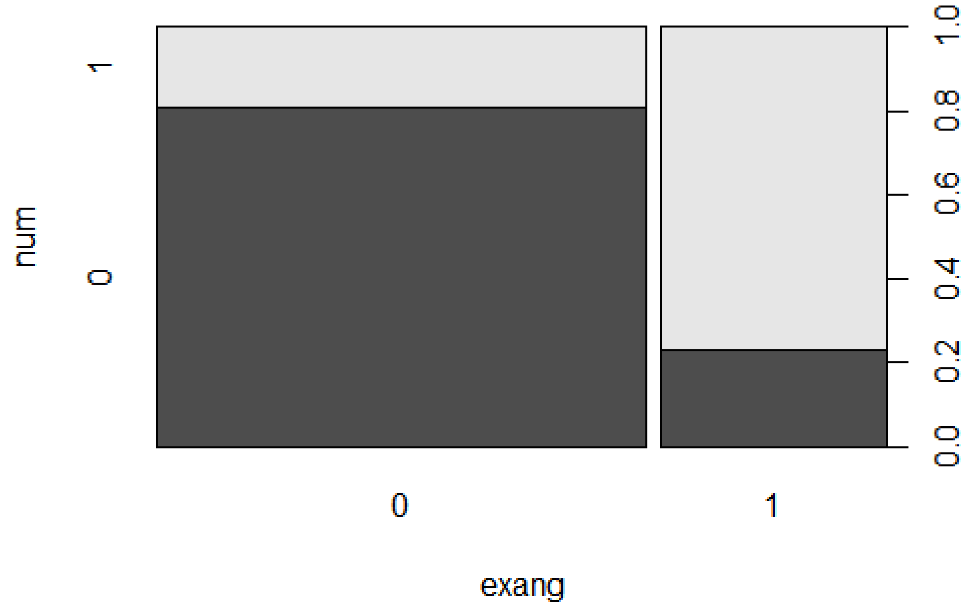
```r
plot(num ~ restecg, data = heart.df)
```



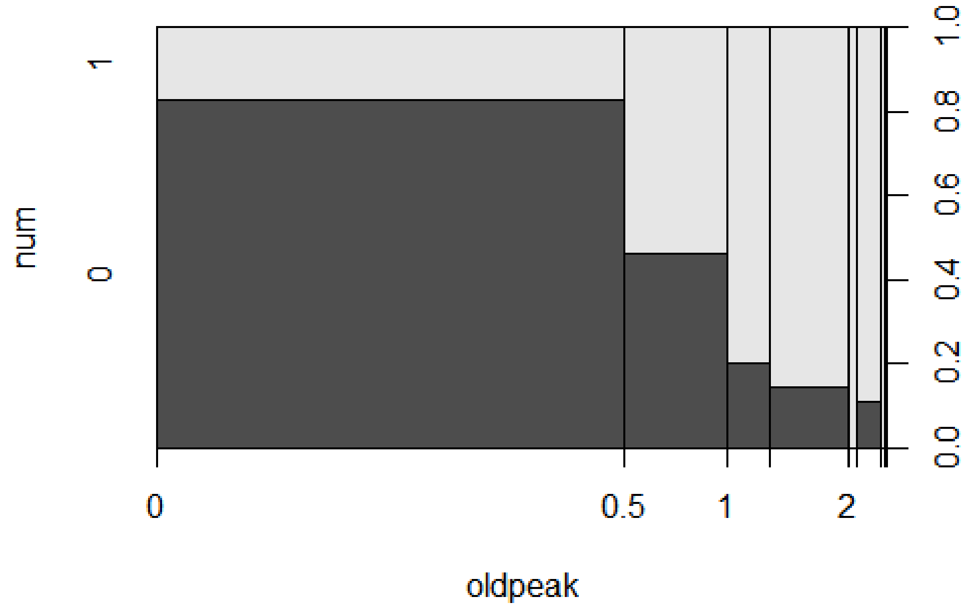
```r
plot(num ~ thalach, data = heart.df)
```



```r
plot(num ~ exang, data = heart.df)
```



```r
plot(num ~ oldpeak, data = heart.df)
```



We observe that the younger the person, higher the proportion of diagnosis of absent heart disease.

We observe that female diagnosed with no heart disease is more than the male. We observe that people with chest pain type 2, type 3, type 1, type 4 have the least proportion of present heart disease respectively.

We observe that the people with trestbps of 100 to 130 (in mm Hg) tend to have least heart disease diagnosis proportion.

We observe that higher the cholesterol (in mg/dl), the proportion of people with diagnosis of present heart disease decreases.

We observe that there are higher proportion of diagnosis of absent heart disease when people's fasting blood sugar is under 120 mg/dl.

We observe that the people showing left ventricular hypertrophy have higher proportion of diagnosis of absent heart disease. Whereas people with normal resting electrocardiographic result and wave abnormality result have around the similar proportion of diagnosis of absent heart disease, 0.6. We observe that the higher the maximum heart rate achieved by the person, the person tend to have a higher chance of diagnosed with absent heart disease.

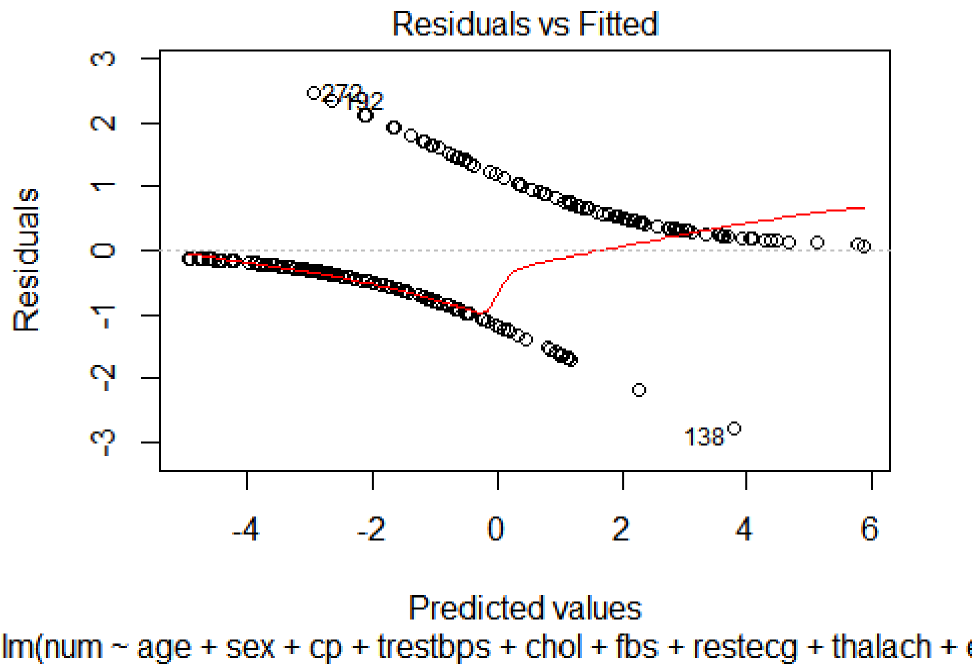
We observe that around 80% people who experienced angina is diagnosed with present heart disease, whereas 20% people who expereienced angina is diagnosed with absent heart disease.

We observe that higher the depression induced by exercise relative to rest, the lower the proportion of diagnosis with absent heart disease.

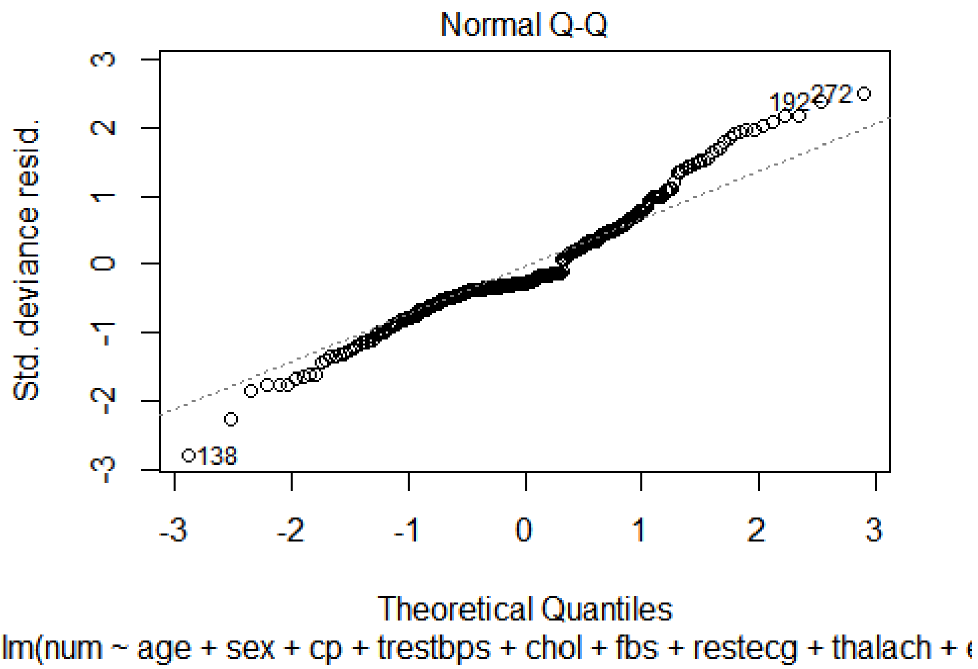
Question 3

Fit an initial model that relates the response to the regressors and do diagnostics. Adjust your model as appropriate.

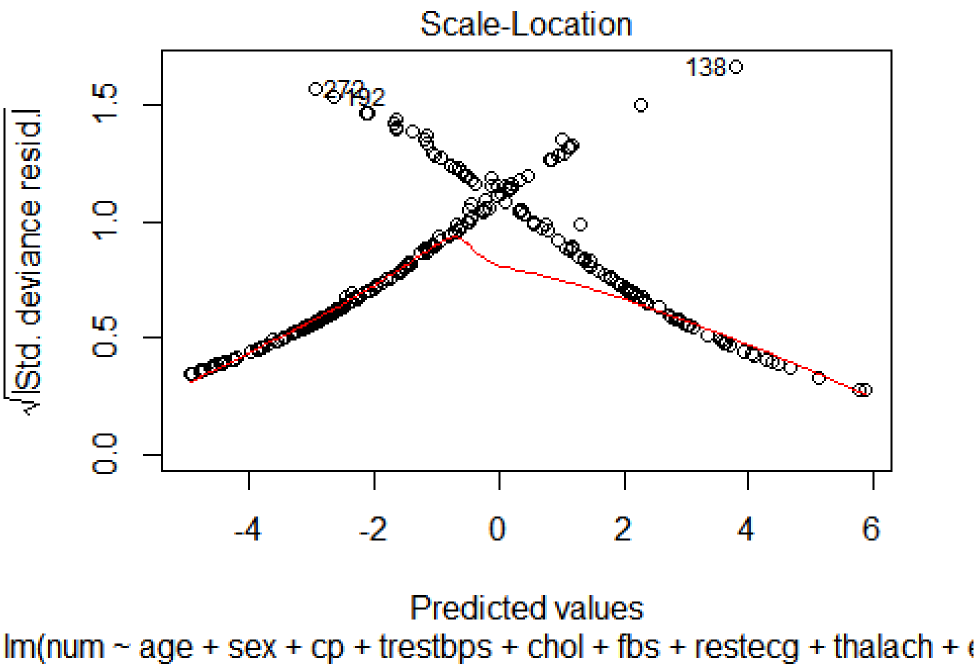
```r
library(mgcv)

## Warning: package 'mgcv' was built under R version 3.5.3

## Loading required package: nlme

## This is mgcv 1.8-29. For overview type 'help("mgcv-package")'.

continuous_columns = c("age", "trestbps", "chol", "thalach", "oldpeak")
heart_model <- glm(num ~ age + sex + cp + trestbps + chol + fbs + restecg +
thalach + exang + oldpeak, family = "binomial", data = heart.df)
plot(heart_model, which = 1)
```



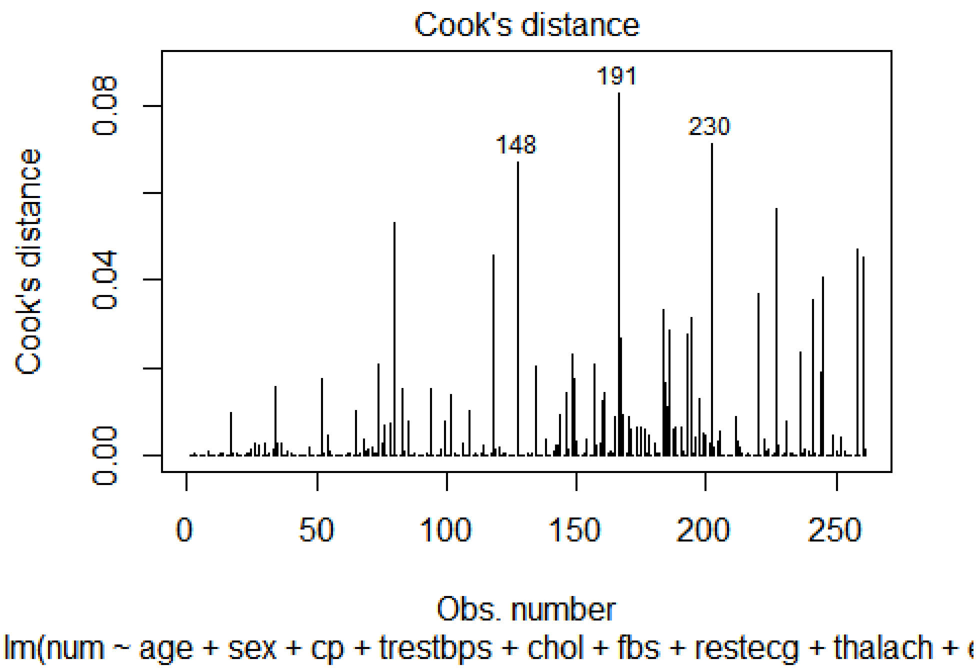
```
plot(heart_model, which = 2)
```



```
plot(heart_model, which = 3)
```



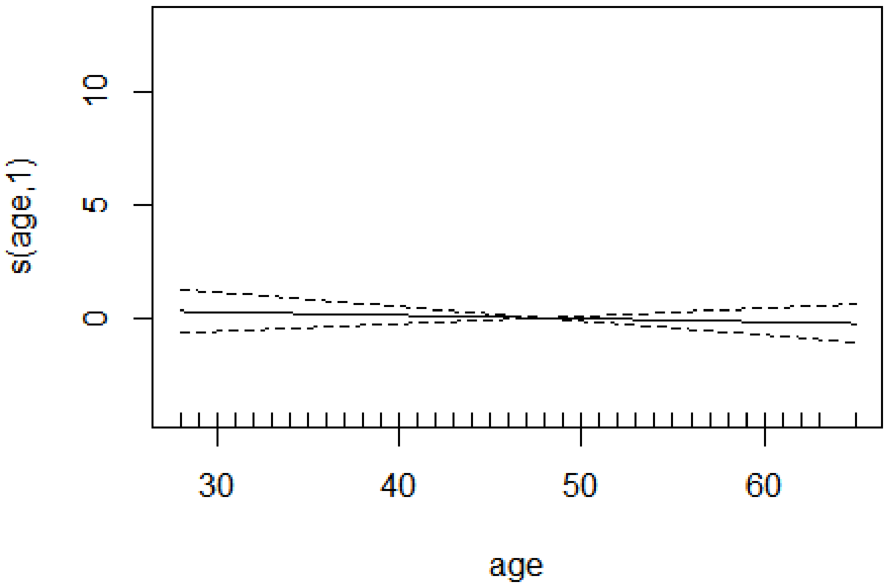
```r
plot(heart_model, which = 4)
```

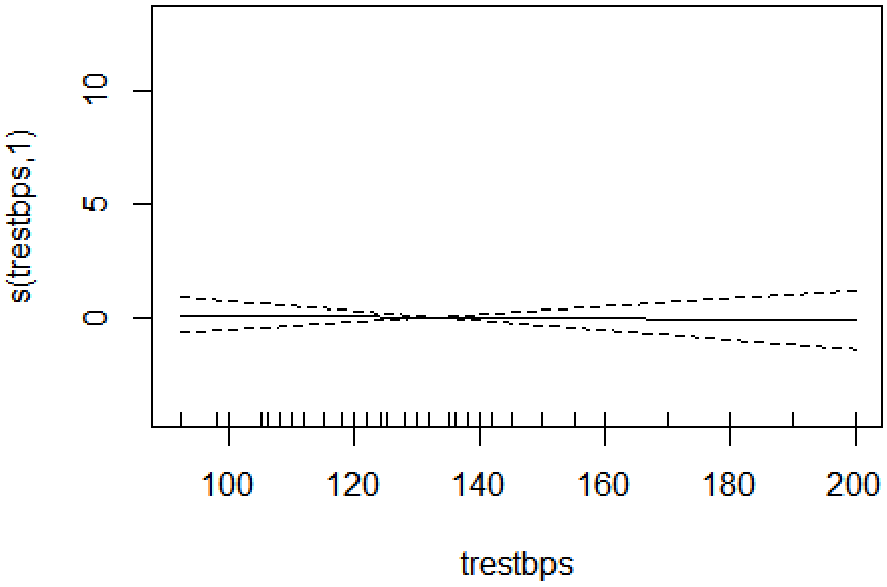


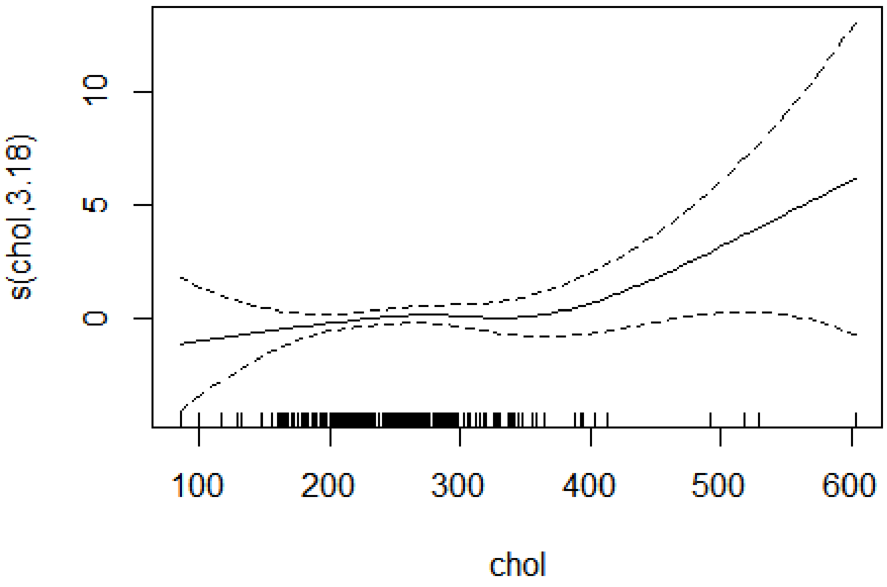
```r
summary(heart_model)
```

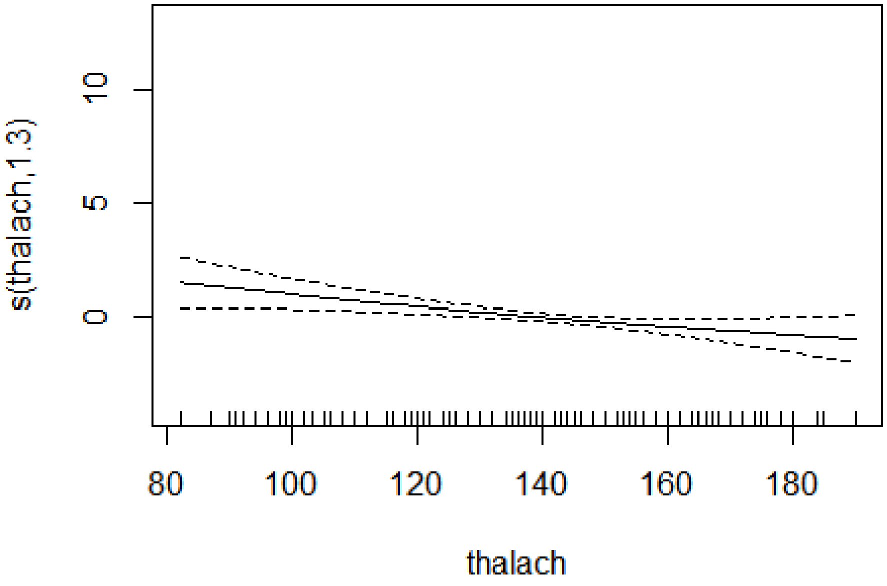
```
##
## Call:
## glm(formula = num ~ age + sex + cp + trestbps + chol + fbs +
##     restecg + thalach + exang + oldpeak, family = "binomial",
##     data = heart.df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7682  -0.4879  -0.2655   0.4366   2.4531
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.669215   2.975271  -0.561  0.57478
## age         -0.009255   0.029526  -0.313  0.75394
## sex1         1.307881   0.481892   2.714  0.00665 **
## cp2         -1.919972   1.022129  -1.878  0.06033 .
## cp3         -0.515823   1.008135  -0.512  0.60889
## cp4          0.368482   0.967228   0.381  0.70323
## trestbps    -0.001603   0.011514  -0.139  0.88925
## chol         0.005218   0.002788   1.872  0.06127 .
## fbs1         1.639892   0.779925   2.103  0.03550 *
## restecg1    -0.391962   0.535431  -0.732  0.46414
## restecg2    -0.862764   1.715431  -0.503  0.61500
## thalach     -0.008658   0.010189  -0.850  0.39549
## exang1       0.943270   0.487140   1.936  0.05283 .
## oldpeak      1.184473   0.280460   4.223 2.41e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 345.46  on 260  degrees of freedom
## Residual deviance: 184.70  on 247  degrees of freedom
## AIC: 212.7
##
## Number of Fisher Scoring iterations: 6
```

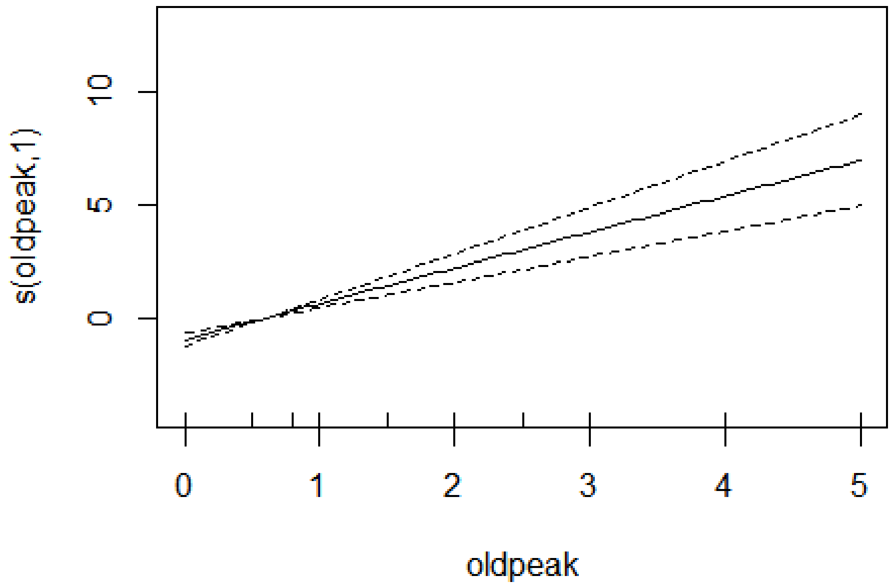
```
gam.fit <- gam(num ~ s(age) + s(trestbps) + s(chol) + s(thalach) +
s(oldpeak), family = "binomial", data = heart.df)
plot(gam.fit)
```


s(age,1)
10
5
0
30
40
50
60
age


s(trestbps,1)
10
5
0
100
120
140
160
180
200
trestbps


s(chol,3.18)
10
5
0
100
200
300
400
500
600
chol


s(thalach,1.3)
10
5
0
80
100
120
140
160
180
thalach

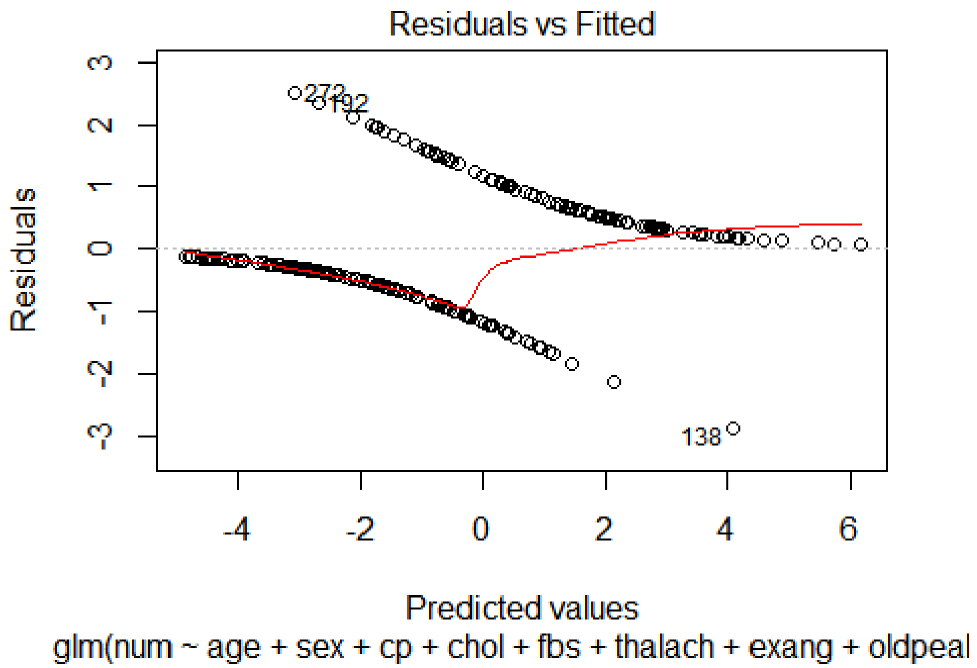


```
poly_heart_model <- glm(num ~ age + sex + cp + trestbps + chol + I(chol^2) +
fbs + thalach + exang + oldpeak, family = "binomial", data = heart.df)
anova(heart_model, poly_heart_model, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: num ~ age + sex + cp + trestbps + chol + fbs + restecg + thalach
+
##     exang + oldpeak
## Model 2: num ~ age + sex + cp + trestbps + chol + I(chol^2) + fbs +
thalach +
##     exang + oldpeak
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       247     184.70
## 2       248     184.24 -1  0.46079

adj_heart_model <- glm(num ~ age + sex + cp  + chol + fbs + thalach + exang +
oldpeak, family = "binomial", data = heart.df)
summary(adj_heart_model)

##
## Call:
## glm(formula = num ~ age + sex + cp + chol + fbs + thalach + exang +
##     oldpeak, family = "binomial", data = heart.df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8677  -0.5280  -0.2681   0.4326   2.4980
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.802327   2.668049  -0.676  0.49934
## age         -0.013877   0.028684  -0.484  0.62854
## sex1         1.314506   0.477861   2.751  0.00594 **
## cp2         -1.892794   1.003766  -1.886  0.05934 .
## cp3         -0.486961   0.986806  -0.493  0.62168
## cp4          0.408731   0.950912   0.430  0.66732
## chol         0.005029   0.002751   1.828  0.06755 .
## fbs1         1.588445   0.778663   2.040  0.04135 *
## thalach     -0.008037   0.010038  -0.801  0.42337
## exang1       0.903555   0.484877   1.863  0.06240 .
## oldpeak      1.199613   0.280276   4.280 1.87e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 345.46  on 260  degrees of freedom
## Residual deviance: 185.51  on 250  degrees of freedom
## AIC: 207.51
##
## Number of Fisher Scoring iterations: 5
```

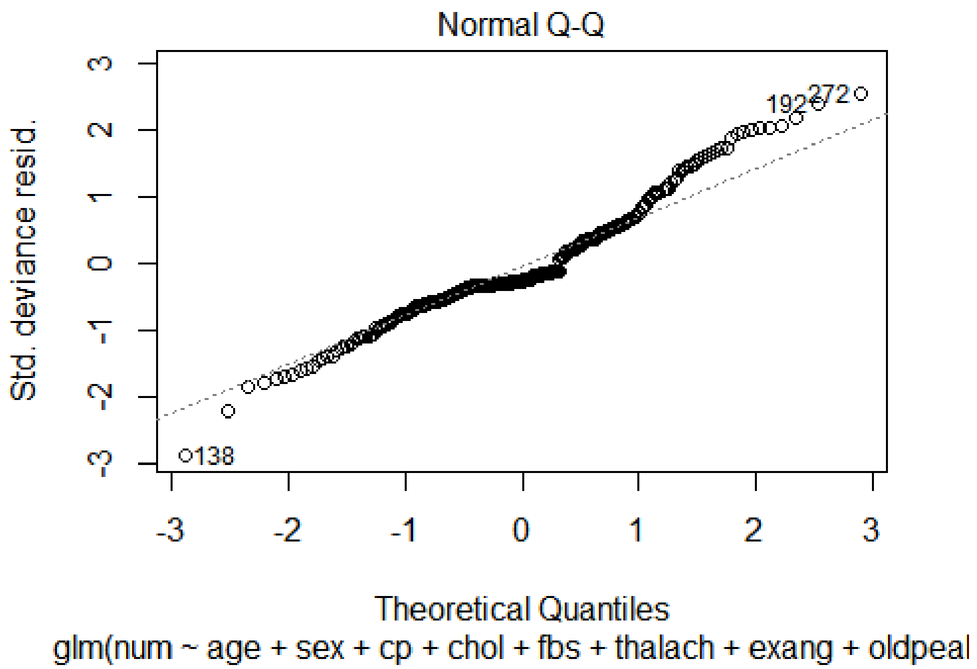
```r
1 - pchisq(185.51, 250)
```

```
## [1] 0.9991758
```

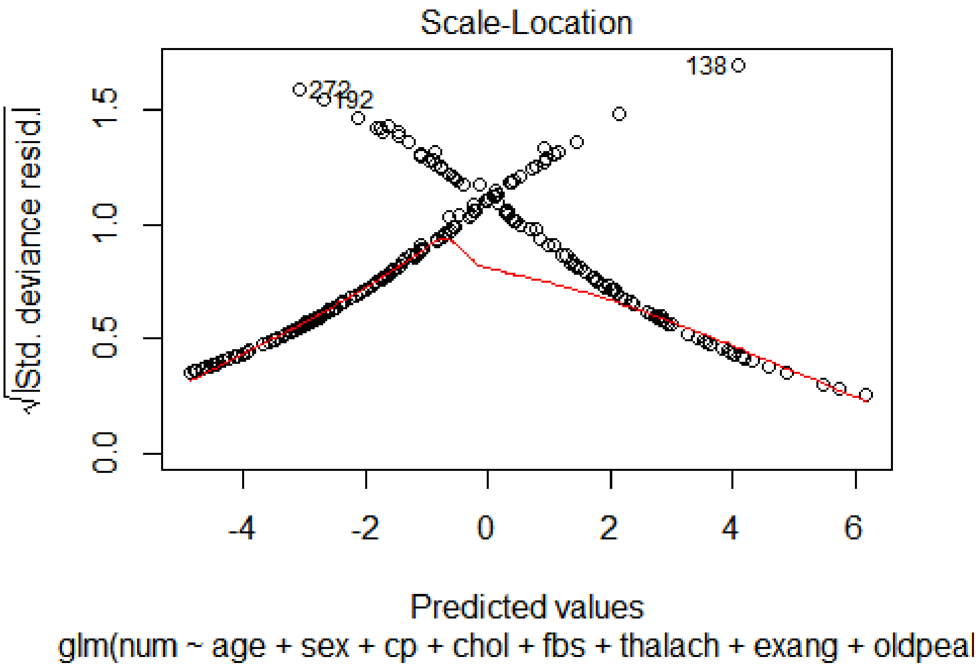
```r
plot(adj_heart_model, which = 1)
```



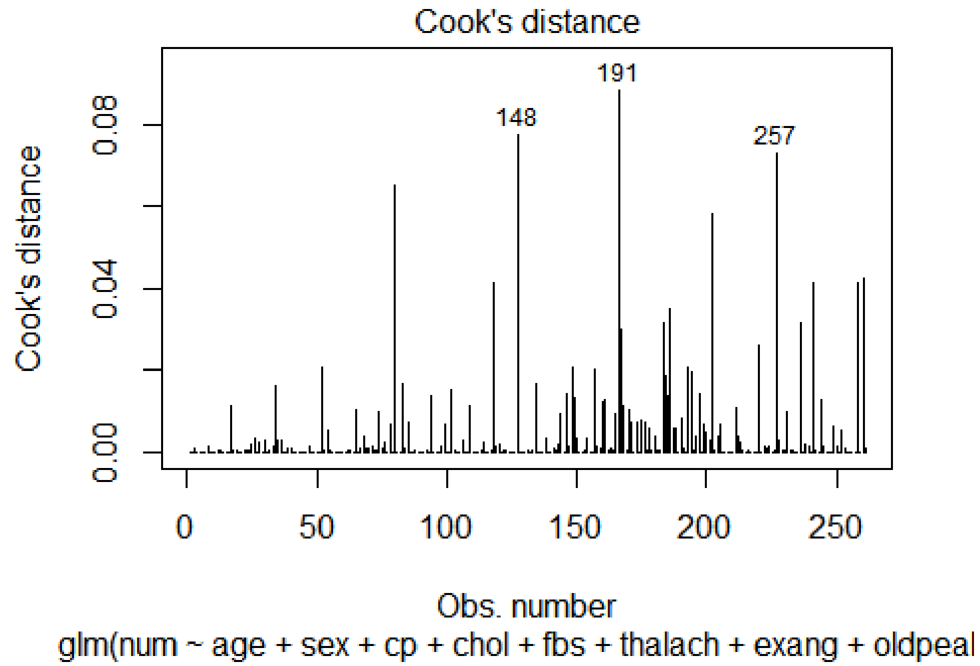
```
plot(adj_heart_model, which = 2)
```



```
plot(adj_heart_model, which = 3)
```



```r
plot(adj_heart_model, which = 4)
```



I have fitted a binomial model with all the explanatory variables in heart data initially, then I have fitted gam plot to potential polynomial term in any of the numeric explanatory variables. Gam plot suggested quadratic term for the cholestrol variable but the anova test provided evidence that we do not need quadratic term in our model. I've also excluded trestbps and restecg as they weren't statistically significant (p - value = 0.89, p - value = 0.46 respectively). Also, the trestbps and num plot showed similar proportion of diagnosis of absent heart disease regardless of trestbps (mm Hg). Similarly, restecg and num plot showed almost the same proportion of diagnosis of absent heart disease whether people had normal ectrocardiographic result or wave abnormality result. Although, people with left ventricular hypertrophy showed a significant proportion of absent heart disease, the sample of people with the resut is too low. Both the heart model and adjusted heart model's diagnostic plots were similar and both were ok. For both of the models, the cooks model had several observations that was over 0.04 so further invstigation is required.

Question 4

Use dredge to produce a “short list” of promising models.

```
library(MuMIn)

## Warning: package 'MuMIn' was built under R version 3.5.3

options(na.action = "na.fail")

all.fits <- dredge(adj_heart_model)

## Fixed term is "(Intercept)"

head(all.fits)

## Global model call: glm(formula = num ~ age + sex + cp + chol + fbs + 
thalach + exang + 
##     oldpeak, family = "binomial", data = heart.df)
## ---
## Model selection table 
##     (Intrc)       age     chol cp exang fbs oldpk sex     thlch df  logLik
## 127  -3.629          0.005109  +     +   + 1.199   +            9 -93.081
## 125  -2.422                    +     +   + 1.182   +            8 -94.905
## 255  -2.790          0.005104  +     +   + 1.195   + -0.005812 10 -92.871
## 119  -3.780          0.005256  +         + 1.502   +            8 -95.189
## 128  -3.469 -0.003379 0.005091  +     +   + 1.200   +           10 -93.073
## 111  -3.548          0.005132  +     +     1.169   +            8 -95.372
##      AICc delta weight
## 127 204.9  0.00  0.347
## 125 206.4  1.50  0.164
## 255 206.6  1.74  0.145
## 119 206.9  2.07  0.123
## 128 207.0  2.15  0.119
## 111 207.3  2.44  0.103
## Models ranked by AICc(x)
```

```
all.fits_2 <- dredge(heart_model, rank = "BIC")
```

```
## Fixed term is "(Intercept)"
```

```
head(all.fits_2)
```

```
## Global model call: glm(formula = num ~ age + sex + cp + trestbps + chol + 
fbs + 
##     restecg + thalach + exang + oldpeak, family = "binomial", 
##     data = heart.df)
## ---
## Model selection table 
##     (Intrc)     chol cp exang fbs oldpk sex df  logLik   BIC delta weight
## 181  -2.545           +         + 1.499   +  7 -97.094 233.1  0.00  0.243
## 165  -2.480           +           1.501   +  6 -99.995 233.4  0.24  0.216
## 173  -2.324           +     +     1.159   +  7 -97.305 233.6  0.42  0.197
## 189  -2.422           +     +   + 1.182   +  8 -94.905 234.3  1.18  0.135
## 167  -3.742 0.005341  +           1.492   +  7 -97.909 234.8  1.63  0.108
## 183  -3.780 0.005256  +         + 1.502   +  8 -95.189 234.9  1.75  0.101
## Models ranked by BIC(x)
```

```
first.model <- get.models(all.fits, 1)[[1]]
summary(first.model)
```

```
## 
## Call:
## glm(formula = num ~ chol + cp + exang + fbs + oldpeak + sex + 
##     1, family = "binomial", data = heart.df)
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -2.8459  -0.5271  -0.2771   0.4402   2.4419  
## 
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)    
## (Intercept) -3.629203   1.163734  -3.119  0.00182 ** 
## chol         0.005109   0.002720   1.878  0.06037 .  
## cp2         -1.923359   1.008316  -1.907  0.05646 .  
## cp3         -0.540244   0.989841  -0.546  0.58521    
## cp4          0.463266   0.951843   0.487  0.62647    
## exang1       0.963505   0.464033   2.076  0.03786 *  
## fbs1         1.631055   0.775590   2.103  0.03547 *  
## oldpeak      1.198613   0.277870   4.314 1.61e-05 ***
## sex1         1.310276   0.480069   2.729  0.00635 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 345.46  on 260  degrees of freedom
## Residual deviance: 186.16  on 252  degrees of freedom
## AIC: 204.16
##
## Number of Fisher Scoring iterations: 5
```

Question 5

Evaluate the top models from your short list using cross validation and choose a predictive model. Explain your choice.

```
library("crossval")

## Warning: package 'crossval' was built under R version 3.5.3

library("pROC")

## Warning: package 'pROC' was built under R version 3.5.3

## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var

set.seed(12345)

predfun.lm <- function(train.x, train.y, test.x, test.y) {
  lm.fit <- glm(train.y ~ chol + cp + exang + fbs + oldpeak + sex, family =
"binomial", data = train.x)
  ynew <- predict(lm.fit, newdata = test.x, type = "response")
  my.roc = roc(response = test.y, predictor = ynew)
  out <- my.roc$auc

  lm.fit2 <- glm(train.y ~ cp + exang + fbs + oldpeak + sex, family =
"binomial", data = train.x)
  ynew <- predict(lm.fit2, newdata = test.x, type = "response")
  out2 <- roc(response = test.y, predictor = ynew)$auc

  lm.fit3 <- glm(train.y ~ chol + cp + exang + fbs + oldpeak + sex + thalach,
family = "binomial", data = train.x)
  ynew <- predict(lm.fit3, newdata = test.x, type = "response")
  out3 <- roc(response = test.y, predictor = ynew)$auc

  lm.fit4 <- glm(train.y ~ chol + cp + fbs + oldpeak + sex, family =
"binomial", data = train.x)
  ynew <- predict(lm.fit4, newdata = test.x, type = "response")
  out4 <- roc(response = test.y, predictor = ynew)$auc

  lm.fit5 <- glm(train.y ~ age + chol + cp + exang + fbs + oldpeak + sex,
family = "binomial", data = train.x)
  ynew <- predict(lm.fit5, newdata = test.x, type = "response")
  out5 <- roc(response = test.y, predictor = ynew)$auc

  lm.fit6 <- glm(train.y ~ chol + cp + exang + oldpeak + sex, family =
"binomial", data = train.x)
  ynew <- predict(lm.fit6, newdata = test.x, type = "response")
  out6 <- roc(response = test.y, predictor = ynew)$auc
  return(c(out, out2, out3, out4, out5, out6))
}
cv.out <- crossval(predfun.lm, X  = heart.df[, 1:10], Y = heart.df[, 11], K =
10, B = 10, verbose = FALSE)
```

```
= 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
##
```

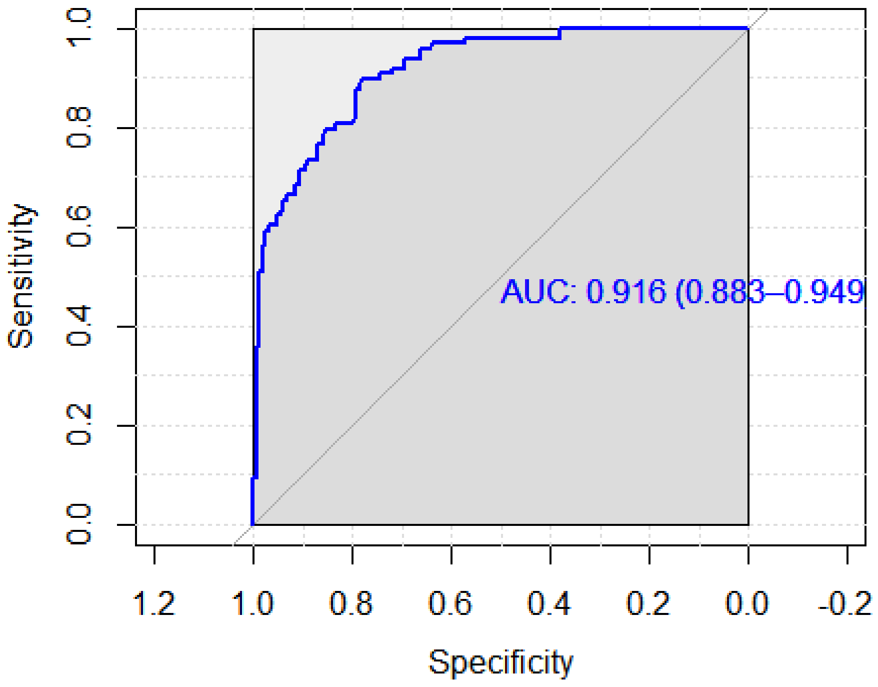
```
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
round(cv.out$stat, 4)
## [1] 0.8912 0.8894 0.8881 0.8880 0.8885 0.8895
round(cv.out$stat.se, 4)
## [1] 0.0066 0.0067 0.0067 0.0065 0.0066 0.0066
```

I've used dredge function to get the top 6 best model using AICc rank, because BIC rank seemed to prefer models with significantly less variables which omitted variables like chol and exang. Using 100 fold cross validation on the top 6 models, we've attained the highest estimate of AUC (0.9839) which orresponds to model 1 from the AICc list.

Question 6

Produce the ROC curve for the model you chose. Comment on the model's predictive ability. Find the value of the threshold c that maximizes sensitivity + specificity.

```
best.pred.model <- glm(num ~ chol + cp + exang + fbs + oldpeak + sex, family
= "binomial", data = heart.df)

heart.roc = roc(response = heart.df$num, predictor =
fitted.values(best.pred.model), ci = TRUE)

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

plot(heart.roc, col = "blue", print.auc = TRUE, auc.polygon = TRUE,
max.auc.polygon = TRUE, auc.polygoncol = 'yellow', grid = TRUE, lwd = 2.5,
print.thres.cex = 0.5)
```



The c value is 0.915 (0.881 ~ 0.948). It maximises sensitiy + specificity