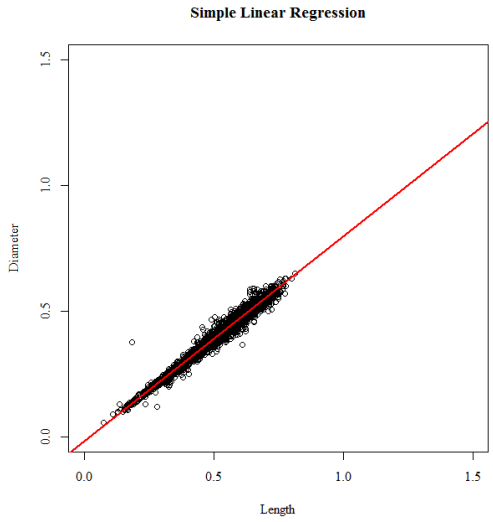
Q1.

```
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)
```

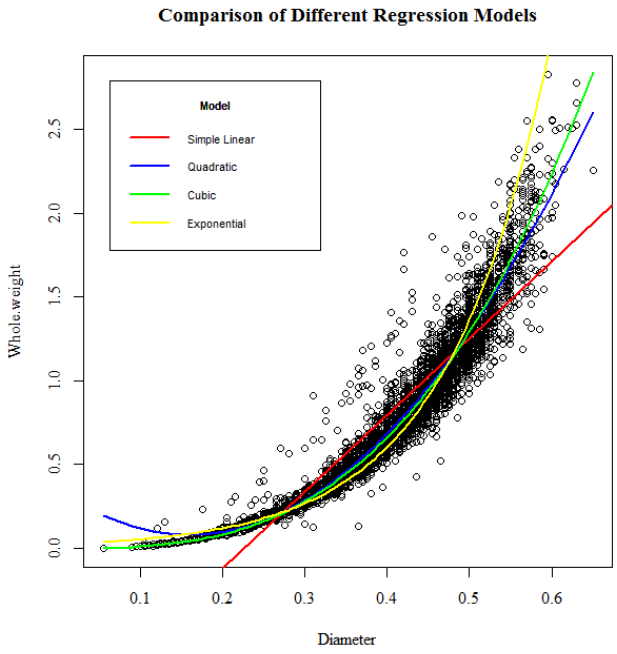
```
Call:
lm(formula = Whole.weight ~ I(Diameter^3) - 1)

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.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

training.data<-data.frame(abalone$Length,abalone$Whole.weight,abalone$Rings)
training.data$prob.length<-predict(fit.length,newdata =
training.data,type="response")
training.data$prob.whole.weight<-predict(fit.Whole.weight,newdata =
training.data,type="response")
training.data$prob.rings<-predict(fit.rings,newdata = training.data,type="response")
training.data$prob.full<-predict(fit.full,newdata = training.data,type="response")
training.data$pre.result.length[training.data$prob.length>0.5]<-1
training.data$pre.result.length[training.data$prob.length<=0.5]<-0
training.data$pre.result.whole.weight[training.data$prob.whole.weight>0.5]<-1
training.data$pre.result.whole.weight[training.data$prob.whole.weight<=0.5]<-0
training.data$pre.result.rings[training.data$prob.rings>0.5]<-1
training.data$pre.result.rings[training.data$prob.rings<=0.5]<-0
training.data$pre.result.full[training.data$prob.full>0.5]<-1
training.data$pre.result.full[training.data$prob.full<=0.5]<-0
training.data$pre.result.length<-
factor(training.data$pre.result.length,levels=c(0,1),labels=c("Infant","Adult"))
training.data$pre.result.whole.weight<-
factor(training.data$pre.result.whole.weight,levels=c(0,1),labels=c("Infant","Adult")
)
training.data$pre.result.rings<-
factor(training.data$pre.result.rings,levels=c(0,1),labels=c("Infant","Adult"))
training.data$pre.result.full<-
factor(training.data$pre.result.full,levels=c(0,1),labels=c("Infant","Adult"))
compare<-
data.frame(abalone$Age.class,training.data$pre.result.length,training.data$pre.result
.whole.weight,training.data$pre.result.rings,training.data$pre.result.full)
library(dplyr)
attach(compare)
summary(filter(compare,abalone.Age.class == training.data.pre.result.length))
 abalone.Age.class training.data.pre.result.length
training.data.pre.result.whole.weight
 Infant: 722      Infant: 722                 Infant: 819
 Adult :2550      Adult :2550                 Adult :2453
 training.data.pre.result.rings training.data.pre.result.full
 Infant: 652                 Infant: 841
 Adult :2620                 Adult :2431
summary(filter(compare,abalone.Age.class == training.data.pre.result.whole.weight))
 abalone.Age.class training.data.pre.result.length
training.data.pre.result.whole.weight
 Infant: 868      Infant: 725                 Infant: 868
 Adult :2456      Adult :2599                 Adult :2456
 training.data.pre.result.rings training.data.pre.result.full
 Infant: 687                 Infant: 904
 Adult :2637                 Adult :2420
```

```
summary(filter(compare,abalone.Age.class == training.data.pre.result.rings))
 abalone.Age.class training.data.pre.result.length
training.data.pre.result.whole.weight
 Infant: 648      Infant: 746                 Infant: 875
 Adult :2644      Adult :2546                 Adult :2417
 training.data.pre.result.rings training.data.pre.result.full
 Infant: 648                 Infant: 873
 Adult :2644                 Adult :2419
summary(filter(compare,abalone.Age.class == training.data.pre.result.full))
 abalone.Age.class training.data.pre.result.length
training.data.pre.result.whole.weight
 Infant: 978      Infant: 750                 Infant: 907
 Adult :2459      Adult :2687                 Adult :2530
 training.data.pre.result.rings training.data.pre.result.full
 Infant: 688                 Infant: 978
 Adult :2749                 Adult :2459
summary(compare)
 abalone.Age.class training.data.pre.result.length
training.data.pre.result.whole.weight
 Infant:1342      Infant:1007                 Infant:1247
 Adult :2835      Adult :3170                 Adult :2930
 training.data.pre.result.rings training.data.pre.result.full
 Infant: 839                 Infant:1354
 Adult :3338                 Adult :2823
```

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"))
adult$workclass[adult$workclass == " ?"]<-NA
adult$occupation[adult$occupation == " ?"]<-NA
adult$native.country[adult$native.country == " ?"]<-NA
adult.new<-na.omit(adult)
adult.new$sex[adult.new$sex == " Male"]<-1
adult.new$sex[adult.new$sex == " Female"]<-0
adult.new$sex<-factor(adult.new$sex,levels=c(0,1),labels = c("Female","Male"))
attach(adult.new)
fit.full<-
glm(sex~age+workclass+fnlwgt+education+education.num+marital.status+occupation+relati
onship+race+capital.gain+capital.loss+hours.per.week+native.country+class,data=adult.
new, family = binomial())
summary(fit.full)

Call:
glm(formula = sex ~ age + workclass + fnlwgt + education.num +
    marital.status + occupation + relationship + race + capital.gain +
    capital.loss + hours.per.week + native.country + class, family = binomial(),
    data = adult.new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.1504  -0.3979   0.0081   0.3191   3.7720

Coefficients:
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                        8.123e+00  1.589e+00   5.113 3.17e-07 ***
age                               -4.227e-03  2.009e-03  -2.104 0.035343 *
workclass Local-gov                -6.306e-01  1.338e-01  -4.711 2.46e-06 ***
workclass Private                  -4.113e-01  1.119e-01  -3.676 0.000237 ***
workclass Self-emp-inc              6.608e-01  1.869e-01   3.535 0.000407 ***
workclass Self-emp-not-inc          3.209e-01  1.413e-01   2.270 0.023199 *
workclass State-gov                -2.153e-01  1.390e-01  -1.549 0.121453
```

```
workclass Without-pay                   -8.453e-01  9.887e-01  -0.855 0.392620
fnlwgt                                 1.152e-06  1.797e-07   6.408 1.48e-10 ***
education.num                         -1.150e-04  9.895e-03  -0.012 0.990731
marital.status Married-AF-spouse         -3.409e+00  4.950e+00  -0.689 0.490986
marital.status Married-civ-spouse         3.162e-01  1.795e-01   1.762 0.078148 .
marital.status Married-spouse-absent      2.041e-01  1.383e-01   1.476 0.140030
marital.status Never-married              5.409e-01  5.450e-02   9.925  < 2e-16 ***
marital.status Separated                  4.941e-02  9.108e-02   0.542 0.587477
marital.status Widowed                   -1.051e+00  1.182e-01  -8.888  < 2e-16 ***
occupation Armed-Forces                   1.359e+01  1.622e+02   0.084 0.933229
occupation Craft-repair                   3.251e+00  9.689e-02  33.554  < 2e-16 ***
occupation Exec-managerial                7.861e-01  7.516e-02  10.459  < 2e-16 ***
occupation Farming-fishing                3.173e+00  1.727e-01  18.380  < 2e-16 ***
occupation Handlers-cleaners              2.876e+00  1.072e-01  26.817  < 2e-16 ***
occupation Machine-op-inspct              1.572e+00  8.621e-02  18.232  < 2e-16 ***
occupation Other-service                  8.353e-01  6.693e-02  12.479  < 2e-16 ***
occupation Priv-house-serv               -1.525e+00  3.812e-01  -4.001 6.31e-05 ***
occupation Prof-specialty                 6.153e-01  7.452e-02   8.257  < 2e-16 ***
occupation Protective-serv                2.379e+00  1.615e-01  14.730  < 2e-16 ***
occupation Sales                          8.143e-01  6.951e-02  11.714  < 2e-16 ***
occupation Tech-support                   8.713e-01  1.065e-01   8.179 2.85e-16 ***
occupation Transport-moving               3.381e+00  1.427e-01  23.696  < 2e-16 ***
relationship Not-in-family               -8.867e+00  1.020e+00  -8.690  < 2e-16 ***
relationship Other-relative              -8.837e+00  1.020e+00  -8.664  < 2e-16 ***
relationship Own-child                   -8.789e+00  1.021e+00  -8.613  < 2e-16 ***
relationship Unmarried                   -1.011e+01  1.022e+00  -9.893  < 2e-16 ***
relationship Wife                        -1.676e+01  1.420e+00 -11.803  < 2e-16 ***
race Asian-Pac-Islander                   3.799e-02  2.453e-01   0.155 0.876920
race Black                               -3.061e-01  1.997e-01  -1.533 0.125293
race Other                               -8.536e-02  2.815e-01  -0.303 0.761716
race White                                2.223e-02  1.924e-01   0.116 0.908022
capital.gain                             -1.654e-08  4.234e-06  -0.004 0.996884
capital.loss                              7.215e-05  5.652e-05   1.277 0.201753
hours.per.week                            2.181e-02  1.770e-03  12.322  < 2e-16 ***
native.country Canada                    -9.565e-01  1.234e+00  -0.775 0.438375
native.country China                     -8.430e-01  1.251e+00  -0.674 0.500403
native.country Columbia                  -1.338e+00  1.252e+00  -1.068 0.285450
native.country Cuba                      -2.046e+00  1.250e+00  -1.637 0.101646
native.country Dominican-Republic        -1.735e+00  1.237e+00  -1.402 0.160840
native.country Ecuador                   -1.290e+00  1.331e+00  -0.969 0.332451
native.country El-Salvador               -7.454e-01  1.227e+00  -0.607 0.543610
native.country England                   -7.779e-01  1.236e+00  -0.629 0.529208
native.country France                    -1.123e+00  1.335e+00  -0.841 0.400141
native.country Germany                   -1.608e+00  1.231e+00  -1.306 0.191444
native.country Greece                    -3.180e-01  1.419e+00  -0.224 0.822600
native.country Guatemala                 -7.103e-01  1.254e+00  -0.566 0.571103
native.country Haiti                     -4.984e-01  1.267e+00  -0.394 0.693927
native.country Holand-Netherlands        -1.550e+01  5.354e+02  -0.029 0.976901
native.country Honduras                  -9.707e-01  1.430e+00  -0.679 0.497177
native.country Hong                      -9.338e-01  1.554e+00  -0.601 0.547816
native.country Hungary                   -1.296e+00  1.506e+00  -0.861 0.389228
native.country India                      3.028e-01  1.254e+00   0.241 0.809213
native.country Iran                       6.190e-01  1.360e+00   0.455 0.649085
native.country Ireland                   -1.750e+00  1.320e+00  -1.326 0.184890
native.country Italy                     -1.430e+00  1.291e+00  -1.107 0.268265
native.country Jamaica                   -1.482e+00  1.238e+00  -1.197 0.231173
native.country Japan                     -1.185e+00  1.257e+00  -0.942 0.345958
native.country Laos                      -1.826e+00  1.458e+00  -1.252 0.210396
native.country Mexico                    -7.786e-01  1.199e+00  -0.649 0.516247
native.country Nicaragua                 -7.222e-01  1.289e+00  -0.560 0.575359
native.country Outlying-US(Guam-USVI-etc) -1.429e+00  1.367e+00  -1.045 0.295845
native.country Peru                      -1.858e+00  1.303e+00  -1.426 0.153867
native.country Philippines               -1.254e+00  1.202e+00  -1.043 0.296988
native.country Poland                    -1.406e+00  1.271e+00  -1.106 0.268691
native.country Portugal                  -1.694e+00  1.365e+00  -1.241 0.214775
native.country Puerto-Rico               -1.194e+00  1.228e+00  -0.973 0.330665
```

```
native.country Scotland                  4.156e-01  1.558e+00   0.267 0.789719
native.country South                    -1.004e+00  1.238e+00  -0.811 0.417204
native.country Taiwan                   -6.008e-01  1.312e+00  -0.458 0.646998
native.country Thailand                 -1.898e+00  1.412e+00  -1.344 0.179021
native.country Trinadad&Tobago           -2.651e+00  1.636e+00  -1.621 0.105038
native.country United-States             -1.182e+00  1.190e+00  -0.993 0.320777
native.country Vietnam                  -9.063e-01  1.233e+00  -0.735 0.462428
native.country Yugoslavia                -9.325e-01  1.562e+00  -0.597 0.550571
class >50K                              8.366e-01  8.132e-02  10.289  < 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: 38009  on 30161  degrees of freedom
Residual deviance: 17141  on 30080  degrees of freedom
AIC: 17305

Number of Fisher Scoring iterations: 12


fit.reduced<-
glm(sex~workclass+fnlwgt+marital.status+occupation+relationship+race+hours.per.week+n
ative.country+class,data=adult.new,family = binomial())
summary(fit.reduced)

Call:
glm(formula = sex ~ workclass + fnlwgt + marital.status + occupation +
    relationship + race + hours.per.week + native.country + class,
    family = binomial(), data = adult.new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.1456  -0.3996   0.0081   0.3206   3.7810

Coefficients:
                                       Estimate Std. Error z value Pr(>|z|)
(Intercept)                           7.921e+00  1.583e+00   5.003 5.63e-07 ***
workclass Local-gov                   -6.309e-01  1.338e-01  -4.715 2.42e-06 ***
workclass Private                     -4.013e-01  1.117e-01  -3.594 0.000326 ***
workclass Self-emp-inc                 6.526e-01  1.867e-01   3.495 0.000474 ***
workclass Self-emp-not-inc             3.111e-01  1.412e-01   2.204 0.027512 *
workclass State-gov                   -2.118e-01  1.389e-01  -1.524 0.127442
workclass Without-pay                 -8.638e-01  9.803e-01  -0.881 0.378251
fnlwgt                                1.173e-06  1.794e-07   6.538 6.23e-11 ***
marital.status Married-AF-spouse          -3.358e+00  4.887e+00  -0.687 0.491999
marital.status Married-civ-spouse          3.415e-01  1.788e-01   1.910 0.056168 .
marital.status Married-spouse-absent       2.126e-01  1.383e-01   1.538 0.124073
marital.status Never-married               5.897e-01  4.905e-02  12.022  < 2e-16 ***
marital.status Separated                   6.208e-02  9.081e-02   0.684 0.494180
marital.status Widowed                    -1.109e+00  1.146e-01  -9.679  < 2e-16 ***
occupation Armed-Forces                    1.362e+01  1.624e+02   0.084 0.933186
occupation Craft-repair                    3.255e+00  9.646e-02  33.744  < 2e-16 ***
occupation Exec-managerial                 7.853e-01  7.465e-02  10.519  < 2e-16 ***
occupation Farming-fishing                 3.170e+00  1.716e-01  18.475  < 2e-16 ***
occupation Handlers-cleaners               2.883e+00  1.064e-01  27.104  < 2e-16 ***
occupation Machine-op-inspct               1.571e+00  8.508e-02  18.468  < 2e-16 ***
occupation Other-service                   8.395e-01  6.617e-02  12.687  < 2e-16 ***
occupation Priv-house-serv                -1.535e+00  3.804e-01  -4.036 5.44e-05 ***
occupation Prof-specialty                  6.126e-01  7.162e-02   8.553  < 2e-16 ***
occupation Protective-serv                 2.385e+00  1.614e-01  14.777  < 2e-16 ***
occupation Sales                           8.229e-01  6.940e-02  11.858  < 2e-16 ***
occupation Tech-support                    8.769e-01  1.062e-01   8.254  < 2e-16 ***
occupation Transport-moving                3.380e+00  1.421e-01  23.785  < 2e-16 ***
relationship Not-in-family                -8.847e+00  1.020e+00  -8.670  < 2e-16 ***
relationship Other-relative               -8.807e+00  1.020e+00  -8.635  < 2e-16 ***
```

```
relationship Own-child                  -8.740e+00  1.020e+00  -8.566  < 2e-16 ***
relationship Unmarried                  -1.008e+01  1.022e+00  -9.868  < 2e-16 ***
relationship Wife                       -1.673e+01  1.420e+00 -11.787  < 2e-16 ***
race Asian-Pac-Islander                  3.148e-02  2.452e-01   0.128 0.897839
race Black                              -3.159e-01  1.997e-01  -1.582 0.113668
race Other                              -7.397e-02  2.816e-01  -0.263 0.792784
race White                               2.169e-02  1.925e-01   0.113 0.910273
hours.per.week                           2.163e-02  1.754e-03  12.332  < 2e-16 ***
native.country Canada                   -9.604e-01  1.234e+00  -0.778 0.436466
native.country China                    -8.553e-01  1.251e+00  -0.684 0.494186
native.country Columbia                 -1.354e+00  1.252e+00  -1.081 0.279740
native.country Cuba                     -2.090e+00  1.249e+00  -1.673 0.094332 .
native.country Dominican-Republic       -1.751e+00  1.237e+00  -1.416 0.156860
native.country Ecuador                  -1.298e+00  1.331e+00  -0.975 0.329361
native.country El-Salvador              -7.527e-01  1.227e+00  -0.614 0.539523
native.country England                  -7.791e-01  1.236e+00  -0.630 0.528479
native.country France                   -1.133e+00  1.335e+00  -0.849 0.395991
native.country Germany                  -1.611e+00  1.231e+00  -1.309 0.190513
native.country Greece                   -3.106e-01  1.419e+00  -0.219 0.826713
native.country Guatemala                -7.130e-01  1.254e+00  -0.569 0.569524
native.country Haiti                    -5.145e-01  1.266e+00  -0.406 0.684535
native.country Holand-Netherlands       -1.536e+01  5.354e+02  -0.029 0.977115
native.country Honduras                 -9.606e-01  1.428e+00  -0.672 0.501276
native.country Hong                     -9.228e-01  1.554e+00  -0.594 0.552490
native.country Hungary                  -1.330e+00  1.507e+00  -0.882 0.377566
native.country India                     3.170e-01  1.254e+00   0.253 0.800487
native.country Iran                      6.140e-01  1.360e+00   0.451 0.651752
native.country Ireland                  -1.766e+00  1.320e+00  -1.338 0.180840
native.country Italy                    -1.461e+00  1.291e+00  -1.132 0.257521
native.country Jamaica                  -1.481e+00  1.238e+00  -1.197 0.231328
native.country Japan                    -1.180e+00  1.257e+00  -0.938 0.348109
native.country Laos                     -1.846e+00  1.459e+00  -1.265 0.205710
native.country Mexico                   -7.794e-01  1.199e+00  -0.650 0.515668
native.country Nicaragua                -7.190e-01  1.289e+00  -0.558 0.577051
native.country Outlying-US(Guam-USVI-etc) -1.454e+00  1.366e+00  -1.064 0.287188
native.country Peru                     -1.888e+00  1.302e+00  -1.450 0.147111
native.country Philippines              -1.255e+00  1.202e+00  -1.044 0.296493
native.country Poland                   -1.417e+00  1.271e+00  -1.115 0.264769
native.country Portugal                 -1.707e+00  1.366e+00  -1.250 0.211334
native.country Puerto-Rico              -1.218e+00  1.227e+00  -0.993 0.320838
native.country Scotland                  4.003e-01  1.556e+00   0.257 0.797013
native.country South                    -9.904e-01  1.238e+00  -0.800 0.423647
native.country Taiwan                   -6.008e-01  1.312e+00  -0.458 0.646958
native.country Thailand                 -1.899e+00  1.413e+00  -1.344 0.178870
native.country Trinadad&Tobago          -2.664e+00  1.637e+00  -1.627 0.103651
native.country United-States            -1.184e+00  1.190e+00  -0.995 0.319734
native.country Vietnam                  -8.976e-01  1.234e+00  -0.728 0.466807
native.country Yugoslavia               -9.393e-01  1.562e+00  -0.601 0.547567
class >50K                               8.273e-01  7.633e-02  10.839  < 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: 38009  on 30161  degrees of freedom
Residual deviance: 17147  on 30084  degrees of freedom
AIC: 17303

Number of Fisher Scoring iterations: 12

anova(fit.reduced,fit.full,test="Chisq")
Analysis of Deviance Table

Model 1: sex ~ workclass + fnlwgt + marital.status + occupation + relationship +
    race + hours.per.week + native.country + class
Model 2: sex ~ age + workclass + fnlwgt + education.num + marital.status +
```

```
   occupation + relationship + race + capital.gain + capital.loss +
   hours.per.week + native.country + class
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     30084     17147
2     30080     17141  4    5.966   0.2017
```

```
adult.training.data<-
data.frame(adult.new$workclass,adult.new$fnlwgt,adult.new$marital.status,adult.new$oc
cupation,adult.new$relationship,adult.new$race,adult.new$hours.per.week,adult.new$nat
ive.country,adult.new$class)
adult.training.data$prob<-predict(fit.reduced,newdata =
adult.training.data,type="response")
adult.training.data$pre.result[adult.training.data$prob>0.5]<-1
adult.training.data$pre.result[adult.training.data$prob<=0.5]<-2
adult.training.data$pre.result<-
factor(adult.training.data$pre.result,levels=c(1,2),labels=c(" Male"," Female"))
compare<-data.frame(adult.new$sex,adult.training.data$pre.result)
library(dplyr)
attach(compare)
summary(filter(compare,adult.new.sex == adult.training.data.pre.result))
 adult.new.sex   adult.training.data.pre.result
  Female: 7983    Male  :17736
  Male  :17736    Female: 7983
```

Q5 Supplement:

```
fit.reduced<-glm(sex~
fnlwgt+marital.status+occupation+relationship+race+hours.per.week+native.country+clas
s,data=adult.new,family = binomial())
< 2.2e-16 ***
fit.reduced<-glm(sex~workclass+
marital.status+occupation+relationship+race+hours.per.week+native.country+class,data=
adult.new,family = binomial())
2.379e-09 ***

fit.reduced<-glm(sex~workclass+fnlwgt
+occupation+relationship+race+hours.per.week+native.country+class,data=adult.new,fami
ly = binomial())
< 2.2e-16 ***

fit.reduced<-glm(sex~workclass+fnlwgt+marital.status
+relationship+race+hours.per.week+native.country+class,data=adult.new,family        =
binomial())
< 2.2e-16 ***

fit.reduced<-glm(sex~workclass+fnlwgt+marital.status+occupation
+race+hours.per.week+native.country+class,data=adult.new,family = binomial())
< 2.2e-16 ***

fit.reduced<-glm(sex~workclass+fnlwgt+marital.status+occupation+relationship
+hours.per.week+native.country+class,data=adult.new,family = binomial())
1.179e-05 ***

fit.reduced<-glm(sex~workclass+fnlwgt+marital.status+occupation+relationship+race
+native.country+class,data=adult.new,family = binomial())
< 2.2e-16 ***

fit.reduced<-
glm(sex~workclass+fnlwgt+marital.status+occupation+relationship+race+hours.per.week
+class,data=adult.new,family = binomial())
0.00354 **

fit.reduced<-
glm(sex~workclass+fnlwgt+marital.status+occupation+relationship+race+hours.per.week+n
ative.country,data=adult.new,family = binomial())
```

< 2.2e-16 ***

```
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"))
adult$workclass[adult$workclass == " ?"]<-NA
adult$occupation[adult$occupation == " ?"]<-NA
adult$native.country[adult$native.country == " ?"]<-NA
adult.new<-na.omit(adult)
adult.new$workclass.Private[adult.new$workclass == " Private"]<-1
adult.new$workclass.Self.emp.not.inc[adult.new$workclass == " Self-emp-not-inc"]<-1
adult.new$workclass.Self.emp.inc[adult.new$workclass == " Self-emp-inc"]<-1
adult.new$workclass.Federal.gov[adult.new$workclass == " Federal-gov"]<-1
adult.new$workclass.Local.gov[adult.new$workclass == " Local-gov"]<-1
adult.new$workclass.State.gov[adult.new$workclass == " State-gov"]<-1
adult.new$workclass.Without.pay[adult.new$workclass == " Without-pay"]<-1
adult.new$workclass.Never.worked[adult.new$workclass == " Never-worked"]<-1
adult.new$marital.status.Married.civ.spouse[adult.new$marital.status==" Married-civ-
spouse"]<-1
adult.new$marital.status.Divorced[adult.new$marital.status==" Divorced"]<-1
adult.new$marital.status.Never.married[adult.new$marital.status==" Never-married"]<-1
adult.new$marital.status.Separated[adult.new$marital.status==" Separated"]<-1
adult.new$marital.status.Widowed[adult.new$marital.status==" Widowed"]<-1
adult.new$marital.status.Married.spouse.absent[adult.new$marital.status==" Married-
spouse-absent"]<-1
adult.new$marital.status.Married.AF.spouse[adult.new$marital.status==" Married-AF-
spouse"]<-1
adult.new$occupation.Tech.support[adult.new$occupation==" Tech-support"]<-1
adult.new$occupation.Craft.repair[adult.new$occupation==" Craft-repair"]<-1
adult.new$occupation.Other.service[adult.new$occupation==" Other-service"]<-1
adult.new$occupation.Sales[adult.new$occupation==" Sales"]<-1
adult.new$occupation.Exec.managerial[adult.new$occupation==" Exec-managerial"]<-1
adult.new$occupation.Prof.specialty[adult.new$occupation==" Prof-specialty"]<-1
adult.new$occupation.Handlers.cleaners[adult.new$occupation==" Handlers-cleaners"]<-1
adult.new$occupation.Machine.op.inspct[adult.new$occupation==" Machine-op-inspct"]<-1
adult.new$occupation.Adm.clerical[adult.new$occupation==" Adm-clerical"]<-1
adult.new$occupation.Farming.fishing[adult.new$occupation==" Farming-fishing"]<-1
adult.new$occupation.Transport.moving[adult.new$occupation==" Transport-moving"]<-1
adult.new$occupation.Priv.house.serv[adult.new$occupation==" Priv-house-serv"]<-1
adult.new$occupation.Protective.serv[adult.new$occupation==" Protective-serv"]<-1
adult.new$occupation.Armed.Forces[adult.new$occupation==" Armed-Forces"]<-1
adult.new$relationship.Wife[adult.new$relationship == " Wife"]<-1
adult.new$relationship.Own.child[adult.new$relationship == " Own-child"]<-1
adult.new$relationship.Husband[adult.new$relationship == " Husband"]<-1
adult.new$relationship.Not.in.family[adult.new$relationship == " Not-in-family"]<-1
adult.new$relationship.Other.relative[adult.new$relationship == " Other-relative"]<-1
adult.new$relationship.Unmarried[adult.new$relationship == " Unmarried"]<-1
adult.new$race.White[adult.new$race==" White"]<-1
adult.new$race.Asian.Pac.Islander[adult.new$race==" Asian-Pac-Islander"]<-1
adult.new$race.Amer.Indian.Eskimo[adult.new$race==" Amer-Indian-Eskimo"]<-1
adult.new$race.Other[adult.new$race==" Other"]<-1
adult.new$race.Black[adult.new$race==" Black"]<-1
adult.new$native.country.United.States[adult.new$native.country==" United-States"]<-1
adult.new$native.country.Cambodia[adult.new$native.country==" Cambodia"]<-1
adult.new$native.country.England[adult.new$native.country==" England"]<-1
adult.new$native.country.Puerto.Rico[adult.new$native.country==" Puerto-Rico"]<-1
adult.new$native.country.Canada[adult.new$native.country==" Canada"]<-1
adult.new$native.country.Germany[adult.new$native.country==" Germany"]<-1
adult.new$native.country.Outlying.US.Guam.USVI.etc[adult.new$native.country==
Outlying-US(Guam-USVI-etc)"]<-1
adult.new$native.country.India[adult.new$native.country==" India"]<-1
adult.new$native.country.Japan[adult.new$native.country==" Japan"]<-1
adult.new$native.country.Greece[adult.new$native.country==" Greece"]<-1
adult.new$native.country.South[adult.new$native.country==" South"]<-1
```

```
adult.new$native.country.China[adult.new$native.country==" China"]<-1
adult.new$native.country.Cuba[adult.new$native.country==" Cuba"]<-1
adult.new$native.country.Iran[adult.new$native.country==" Iran"]<-1
adult.new$native.country.Honduras[adult.new$native.country==" Honduras"]<-1
adult.new$native.country.Philippines[adult.new$native.country==" Philippines"]<-1
adult.new$native.country.Italy[adult.new$native.country==" Italy"]<-1
adult.new$native.country.Poland[adult.new$native.country==" Poland"]<-1
adult.new$native.country.Jamaica[adult.new$native.country==" Jamaica"]<-1
adult.new$native.country.Vietnam[adult.new$native.country==" Vietnam"]<-1
adult.new$native.country.Mexico[adult.new$native.country==" Mexico"]<-1
adult.new$native.country.Portugal[adult.new$native.country==" Portugal"]<-1
adult.new$native.country.Ireland[adult.new$native.country==" Ireland"]<-1
adult.new$native.country.France[adult.new$native.country==" France"]<-1
adult.new$native.country.Dominican.Republic[adult.new$native.country==" Dominican-
Republic"]<-1
adult.new$native.country.Laos[adult.new$native.country==" Laos"]<-1
adult.new$native.country.Ecuador[adult.new$native.country==" Ecuador"]<-1
adult.new$native.country.Taiwan[adult.new$native.country==" Taiwan"]<-1
adult.new$native.country.Haiti[adult.new$native.country==" Haiti"]<-1
adult.new$native.country.Columbia[adult.new$native.country==" Columbia"]<-1
adult.new$native.country.Hungary[adult.new$native.country==" Hungary"]<-1
adult.new$native.country.Guatemala[adult.new$native.country==" Guatemala"]<-1
adult.new$native.country.Nicaragua[adult.new$native.country==" Nicaragua"]<-1
adult.new$native.country.Scotland[adult.new$native.country==" Scotland"]<-1
adult.new$native.country.Thailand[adult.new$native.country==" Thailand"]<-1
adult.new$native.country.Yugoslavia[adult.new$native.country==" Yugoslavia"]<-1
adult.new$native.country.El.Salvador[adult.new$native.country==" El-Salvador"]<-1
adult.new$native.country.Trinadad.Tobago[adult.new$native.country==""
Trinadad&Tobago"]<-1
adult.new$native.country.Peru[adult.new$native.country==" Peru"]<-1
adult.new$native.country.Hong[adult.new$native.country==" Hong"]<-1
adult.new$native.country.Holand.Netherlands[adult.new$native.country==" Holand-
Netherlands"]<-1
adult.new$class.less50K[adult.new$class ==" <=50K"]<-1
adult.new$class.more50K[adult.new$class ==" >50K"]<-1
adult.new[is.na(adult.new)]<-0
fit.all<-
glm(sex~age+fnlwgt+education.num+capital.gain+capital.loss+hours.per.week+workclass.P
rivate+workclass.Self.emp.not.inc+workclass.Self.emp.inc+workclass.Federal.gov+workcl
ass.Local.gov+workclass.State.gov+workclass.Without.pay+workclass.Never.worked+marita
l.status.Married.civ.spouse+marital.status.Divorced+marital.status.Never.married+mari
tal.status.Separated+marital.status.Widowed+marital.status.Married.spouse.absent+mari
tal.status.Married.AF.spouse+occupation.Tech.support+occupation.Craft.repair+occupati
on.Other.service+occupation.Sales+occupation.Exec.managerial+occupation.Prof.specialt
y+occupation.Handlers.cleaners+occupation.Machine.op.inspct+occupation.Adm.clerical+o
ccupation.Farming.fishing+occupation.Transport.moving+occupation.Priv.house.serv+occu
pation.Protective.serv+occupation.Armed.Forces+relationship.Wife+relationship.Own.chi
ld+relationship.Husband+relationship.Not.in.family+relationship.Other.relative+relati
onship.Unmarried+race.White+race.Asian.Pac.Islander+race.Amer.Indian.Eskimo+race.Othe
r+race.Black+native.country.United.States+native.country.Cambodia+native.country.Engl
and+native.country.Puerto.Rico+native.country.Canada+native.country.Germany+native.co
untry.Outlying.US.Guam.USVI.etc+native.country.India+native.country.Japan+native.coun
try.Greece+native.country.South+native.country.China+native.country.Cuba+native.count
ry.Iran+native.country.Honduras+native.country.Philippines+native.country.Italy+nativ
e.country.Poland+native.country.Jamaica+native.country.Vietnam+native.country.Mexico+
native.country.Portugal+native.country.Ireland+native.country.France+native.country.D
ominican.Republic+native.country.Laos+native.country.Ecuador+native.country.Taiwan+na
tive.country.Haiti+native.country.Columbia+native.country.Hungary+native.country.Guat
emala+native.country.Nicaragua+native.country.Scotland+native.country.Thailand+native
.country.Yugoslavia+native.country.El.Salvador+native.country.Trinadad.Tobago+native.
country.Peru+native.country.Hong+native.country.Holand.Netherlands+class.less50K+clas
s.more50K,data=adult.new,family = binomial())
```