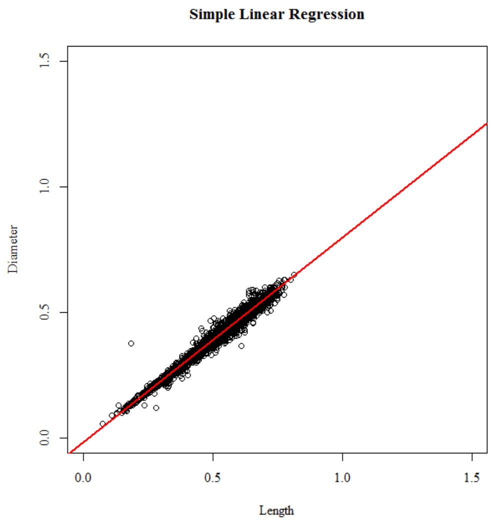
CODE FOR CS910 EXERCISE 3

Q1.

```
> abalone<-read.csv("abalone.data", header = FALSE,
col.names=c("Sex","Length","Diameter","Height","Whole.weight","Shucked.weight","Visce
ra.weight","Shell.weight","Rings"))
attach(abalone)
> plot(Length,Diameter,xlim=c(0,1.5),ylim = c(0,1.5),main="Simple Linear
Regression",family="serif")
> fit.simple.linear<-lm(Diameter~Length)
> abline(fit.simple.linear,col="red",lwd=2)
> summary(fit.simple.linear)
```




```
Call:
lm(formula = Diameter ~ Length)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.113017 -0.008703 -0.000549  0.008678  0.243553 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.019414   0.001113  -17.44   <2e-16 ***
Length       0.815461   0.002070  393.90   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01607 on 4175 degrees of freedom
Multiple R-squared:  0.9738,  Adjusted R-squared:  0.9738 
F-statistic: 1.552e+05 on 1 and 4175 DF,  p-value: < 2.2e-16
```

Q2.

```
> fit.multilinear<-lm(Whole.weight~Shucked.weight+Viscera.weight+Shell.weight)
> summary(fit.multilinear)

Call:
lm(formula = Whole.weight ~ Shucked.weight + Viscera.weight + 
    Shell.weight)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.54690 -0.01708 -0.00195  0.00903  0.51721 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    -0.007830   0.001452  -5.393 7.32e-08 ***
Shucked.weight  0.936560   0.009294 100.770  < 2e-16 ***
Viscera.weight  1.111650   0.021079  52.737  < 2e-16 ***
Shell.weight    1.252962   0.012802  97.876  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0469 on 4173 degrees of freedom
Multiple R-squared:  0.9909,  Adjusted R-squared:  0.9909 
F-statistic: 1.508e+05 on 3 and 4173 DF,  p-value: < 2.2e-16
```

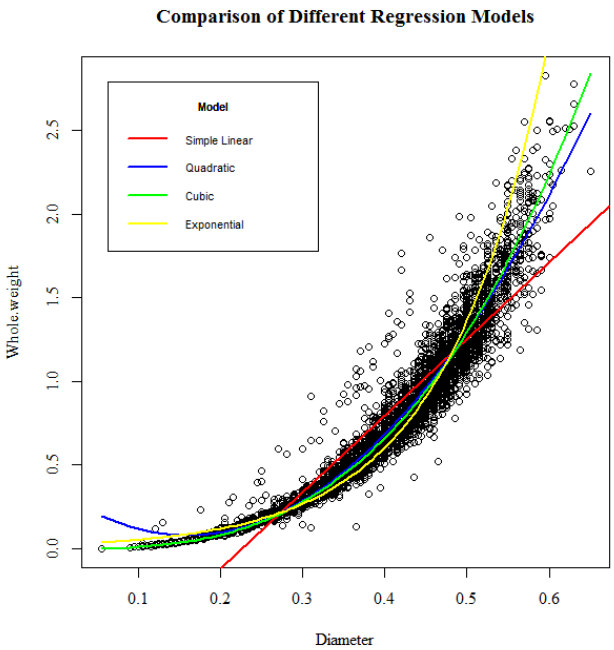
Q3.

```
> fit.a<-lm(Whole.weight~Diameter)
> fit.b<-lm(Whole.weight~Diameter+I(Diameter^2))
> fit.c<-lm(Whole.weight~I(Diameter^3)-1)
> fit.d<-lm(log(Whole.weight)~Diameter)
> summary(fit.a)

Call:
```

```
lm(formula = Whole.weight ~ Diameter)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.56745 -0.12307 -0.03997  0.07213  1.14105 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.03653    0.01216  -85.22   <2e-16 ***
Diameter     4.57308    0.02897  157.83   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1858 on 4175 degrees of freedom
Multiple R-squared:  0.8565,	Adjusted R-squared:  0.8564 
F-statistic: 2.491e+04 on 1 and 4175 DF,  p-value: < 2.2e-16

> summary(fit.b)

Call:
lm(formula = Whole.weight ~ Diameter + I(Diameter^2))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.66801 -0.06579 -0.00611  0.04589  0.97396 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)    0.34772    0.02353   14.78   <2e-16 ***
Diameter      -3.35552    0.12696  -26.43   <2e-16 ***
I(Diameter^2) 10.49681    0.16583   63.30   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1327 on 4174 degrees of freedom
Multiple R-squared:  0.9268,	Adjusted R-squared:  0.9267 
F-statistic: 2.641e+04 on 2 and 4174 DF,  p-value: < 2.2e-16

> summary(fit.c)

Call:
lm(formula = Whole.weight ~ I(Diameter^3) - 1)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.76061 -0.04998  0.00575  0.05708  0.99811 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
I(Diameter^3) 10.33761    0.02233   462.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1332 on 4176 degrees of freedom
Multiple R-squared:  0.9809,	Adjusted R-squared:  0.9809 
F-statistic: 2.143e+05 on 1 and 4176 DF,  p-value: < 2.2e-16

> summary(fit.d)

Call:
lm(formula = log(Whole.weight) ~ Diameter)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.91005 -0.09511  0.01132  0.12019  1.14050 
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.75098    0.01465  -256.1   <2e-16 ***
Diameter     8.11667    0.03490   232.6   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2238 on 4175 degrees of freedom
Multiple R-squared:  0.9284,  Adjusted R-squared:  0.9283
F-statistic: 5.41e+04 on 1 and 4175 DF,  p-value: < 2.2e-16
```

```
> plot(Diameter,Whole.weight,main="Comparison of Different Regression
Models",family="serif")
> abline(fit.a,col="red",lwd=2)
> lines(Diameter[order(Diameter)],fitted(fit.b)[order(Diameter)],col="blue",lwd=2)
> lines(Diameter[order(Diameter)],fitted(fit.c)[order(Diameter)],col="green",lwd=2)
>
lines(Diameter[order(Diameter)],exp(fitted(fit.d))[order(Diameter)],col="yellow",lwd=
2)
> legend("topleft",title="Model",legend = c("Simple
Linear","Quadratic","Cubic","Exponential"),lwd=c(2,2,2,2),col=c("red","blue","green",
"yellow"),cex=c(0.7,0.7,0.7,0.7),inset=0.05)
```



Q4.

```
> abalone$Age.class[abalone$Sex == "I"] <-0
> abalone$Age.class[abalone$Sex == "F"] <-1
> abalone$Age.class[abalone$Sex == "M"] <-1
> abalone$Age.class<-
factor(abalone$Age.class,levels=c(0,1),labels=c("Infant","Adult"))
> attach(abalone)

> fit.length<-glm(Age.class~Length,family = binomial(link="logit"))
> fit.whole.weight<-glm(Age.class~Whole.weight,family = binomial(link="logit"))
> fit.rings<-glm(Age.class~Rings,family = binomial(link="logit"))
> fit.full<-glm(Age.class~Length+Whole.weight+Rings,family = binomial(link="logit"))

> summary(fit.length)

Call:
glm(formula = Age.class ~ Length, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
```

```
-2.6677  -0.7024   0.4312   0.6697   2.7205

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -5.6347     0.2139  -26.35   <2e-16 ***
Length       12.6402     0.4237   29.83   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5244.9  on 4176  degrees of freedom
Residual deviance: 3849.9  on 4175  degrees of freedom
AIC: 3853.9

Number of Fisher Scoring iterations: 5

> summary(fit.whole.weight)

Call:
glm(formula = Age.class ~ Whole.weight, family = binomial(link = "logit"))

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-3.5916  -0.6897   0.2814   0.6350   2.0780  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -2.10113    0.09061  -23.19   <2e-16 ***
whole.weight  4.17141    0.13588   30.70   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5244.9  on 4176  degrees of freedom
Residual deviance: 3534.7  on 4175  degrees of freedom
AIC: 3538.7

Number of Fisher Scoring iterations: 5

> summary(fit.rings)

Call:
glm(formula = Age.class ~ Rings, family = binomial(link = "logit"))

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-3.6764  -0.8390   0.4618   0.7357   2.2249  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) -3.91107    0.17631  -22.18   <2e-16 ***
Rings        0.50799    0.01973   25.75   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5244.9  on 4176  degrees of freedom
Residual deviance: 4158.3  on 4175  degrees of freedom
AIC: 4162.3

Number of Fisher Scoring iterations: 5

> summary(fit.full)
```

```
Call:
glm(formula = Age.class ~ Length + Whole.weight + Rings, family = binomial(link =
"logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.8606  -0.6879   0.2077   0.6172   2.0568

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.27940    0.36156  -0.773     0.44
Length       -10.95785    1.21393  -9.027   <2e-16 ***
Whole.weight   6.56365    0.39869  16.463   <2e-16 ***
Rings          0.22230    0.02157  10.306   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5244.9  on 4176  degrees of freedom
Residual deviance: 3355.9  on 4173  degrees of freedom
AIC: 3363.9

Number of Fisher Scoring iterations: 6



> table(ifelse(predict(fit.length,type="response")>0.5,1,0),abalone$Age.class)

    Infant Adult
  0    722   285
  1    620  2550

> table(ifelse(predict(fit.whole.weight,type="response")>0.5,1,0),abalone$Age.class)

    Infant Adult
  0    868   379
  1    474  2456
> table(ifelse(predict(fit.rings,type="response")>0.5,1,0),abalone$Age.class)

    Infant Adult
  0    648   191
  1    694  2644
> table(ifelse(predict(fit.full,type="response")>0.5,1,0),abalone$Age.class)

    Infant Adult
  0    978   376
  1    364  2459
```

Q5.

```
> adult<-read.csv("adult.data",header = F,col.names =
c("age","workclass","fnlwgt","education","education-num","marital-
status","occupation","relationship","race","sex","capital-gain","capital-
loss","hours-per-week","native-country","class"))
> fit.full<-
glm(sex~age+workclass+fnlwgt+education+marital.status+occupation+relationship+race+ca
pital.gain+capital.loss+hours.per.week+native.country+class,data=adult, family =
binomial())
> summary(fit.full)

Call:
glm(formula = sex ~ age + workclass + fnlwgt + education + marital.status +
```

```
    occupation + relationship + race + capital.gain + capital.loss + 
    hours.per.week + native.country + class, family = binomial(), 
    data = adult)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-4.1635  -0.4325   0.0079   0.3321   3.6235  

Coefficients: (1 not defined because of singularities)
                                     Estimate Std. Error z value Pr(>|z|)    
(Intercept)                         8.501e+00  1.055e+00   8.061 7.55e-16 ***
age                                -1.585e-03  1.913e-03  -0.829 0.407329    
workclass Federal-gov               2.720e+00  1.838e-01  14.799  < 2e-16 ***
workclass Local-gov                 2.112e+00  1.658e-01  12.743  < 2e-16 ***
workclass Never-worked              1.670e+00  1.106e+00   1.509 0.131184    
workclass Private                   2.355e+00  1.475e-01  15.961  < 2e-16 ***
workclass Self-emp-inc              3.373e+00  2.112e-01  15.975  < 2e-16 ***
workclass Self-emp-not-inc          3.082e+00  1.725e-01  17.870  < 2e-16 ***
workclass State-gov                 2.550e+00  1.714e-01  14.881  < 2e-16 ***
workclass Without-pay               1.913e+00  9.668e-01   1.979 0.047860 *  
fnlwgt                              1.186e-06  1.707e-07   6.949 3.67e-12 ***
education 11th                     -1.448e-01  1.284e-01  -1.128 0.259402    
education 12th                      1.654e-01  1.653e-01   1.001 0.317012    
education 1st-4th                   1.886e-01  2.995e-01   0.630 0.528944    
education 5th-6th                  -8.987e-02  2.393e-01  -0.376 0.707265    
education 7th-8th                   3.255e-02  1.784e-01   0.182 0.855253    
education 9th                       1.784e-01  1.763e-01   1.012 0.311611    
education Assoc-acdm               -1.416e-01  1.393e-01  -1.017 0.309332    
education Assoc-voc                -3.594e-01  1.370e-01  -2.623 0.008724 ** 
education Bachelors                 2.213e-02  1.133e-01   0.195 0.845181    
education Doctorate                -3.589e-01  2.323e-01  -1.545 0.122257    
education HS-grad                  -1.752e-01  1.042e-01  -1.681 0.092672 .  
education Masters                  -2.342e-01  1.390e-01  -1.685 0.091917 .  
education Preschool                 2.722e-01  4.325e-01   0.629 0.529071    
education Prof-school               2.295e-01  2.119e-01   1.083 0.278699    
education Some-college             -1.994e-01  1.054e-01  -1.893 0.058378 .  
marital.status Married-AF-spouse   -3.556e+00  5.277e+00  -0.674 0.500393    
marital.status Married-civ-spouse   9.666e-02  1.679e-01   0.576 0.564724    
marital.status Married-spouse-absent  2.890e-01  1.291e-01   2.238 0.025194 *  
marital.status Never-married        5.414e-01  5.302e-02  10.211  < 2e-16 ***
marital.status Separated            4.773e-02  8.702e-02   0.549 0.583344    
marital.status Widowed             -1.060e+00  1.088e-01  -9.743  < 2e-16 ***
occupation Adm-clerical            -3.357e+00  1.418e-01 -23.682  < 2e-16 ***
occupation Armed-Forces             1.028e+01  1.608e+02   0.064 0.949002    
occupation Craft-repair            -1.380e-01  1.546e-01  -0.892 0.372243    
occupation Exec-managerial         -2.613e+00  1.455e-01 -17.955  < 2e-16 ***
occupation Farming-fishing         -2.362e-01  2.095e-01  -1.128 0.259531    
occupation Handlers-cleaners       -5.073e-01  1.623e-01  -3.125 0.001778 ** 
occupation Machine-op-inspct       -1.821e+00  1.487e-01 -12.253  < 2e-16 ***
occupation Other-service           -2.561e+00  1.398e-01 -18.319  < 2e-16 ***
occupation Priv-house-serv         -5.012e+00  4.000e-01 -12.530  < 2e-16 ***
occupation Prof-specialty          -2.803e+00  1.474e-01 -19.019  < 2e-16 ***
occupation Protective-serv         -9.705e-01  2.028e-01  -4.785 1.71e-06 ***
occupation Sales                   -2.565e+00  1.420e-01 -18.065  < 2e-16 ***
occupation Tech-support            -2.506e+00  1.644e-01 -15.245  < 2e-16 ***
occupation Transport-moving                NA         NA      NA       NA    
relationship Not-in-family         -9.174e+00  1.018e+00  -9.012  < 2e-16 ***
relationship Other-relative        -9.105e+00  1.018e+00  -8.946  < 2e-16 ***
relationship Own-child             -9.073e+00  1.018e+00  -8.913  < 2e-16 ***
relationship Unmarried             -1.042e+01  1.019e+00 -10.222  < 2e-16 ***
relationship Wife                  -1.623e+01  1.230e+00 -13.192  < 2e-16 ***
race Asian-Pac-Islander            -1.173e-01  2.269e-01  -0.517 0.605299    
race Black                         -3.574e-01  1.866e-01  -1.915 0.055530 .  
race Other                         -2.774e-01  2.583e-01  -1.074 0.282890    
race White                         -5.016e-02  1.798e-01  -0.279 0.780338    
capital.gain                       -1.149e-07  4.151e-06  -0.028 0.977913    
```

```
capital.loss                            5.966e-05  5.404e-05   1.104 0.269603    
hours.per.week                          2.129e-02  1.643e-03  12.956  < 2e-16 ***
native.country Cambodia                -2.902e-02  9.140e-01  -0.032 0.974674    
native.country Canada                  -2.785e-01  3.336e-01  -0.835 0.403902    
native.country China                   -1.009e-01  4.319e-01  -0.233 0.815374    
native.country Columbia                -6.207e-01  4.082e-01  -1.521 0.128384    
native.country Cuba                    -1.418e+00  4.038e-01  -3.512 0.000445 ***
native.country Dominican-Republic      -1.170e+00  3.697e-01  -3.163 0.001559 ** 
native.country Ecuador                 -5.671e-01  6.076e-01  -0.933 0.350644    
native.country El-Salvador             -1.157e-01  3.235e-01  -0.358 0.720710    
native.country England                 -9.662e-02  3.472e-01  -0.278 0.780761    
native.country France                  -7.299e-01  5.894e-01  -1.238 0.215560    
native.country Germany                 -9.840e-01  3.258e-01  -3.020 0.002525 ** 
native.country Greece                   4.557e-01  7.815e-01   0.583 0.559838    
native.country Guatemala               -8.307e-02  4.187e-01  -0.198 0.842711    
native.country Haiti                    5.879e-02  4.478e-01   0.131 0.895560    
native.country Holand-Netherlands      -1.475e+01  5.354e+02  -0.028 0.978018    
native.country Honduras                -3.166e-01  8.036e-01  -0.394 0.693619    
native.country Hong                    -3.289e-01  1.015e+00  -0.324 0.745933    
native.country Hungary                 -7.179e-01  9.404e-01  -0.763 0.445244    
native.country India                    1.035e+00  4.612e-01   2.245 0.024782 *  
native.country Iran                     9.923e-01  6.290e-01   1.577 0.114682    
native.country Ireland                 -1.065e+00  5.862e-01  -1.816 0.069310 .  
native.country Italy                   -8.404e-01  4.871e-01  -1.725 0.084485 .  
native.country Jamaica                 -8.710e-01  3.628e-01  -2.401 0.016369 *  
native.country Japan                   -5.581e-01  4.326e-01  -1.290 0.196991    
native.country Laos                    -1.175e+00  8.735e-01  -1.346 0.178422    
native.country Mexico                  -2.511e-01  2.019e-01  -1.243 0.213732    
native.country Nicaragua               -3.214e-02  5.150e-01  -0.062 0.950232    
native.country Outlying-US(Guam-USVI-etc) -7.826e-01  6.863e-01  -1.140 0.254143    
native.country Peru                    -1.001e+00  5.253e-01  -1.905 0.056748 .  
native.country Philippines             -4.245e-01  2.860e-01  -1.484 0.137751    
native.country Poland                  -6.124e-01  4.566e-01  -1.341 0.179850    
native.country Portugal                -9.180e-01  6.463e-01  -1.420 0.155493    
native.country Puerto-Rico             -6.390e-01  3.287e-01  -1.944 0.051883 .  
native.country Scotland                 1.135e+00  1.002e+00   1.132 0.257545    
native.country South                   -4.188e-01  3.848e-01  -1.088 0.276408    
native.country Taiwan                  -4.802e-02  4.874e-01  -0.099 0.921532    
native.country Thailand                -1.307e+00  7.804e-01  -1.675 0.093980 .  
native.country Trinadad&Tobago         -2.344e+00  1.109e+00  -2.114 0.034542 *  
native.country United-States           -5.013e-01  1.379e-01  -3.635 0.000278 ***
native.country Vietnam                 -1.229e-01  3.932e-01  -0.313 0.754592    
native.country Yugoslavia              -2.595e-01  1.021e+00  -0.254 0.799256    
class >50K                              8.026e-01  8.068e-02   9.948  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 41336  on 32560  degrees of freedom
Residual deviance: 18942  on 32462  degrees of freedom
AIC: 19140

Number of Fisher Scoring iterations: 12


> sum(diag(table(ifelse(predict(fit.full,type="response")>0.5,1,0),adult$sex)))/32561
[1] 0.8481312


> fit.reduce.age<-
glm(sex~workclass+fnlwgt+education+marital.status+occupation+relationship+race+capita
l.gain+capital.loss+hours.per.week+native.country+class,data=adult, family = 
binomial())
```

```
>
sum(diag(table(ifelse(predict(fit.reduce.age,type="response")>0.5,1,0),adult$sex)))/3
2561
[1] 0.8479162
> fit.reduce.workclass<-
glm(sex~age+fnlwgt+education+marital.status+occupation+relationship+race+capital.gain
+capital.loss+hours.per.week+native.country+class,data=adult, family = binomial())
>
sum(diag(table(ifelse(predict(fit.reduce.workclass,type="response")>0.5,1,0),adult$se
x)))/32561
[1] 0.8463806
> fit.reduce.fnlwgt<-
glm(sex~age+workclass+education+marital.status+occupation+relationship+race+capital.g
ain+capital.loss+hours.per.week+native.country+class,data=adult, family = binomial())
>
sum(diag(table(ifelse(predict(fit.reduce.fnlwgt,type="response")>0.5,1,0),adult$sex))
)/32561
[1] 0.8459507
> fit.reduce.education<-
glm(sex~age+workclass+fnlwgt+marital.status+occupation+relationship+race+capital.gain
+capital.loss+hours.per.week+native.country+class,data=adult, family = binomial())
>
sum(diag(table(ifelse(predict(fit.reduce.education,type="response")>0.5,1,0),adult$se
x)))/32561
[1] 0.8474555
> fit.reduce.marital.status<-
glm(sex~age+workclass+fnlwgt+education+occupation+relationship+race+capital.gain+capi
tal.loss+hours.per.week+native.country+class,data=adult, family = binomial())
>
sum(diag(table(ifelse(predict(fit.reduce.marital.status,type="response")>0.5,1,0),adu
lt$sex)))/32561
[1] 0.845275
> fit.reduce.occupation<-
glm(sex~age+workclass+fnlwgt+education+marital.status+relationship+race+capital.gain+
capital.loss+hours.per.week+native.country+class,data=adult, family = binomial())
>
sum(diag(table(ifelse(predict(fit.reduce.occupation,type="response")>0.5,1,0),adult$s
ex)))/32561
[1] 0.8106324
> fit.reduce.relationship<-
glm(sex~age+workclass+fnlwgt+education+marital.status+occupation+race+capital.gain+ca
pital.loss+hours.per.week+native.country+class,data=adult, family = binomial())
>
sum(diag(table(ifelse(predict(fit.reduce.relationship,type="response")>0.5,1,0),adult
$sex)))/32561
[1] 0.7936181
> fit.reduce.race<-
glm(sex~age+workclass+fnlwgt+education+marital.status+occupation+relationship+capital
.gain+capital.loss+hours.per.week+native.country+class,data=adult, family =
binomial())
>
sum(diag(table(ifelse(predict(fit.reduce.race,type="response")>0.5,1,0),adult$sex)))/
32561
[1] 0.8481005
> fit.reduce.capital.gain<-
glm(sex~age+workclass+fnlwgt+education+marital.status+occupation+relationship+race+ca
pital.loss+hours.per.week+native.country+class,data=adult, family = binomial())
>
sum(diag(table(ifelse(predict(fit.reduce.capital.gain,type="response")>0.5,1,0),adult
$sex)))/32561
[1] 0.8481312
> fit.reduce.capital.loss<-
glm(sex~age+workclass+fnlwgt+education+marital.status+occupation+relationship+race+ca
pital.gain+hours.per.week+native.country+class,data=adult, family = binomial())
```

```
> sum(diag(table(ifelse(predict(fit.reduce.capital.loss,type="response")>0.5,1,0),adult$sex)))/32561
[1] 0.8480698
> fit.reduce.hours.per.week<-glm(sex~age+workclass+fnlwgt+education+marital.status+occupation+relationship+race+capital.gain+capital.loss+native.country+class,data=adult, family = binomial())
> sum(diag(table(ifelse(predict(fit.reduce.hours.per.week,type="response")>0.5,1,0),adult$sex)))/32561
[1] 0.8430331
> fit.reduce.native.country<-glm(sex~age+workclass+fnlwgt+education+marital.status+occupation+relationship+race+capital.gain+capital.loss+hours.per.week+class,data=adult, family = binomial())
> sum(diag(table(ifelse(predict(fit.reduce.native.country,type="response")>0.5,1,0),adult$sex)))/32561
[1] 0.8461349
> fit.reduce.class<-glm(sex~age+workclass+fnlwgt+education+marital.status+occupation+relationship+race+capital.gain+capital.loss+hours.per.week+native.country,data=adult, family = binomial())
> sum(diag(table(ifelse(predict(fit.reduce.class,type="response")>0.5,1,0),adult$sex)))/32561
[1] 0.8469027
```

```
> fit.reduced<-glm(sex~relationship,data=adult, family = binomial())
> sum(diag(table(ifelse(predict(fit.reduced,type="response")>0.5,1,0),adult$sex)))/32561
[1] 0.7744234
> fit.reduced<-glm(sex~occupation+relationship,data=adult, family = binomial())
> sum(diag(table(ifelse(predict(fit.reduced,type="response")>0.5,1,0),adult$sex)))/32561
[1] 0.8220264
> fit.reduced<-glm(sex~occupation+relationship+hours.per.week,data=adult, family = binomial())
> sum(diag(table(ifelse(predict(fit.reduced,type="response")>0.5,1,0),adult$sex)))/32561
[1] 0.83772
> fit.reduced<-glm(sex~occupation+relationship+hours.per.week+marital.status,data=adult, family = binomial())
> sum(diag(table(ifelse(predict(fit.reduced,type="response")>0.5,1,0),adult$sex)))/32561
[1] 0.8396241
> fit.reduced<-glm(sex~occupation+relationship+hours.per.week+native.country,data=adult, family = binomial())
> sum(diag(table(ifelse(predict(fit.reduced,type="response")>0.5,1,0),adult$sex)))/32561
[1] 0.8399619
> table(ifelse(predict(fit.reduced,type="response")>0.5,1,0),adult$sex)

    Female  Male
  0   8733  3173
  1   2038 18617
> summary(fit.reduced)
```

```
Call:
glm(formula = sex ~ occupation + relationship + hours.per.week +
    native.country, family = binomial(), data = adult)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.2693  -0.5134   0.0108   0.3791   3.7177

Coefficients:
                                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                                8.619693   1.012230   8.516  < 2e-16 ***
occupation Adm-clerical                   -0.906405   0.077133 -11.751  < 2e-16 ***
occupation Armed-Forces                   13.229864 160.916364   0.082 0.934475
occupation Craft-repair                    2.305508   0.101183  22.786  < 2e-16 ***
occupation Exec-managerial                 0.006659   0.082335   0.081 0.935544
occupation Farming-fishing                 2.310099   0.168897  13.678  < 2e-16 ***
occupation Handlers-cleaners               1.945980   0.111245  17.493  < 2e-16 ***
occupation Machine-op-inspct               0.534612   0.091225   5.860 4.62e-09 ***
occupation Other-service                  -0.122921   0.074229  -1.656 0.097725 .
occupation Priv-house-serv                -2.653857   0.379211  -6.998 2.59e-12 ***
occupation Prof-specialty                 -0.134119   0.078572  -1.707 0.087833 .
occupation Protective-serv                 1.334663   0.158873   8.401  < 2e-16 ***
occupation Sales                           0.001974   0.076969   0.026 0.979539
occupation Tech-support                    0.037348   0.110570   0.338 0.735530
occupation Transport-moving                2.343232   0.144081  16.263  < 2e-16 ***
relationship Not-in-family                -9.223108   1.000380  -9.220  < 2e-16 ***
relationship Other-relative               -9.157551   1.002978  -9.130  < 2e-16 ***
relationship Own-child                    -8.945313   1.000697  -8.939  < 2e-16 ***
relationship Unmarried                   -10.697977   1.001220 -10.685  < 2e-16 ***
relationship Wife                        -16.027568   1.225894 -13.074  < 2e-16 ***
hours.per.week                             0.025030   0.001560  16.040  < 2e-16 ***
native.country Cambodia                    0.251847   0.927612   0.272 0.786006
native.country Canada                     -0.298985   0.323404  -0.924 0.355230
native.country China                      -0.204863   0.402560  -0.509 0.610821
native.country Columbia                   -0.487228   0.393363  -1.239 0.215485
native.country Cuba                       -1.336411   0.387602  -3.448 0.000565 ***
native.country Dominican-Republic         -1.109256   0.359721  -3.084 0.002045 **
native.country Ecuador                    -0.635108   0.587813  -1.080 0.279938
native.country El-Salvador                 0.159630   0.311291   0.513 0.608092
native.country England                    -0.136229   0.336883  -0.404 0.685933
native.country France                     -0.588852   0.559313  -1.053 0.292427
native.country Germany                    -1.012714   0.318086  -3.184 0.001454 **
native.country Greece                      0.502286   0.771791   0.651 0.515172
native.country Guatemala                   0.247185   0.410382   0.602 0.546955
native.country Haiti                      -0.262615   0.442292  -0.594 0.552673
native.country Holand-Netherlands        -14.564011 535.411193  -0.027 0.978299
native.country Honduras                    0.024914   0.781700   0.032 0.974575
native.country Hong                       -0.096038   1.016551  -0.094 0.924732
native.country Hungary                    -0.749883   0.893923  -0.839 0.401544
native.country India                       1.169370   0.432769   2.702 0.006891 **
native.country Iran                        0.795229   0.581408   1.368 0.171386
native.country Ireland                    -0.997506   0.593866  -1.680 0.093019 .
native.country Italy                      -0.750987   0.462637  -1.623 0.104531
native.country Jamaica                    -1.031841   0.351403  -2.936 0.003321 **
native.country Japan                      -0.499614   0.421163  -1.186 0.235515
native.country Laos                       -0.968648   0.862312  -1.123 0.261304
native.country Mexico                      0.057655   0.191835   0.301 0.763761
native.country Nicaragua                   0.044909   0.504028   0.089 0.929003
native.country Outlying-US(Guam-USVI-etc) -1.047385   0.677234  -1.547 0.121969
native.country Peru                       -0.872517   0.520386  -1.677 0.093607 .
native.country Philippines                -0.432030   0.256205  -1.686 0.091744 .
native.country Poland                     -0.717931   0.442770  -1.621 0.104921
native.country Portugal                   -0.812226   0.619972  -1.310 0.190161
native.country Puerto-Rico                -0.620297   0.318019  -1.951 0.051116 .
native.country Scotland                    0.968902   0.981941   0.987 0.323779
native.country South                      -0.387181   0.358695  -1.079 0.280402
```

```
native.country Taiwan                     0.174196   0.466342   0.374 0.708748
native.country Thailand                  -1.167824   0.751926  -1.553 0.120397
native.country Trinadad&Tobago           -2.117964   1.086985  -1.948 0.051358 .
native.country United-States             -0.509566   0.134099  -3.800 0.000145 ***
native.country Vietnam                   -0.001102   0.376144  -0.003 0.997662
native.country Yugoslavia                -0.324869   1.022339  -0.318 0.750659
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 41336  on 32560  degrees of freedom
Residual deviance: 19686  on 32499  degrees of freedom
AIC: 19810

Number of Fisher Scoring iterations: 12
```