6. A study was performed in a small mining town to investigate the relationship between breathlessness and wheeziness. Subjects in nine five-year age groups ranging from 20 to 65 were evaluated for the presence or absence of breathlessness and wheeziness. Data from the study are given in the accompanying output, where `wheeze` is the number of subjects diagnosed with wheeziness, `total` is the total number of subjects in each breathlessness/age group, `breath` is `B` for the breathless group and `N` otherwise, and `age` is the midpoint of the age interval.

The output shows part of an analysis of these data.

The plot shows the observed data on the 'logit scale' (ie log odds on wheeze) for the breathless and not breathless groups as a function of `log(age)`.

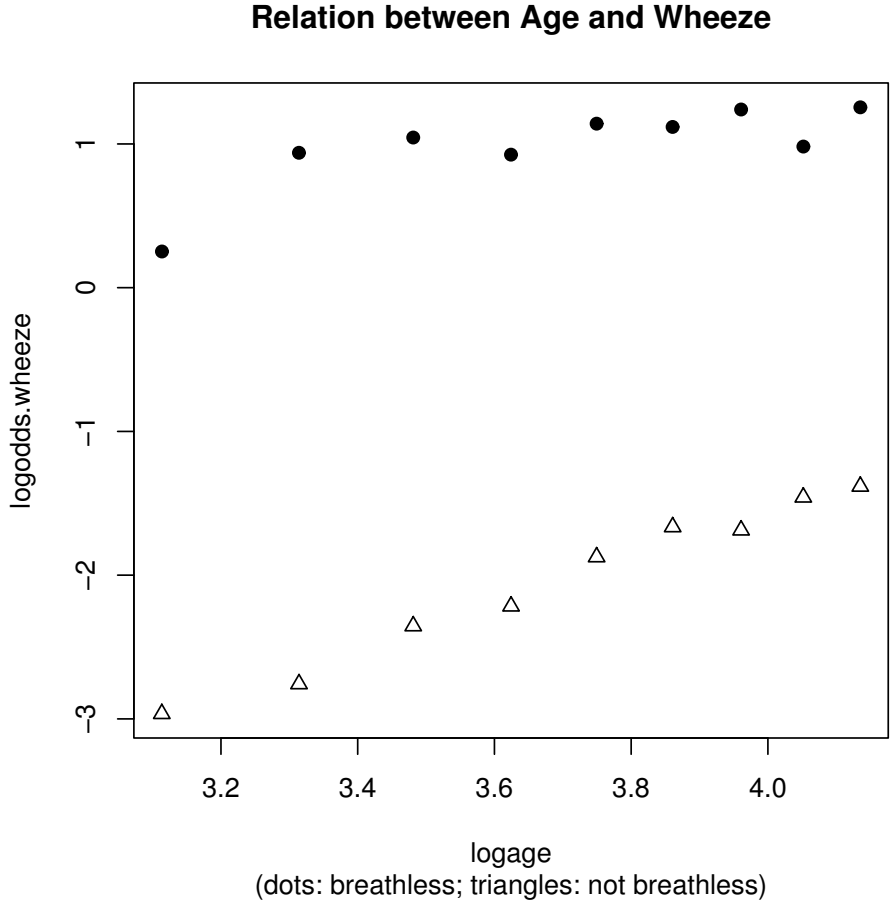
```
> # Data frame "Mining"
> mining
   wheeze total breath  age   logage
1       9    16      B 22.5 3.113515
2      23    32      B 27.5 3.314186
3      54    73      B 32.5 3.481240
4     121   169      B 37.5 3.624341
5     169   223      B 42.5 3.749504
6     269   357      B 47.5 3.860730
7     404   521      B 52.5 3.960813
8     406   558      B 57.5 4.051785
9     372   478      B 62.5 4.135167
10     95  1936      N 22.5 3.113515
11    105  1759      N 27.5 3.314186
12    177  2040      N 32.5 3.481240
13    257  2614      N 37.5 3.624341
14    273  2051      N 42.5 3.749504
15    324  2036      N 47.5 3.860730
16    245  1569      N 52.5 3.960813
17    225  1192      N 57.5 4.051785
18    132   658      N 62.5 4.135167

> logage<-log(age)
> logodds.wheeze<-log(wheeze/(total-wheeze))
> # Plot below, commands edited out
```

. . . continued

```
> # Model 1
> mining.glm<-glm(wheeze/total~breath+logage, binomial, weights=total)



> # Model 2
> mining.glm2<-glm(wheeze/total~breath*logage, binomial, weights=total)
> summary(mining.glm2)

Call:
glm(formula = wheeze/total ~ breath * logage, family = binomial,
    weights = total)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8391  -0.9016   0.1694   0.5459   1.6575

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.3371     0.9330  -0.361    0.718
breathN        -7.7325     0.9954  -7.768 7.94e-15 ***
logage          0.3694     0.2375   1.555    0.120
breathN:logage  1.2632     0.2548   4.957 7.15e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4596.676  on 17  degrees of freedom
Residual deviance:   14.996  on 14  degrees of freedom
AIC: 133.01

Number of Fisher Scoring iterations: 3
```

. . . continued

```
> anova(mining.glm,mining.glm2,test="Chi")
Analysis of Deviance Table

Model 1: wheeze/total ~ breath + logage
Model 2: wheeze/total ~ breath * logage
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1        15     40.241
2        14     14.996  1   25.245 5.048e-07
>
```



(a) Two models were fitted to the data. Only the second model (labelled Model 2) is summarised. Comment on the plot and in particular on the form of any logistic regression modelling that you might carry out. [5]

. . . continued

(b) Algebraically, give a complete specification of the generalized linear model fitted as Model 2 in the output. Match the parameters in your model with those whose estimates are provided in the output. [7]

(c) Comment on the goodness-of-fit of Model 2. [1]

(d) Describe the test performed in the analysis of deviance (obtained by the `anova` command). What does the result of this test suggest? [3]

(e) The hypothesis test of part (d) may be carried out by an alternative method. Identify the information in the output to enable you to carry out this alternative test. [1]

(f) Obtain a formula for the ratio comparing the log odds of wheezing in a breathless individual of age z years with the log odds of a comparably aged individual who is not breathless. What would be the consequence for this calculation under the first and second model fitted in the output? [3]