

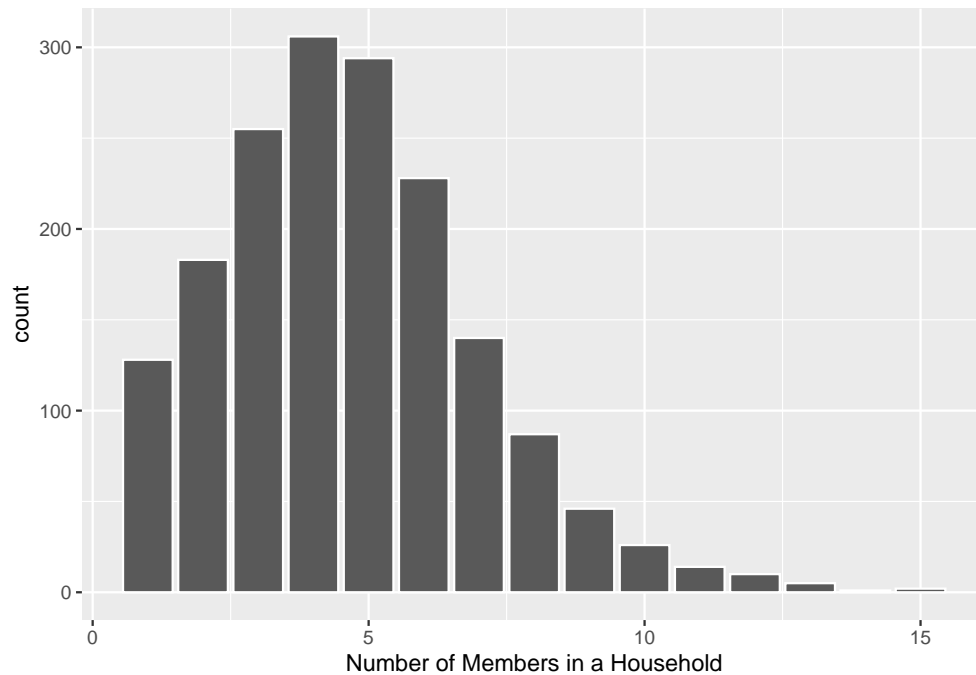
project2_test

Group_01

2021/7/7

Introduction

Data come from the FIES (Family Income and Expenditure Survey) recorded in the Philippines. The survey, which is undertaken every three years, is aimed at providing data on family income and expenditure. The data obtained from this survey are from different regions across the Philippines.



Total.Number.of.Family.members	n	percent
1	128	7.4%
2	183	10.6%
3	255	14.8%
4	306	17.7%
5	294	17.0%
6	228	13.2%
7	140	8.1%
8	87	5.0%
9	46	2.7%
10	26	1.5%

11	14	0.8%
12	10	0.6%
13	5	0.3%
14	1	0.1%
15	2	0.1%
Total	1725	100.0%

	Total.Number.of.Family.members		
Total.Number.of.Family.members	1.00000000		
Total.Household.Income	0.19228742		
Total.Food.Expenditure	0.46924215		
Household.Head.Age	-0.06541636		
House.Floor.Area	-0.01415702		
House.Age	-0.07003586		
Number.of.bedrooms	0.07207630		
Electricity	0.09193871		
	Total.Household.Income	Total.Food.Expenditure	
Total.Number.of.Family.members	0.19228742	0.46924215	
Total.Household.Income	1.00000000	0.611494530	
Total.Food.Expenditure	0.61149453	1.000000000	
Household.Head.Age	0.06280405	-0.051724735	
House.Floor.Area	0.23413840	0.124320633	
House.Age	0.02471720	0.006725185	
Number.of.bedrooms	0.44137375	0.355734454	
Electricity	0.14866655	0.198610366	
	Household.Head.Age	House.Floor.Area	House.Age
Total.Number.of.Family.members	-0.06541636	-0.01415702	-0.070035856
Total.Household.Income	0.06280405	0.23413840	0.024717197
Total.Food.Expenditure	-0.05172474	0.12432063	0.006725185
Household.Head.Age	1.00000000	0.09057216	0.218079293
House.Floor.Area	0.09057216	1.00000000	0.074265080
House.Age	0.21807929	0.07426508	1.000000000
Number.of.bedrooms	0.15415511	0.37399081	0.123180471
Electricity	-0.01304412	0.10693465	0.085327324
	Number.of.bedrooms	Electricity	
Total.Number.of.Family.members	0.0720763	0.09193871	
Total.Household.Income	0.4413738	0.14866655	
Total.Food.Expenditure	0.3557345	0.19861037	
Household.Head.Age	0.1541551	-0.01304412	
House.Floor.Area	0.3739908	0.10693465	
House.Age	0.1231805	0.08532732	
Number.of.bedrooms	1.0000000	0.21376315	
Electricity	0.2137632	1.00000000	

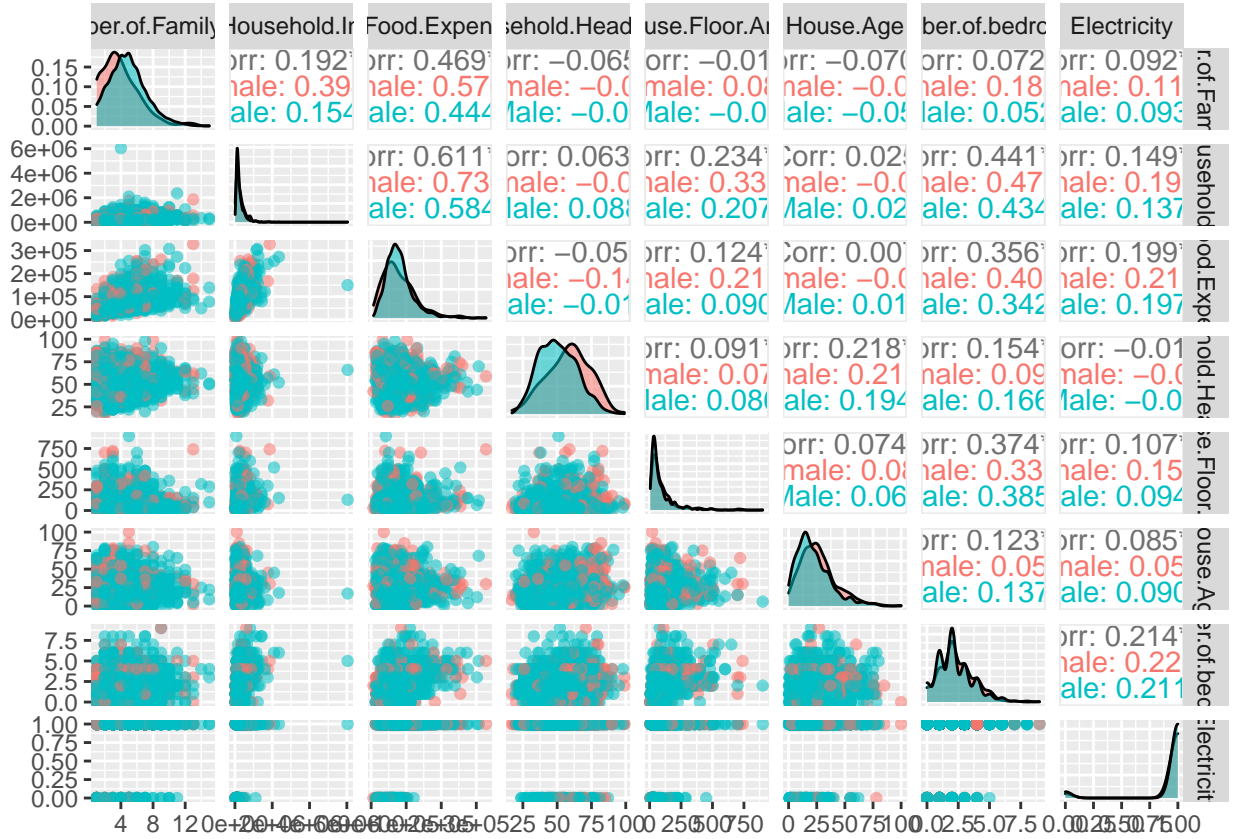


Table 1: Summary statistics of numerical variables

Variable	Missing	Complete	Mean	SD	Min.	1st Q.	Median	3rd Q.
Total.Number.of.Family.members	0	1	4.67	2.33	1	3	4	6
Total.Household.Income	0	1	269540.48	274564.17	11988	118565	188580	328335
Total.Food.Expenditure	0	1	80352.78	41194.36	6781	51922	73578	98493
Household.Head.Age	0	1	52.23	14.52	17	41	52	63
House.Floor.Area	0	1	90.92	99.20	5	32	54	102
House.Age	0	1	22.98	15.32	0	12	20	31
Number.of.bedrooms	0	1	2.26	1.44	0	1	2	3
Electricity	0	1	0.93	0.26	0	1	1	1

Type.of.Household	n	percent
Extended Family	569	33.0%
Single Family	1148	66.6%
Two or More Nonrelated Persons/Members	8	0.5%
Total	1725	100.0%

Household.Head.Sex	n	percent
Female	369	21.4%
Male	1356	78.6%
Total	1725	100.0%

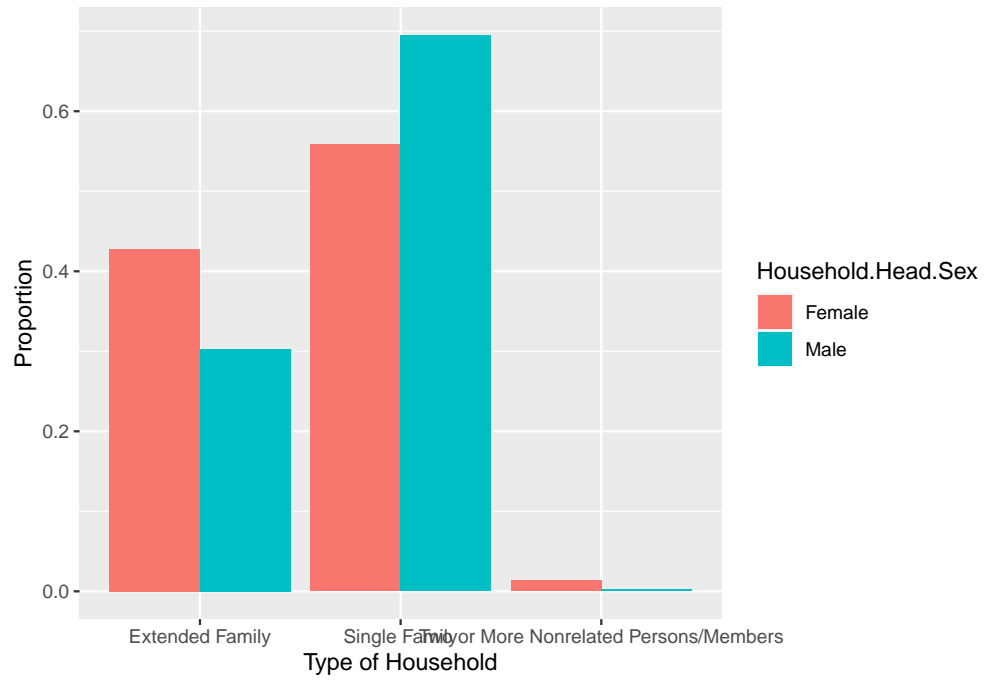
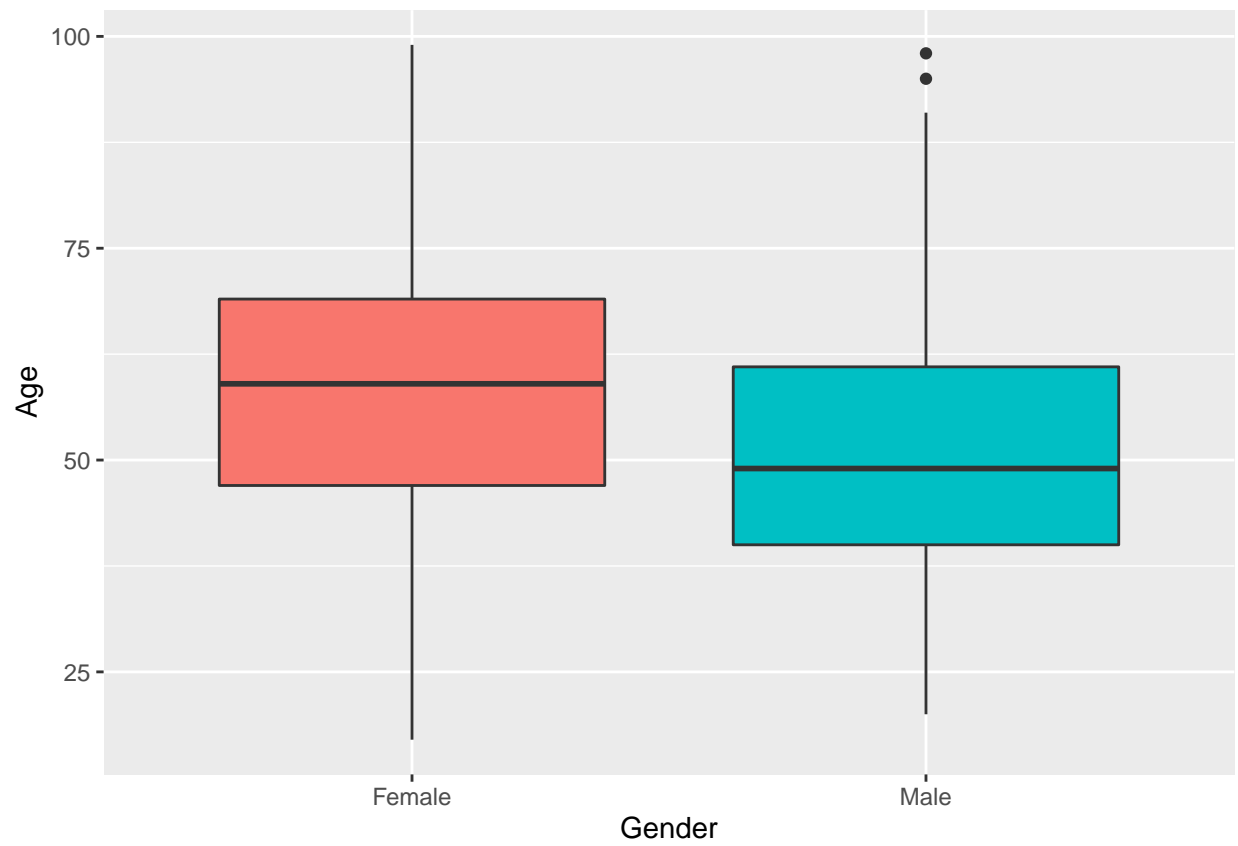
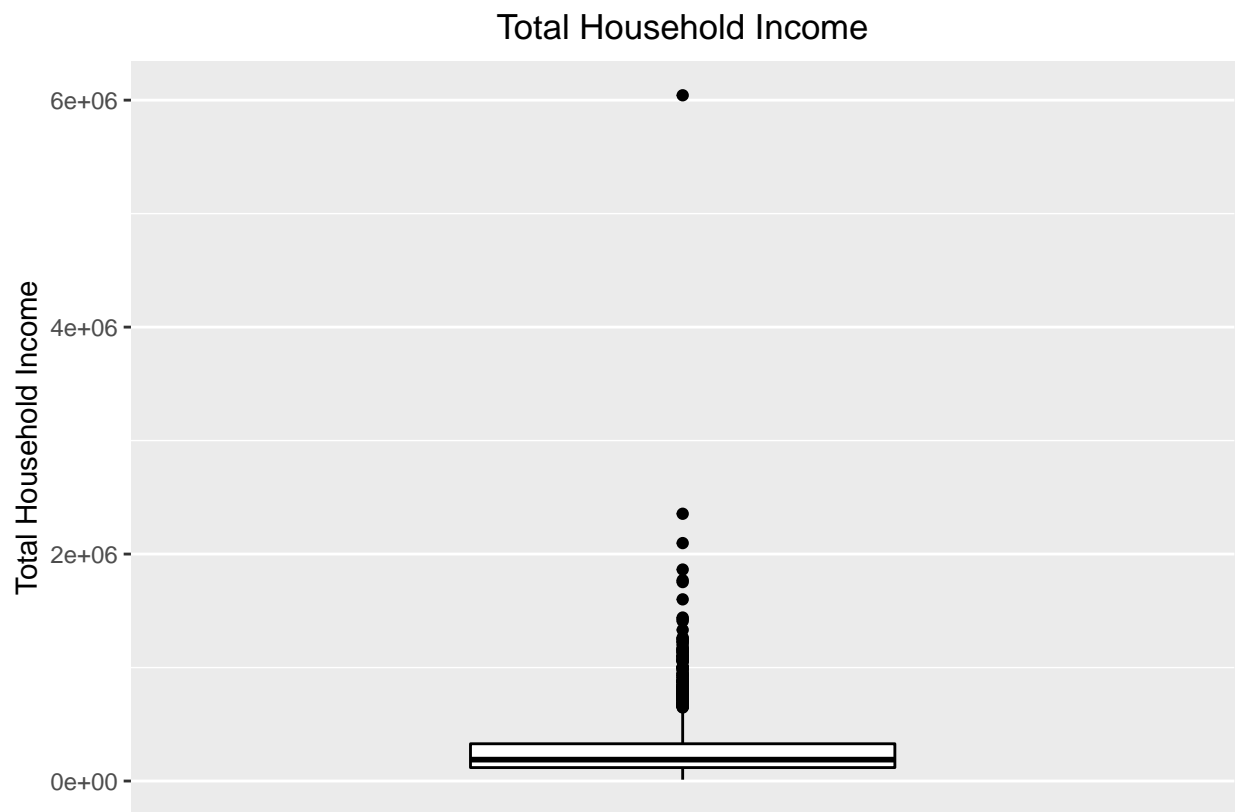


Figure 1: Barplot of household head'ssex by type of household

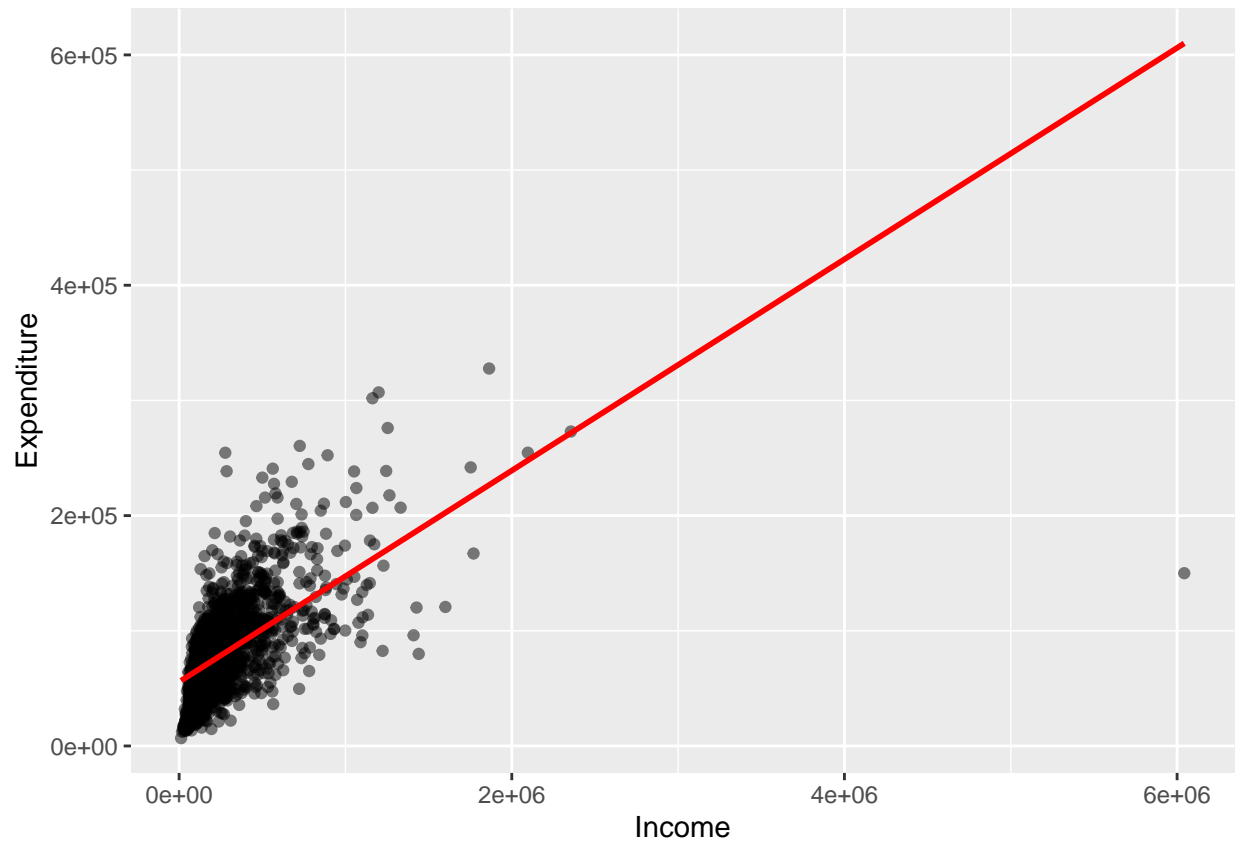
Gender & Age



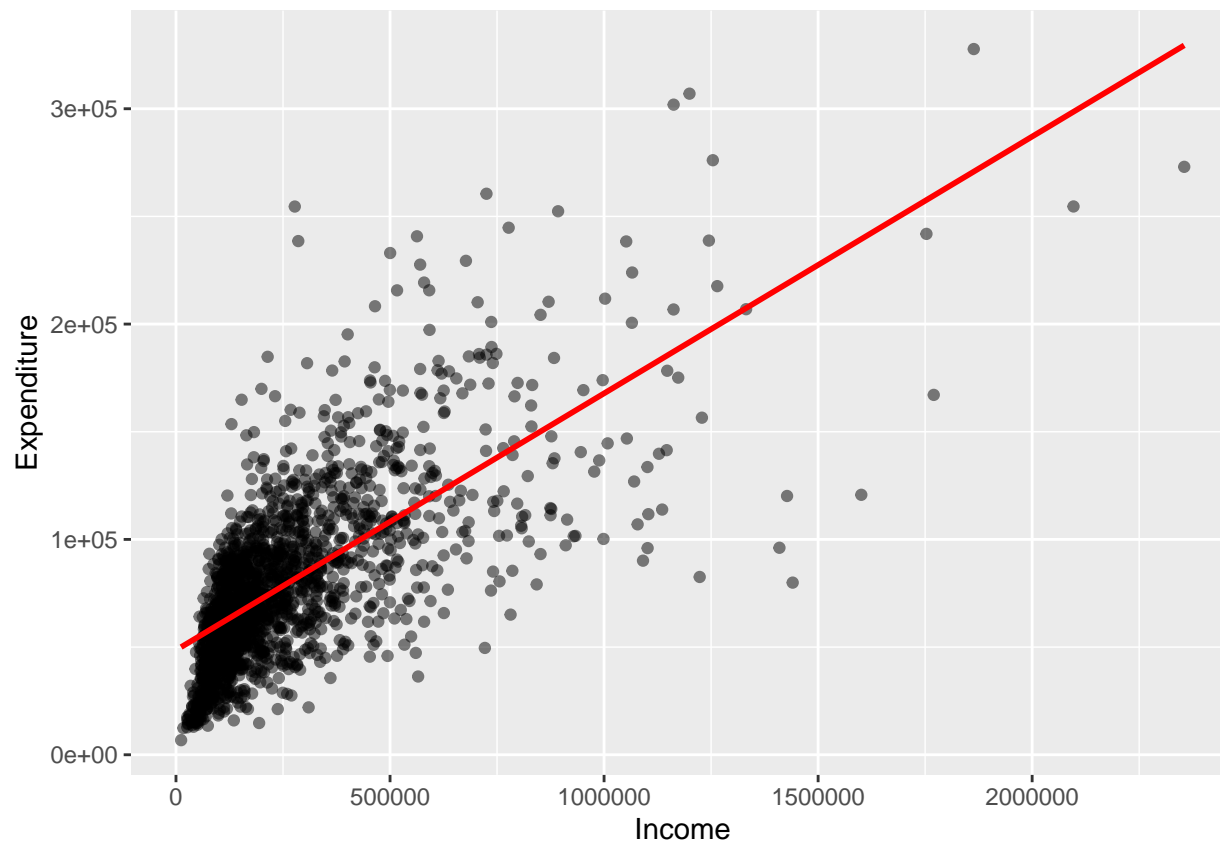
Household Income Balance







The above plot highlights a possible outlier in terms of income, this could be a data entry error or just an outlier. Removing this observation from the data set and plotting provides the following scatter diagram.



Model

Subset selection object

Call: `regsubsets.formula(Total.Number.of.Family.members ~ ., data = data,`
`nvmax = 10)`

10 Variables (and intercept)

	Forced in	Forced out
Total.Household.Income	FALSE	FALSE
Total.Food.Expenditure	FALSE	FALSE
Household.Head.SexMale	FALSE	FALSE
Household.Head.Age	FALSE	FALSE
Type.of.HouseholdSingle Family	FALSE	FALSE
Type.of.HouseholdTwo or More Nonrelated Persons/Members	FALSE	FALSE
House.Floor.Area	FALSE	FALSE
House.Age	FALSE	FALSE
Number.of.bedrooms	FALSE	FALSE
Electricity	FALSE	FALSE

1 subsets of each size up to 10

Selection Algorithm: exhaustive

	Total.Household.Income	Total.Food.Expenditure	Household.Head.SexMale
1 (1)	" "	"*"	" "
2 (1)	" "	"*"	" "
3 (1)	" "	"*"	"*"
4 (1)	"*"	"*"	"*"
5 (1)	"*"	"*"	"*"

```

6 ( 1 ) "*" "*" "*"
7 ( 1 ) "*" "*" "*"
8 ( 1 ) "*" "*" "*"
9 ( 1 ) "*" "*" "*"
10 ( 1 ) "*" "*" "*"

Household.Head.Age Type.of.HouseholdSingle Family
1 ( 1 ) " " " "
2 ( 1 ) " " "*"
3 ( 1 ) " " "*"
4 ( 1 ) " " "*"
5 ( 1 ) " " "*"
6 ( 1 ) " " "*"
7 ( 1 ) "*" "*"
8 ( 1 ) "*" "*"
9 ( 1 ) "*" "*"
10 ( 1 ) "*" "*"

Type.of.HouseholdTwo or More Nonrelated Persons/Members
1 ( 1 ) " "
2 ( 1 ) " "
3 ( 1 ) " "
4 ( 1 ) " "
5 ( 1 ) " "
6 ( 1 ) " "
7 ( 1 ) " "
8 ( 1 ) " "
9 ( 1 ) "*"
10 ( 1 ) "*"

House.Floor.Area House.Age Number.of.bedrooms Electricity
1 ( 1 ) " " " " " "
2 ( 1 ) " " " " " "
3 ( 1 ) " " " " " "
4 ( 1 ) " " " " " "
5 ( 1 ) " " "*" " " "
6 ( 1 ) " " "*" "*" " "
7 ( 1 ) " " "*" "*" " "
8 ( 1 ) "*" "*" "*" " "
9 ( 1 ) "*" "*" "*" " "
10 ( 1 ) "*" "*" "*" "*"

```

```

Adj.R2    CP    BIC
   9      8     6

```

Call:

```

glm(formula = Total.Number.of.Family.members ~ Total.Household.Income +
    Total.Food.Expenditure + Household.Head.Age + House.Floor.Area +
    House.Age + Number.of.bedrooms + Electricity, data = data)

```

Deviance Residuals:

```

      Min      1Q   Median      3Q      Max
-5.5671 -1.4626 -0.3084  1.2037 10.7417

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.746e+00  2.667e-01  10.298  < 2e-16 ***

```

```

Total.Household.Income -1.022e-06  2.384e-07  -4.287  1.91e-05 ***
Total.Food.Expenditure  3.197e-05  1.540e-06  20.759  < 2e-16 ***
Household.Head.Age      -4.491e-04  3.520e-03  -0.128  0.89850
House.Floor.Area        -7.261e-04  5.357e-04  -1.355  0.17550
House.Age               -9.472e-03  3.301e-03  -2.870  0.00416 **
Number.of.bedrooms      -9.756e-02  4.121e-02  -2.367  0.01802 *
Electricity             1.696e-01  1.929e-01   0.879  0.37955

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 4.130968)

```

Null deviance: 9384.0  on 1724  degrees of freedom
Residual deviance: 7092.9  on 1717  degrees of freedom
AIC: 7352.3

```

Number of Fisher Scoring iterations: 2

Call:

```

glm(formula = Total.Number.of.Family.members ~ Total.Household.Income +
    Total.Food.Expenditure + House.Age + Number.of.bedrooms,
    data = data)

```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-5.5796  -1.4561  -0.3048   1.1778  10.6187

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.828e+00  1.375e-01  20.560 < 2e-16 ***
Total.Household.Income -1.061e-06  2.364e-07  -4.487 7.71e-06 ***
Total.Food.Expenditure  3.229e-05  1.513e-06  21.340 < 2e-16 ***
House.Age       -9.507e-03  3.223e-03  -2.950 0.00322 **
Number.of.bedrooms -1.103e-01  3.855e-02  -2.862 0.00425 **

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 4.129948)

```

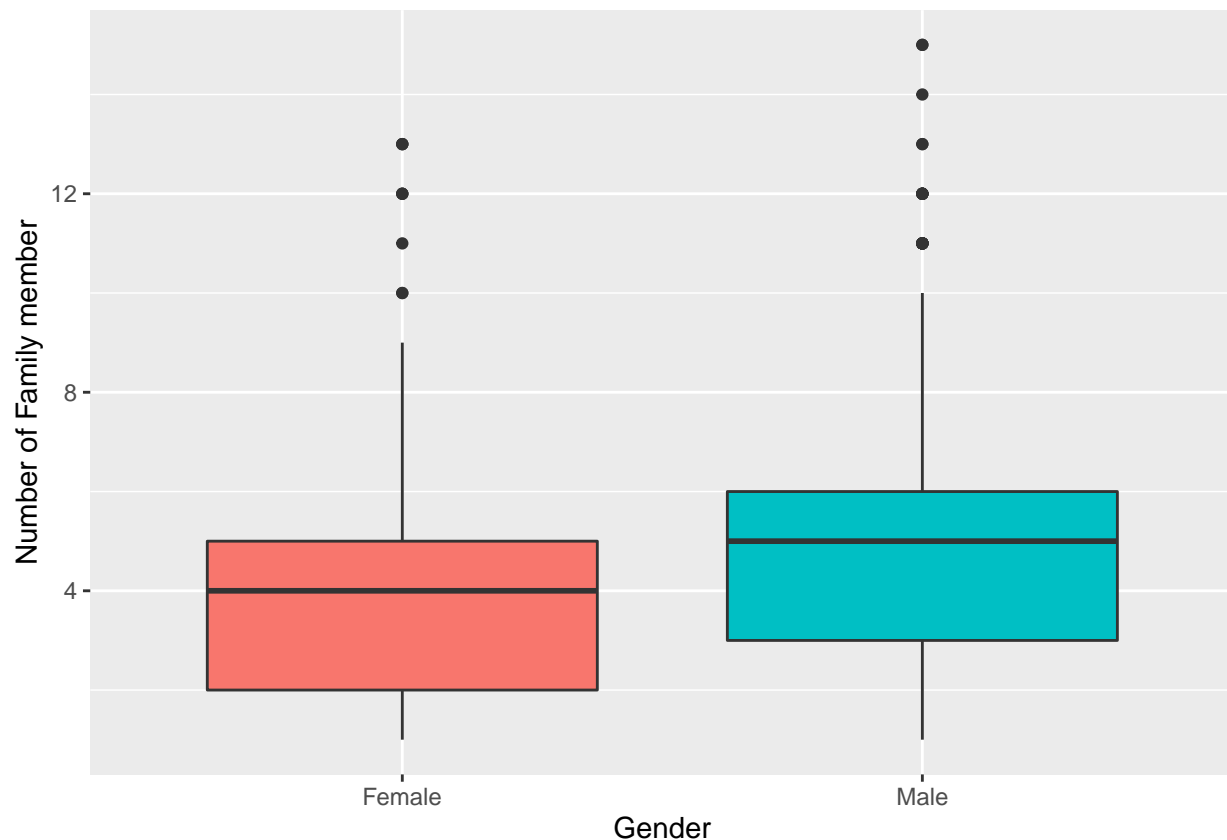
Null deviance: 9384.0  on 1724  degrees of freedom
Residual deviance: 7103.5  on 1720  degrees of freedom
AIC: 7348.8

```

Number of Fisher Scoring iterations: 2

Model	AIC	BIC
Full Model	7352.25	7401.33
Significant Factors Model	7348.84	7381.56

Family Members & Gender



Hence we can see that the male household head's number of family members tend to be more than the female's.

Log-odds

Call:

```
glm(formula = Household.Head.Sex ~ Total.Number.of.Family.members,  
     family = binomial(link = "logit"), data = data.sex_number)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.4219	0.4705	0.6602	0.7163	0.9054

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.49674	0.13174	3.771	0.000163 ***
Total.Number.of.Family.members	0.18319	0.02844	6.442	1.18e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1790.9 on 1724 degrees of freedom

Residual deviance: 1745.4 on 1723 degrees of freedom
AIC: 1749.4

Number of Fisher Scoring iterations: 4

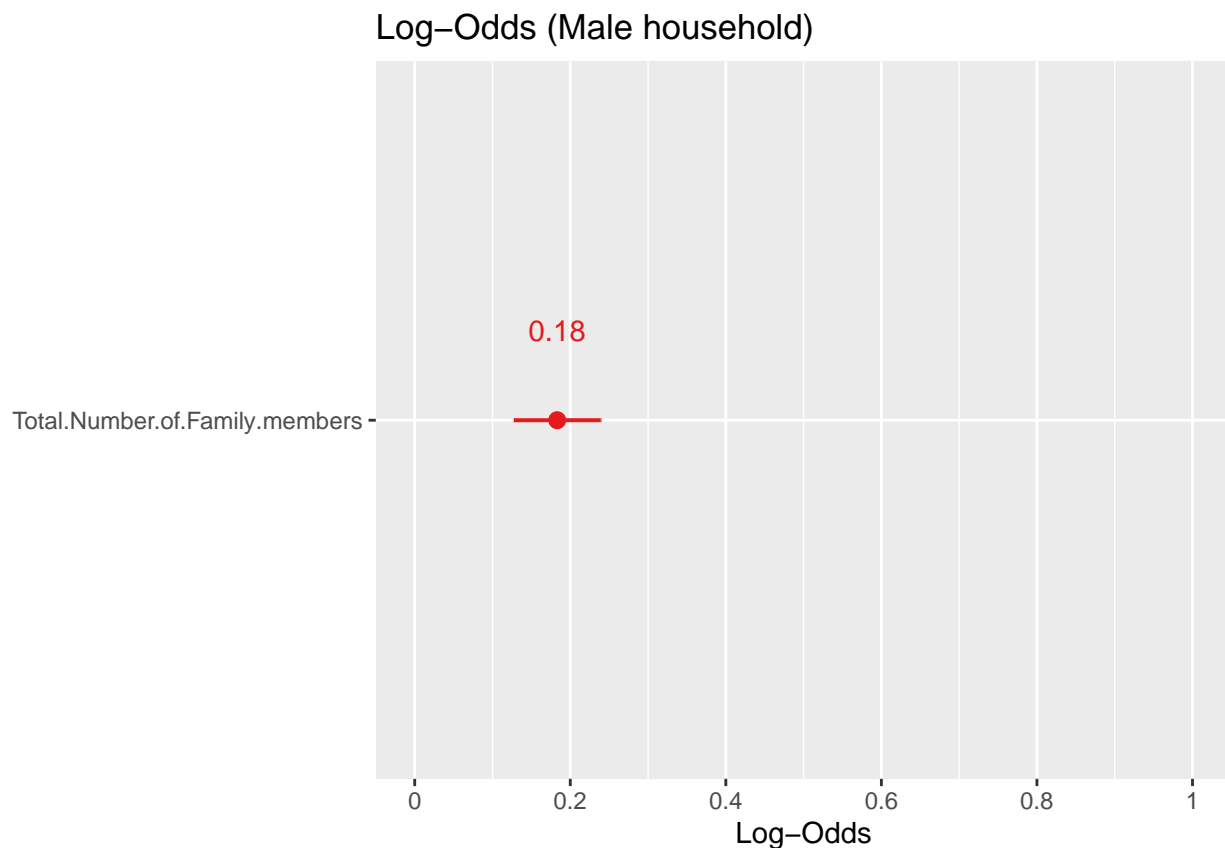
[1] "Female" "Male"

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta \cdot \text{number of family members} = 0.5 + 0.18 \cdot \text{number of family members},$$

Where $p = \text{Prob}(\text{Male})$ and $1 - p = \text{Prob}(\text{Female})$. Hence, the log-odds of the household being male increase by 0.18 for every one unit increase in number of family members. This provides us with a point estimate of how the log-odds changes with age.

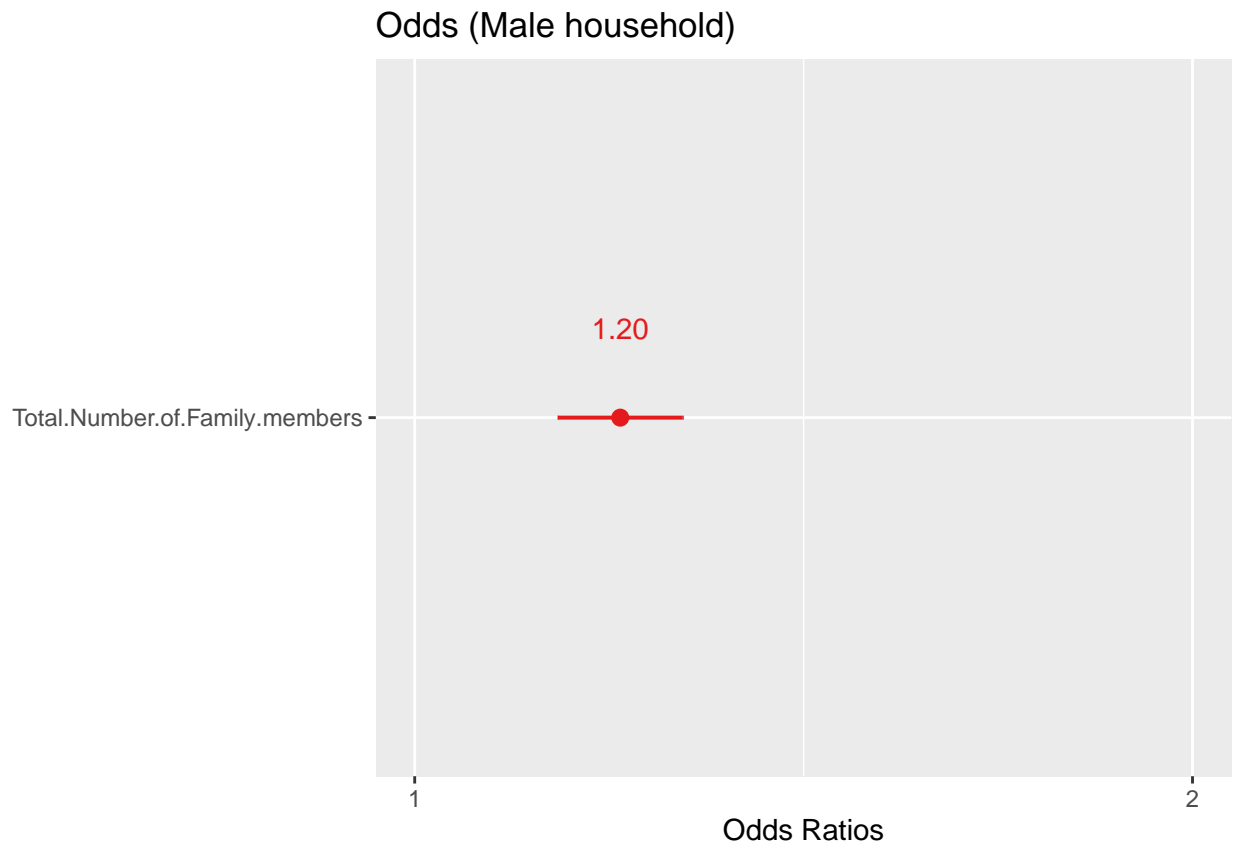
However, we are also interested in producing a 95% confidence interval for these log-odds.

	2.5 %	97.5 %
(Intercept)	0.2388990	0.7555347
Total.Number.of.Family.members	0.1282353	0.2397474



Now, let's add the estimates of the log-odds to our data set:

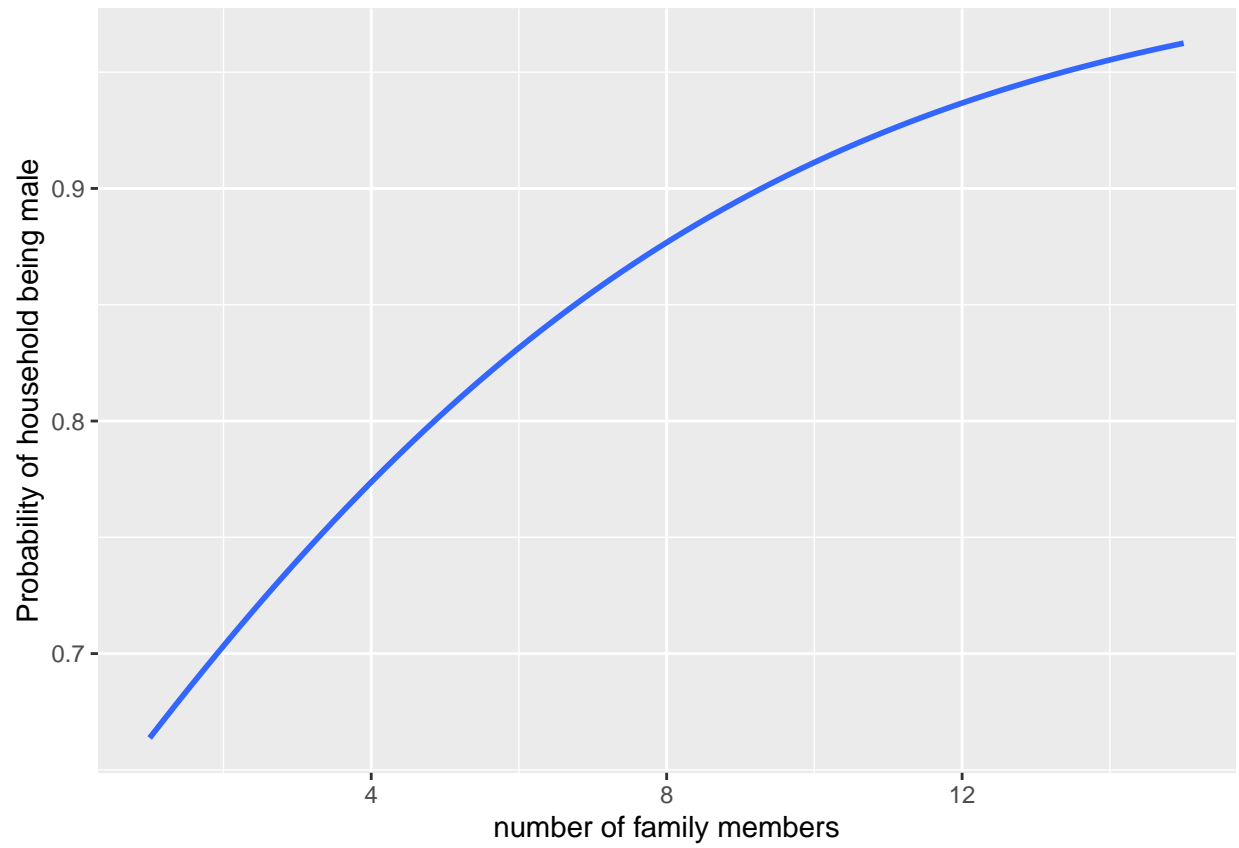
Odds



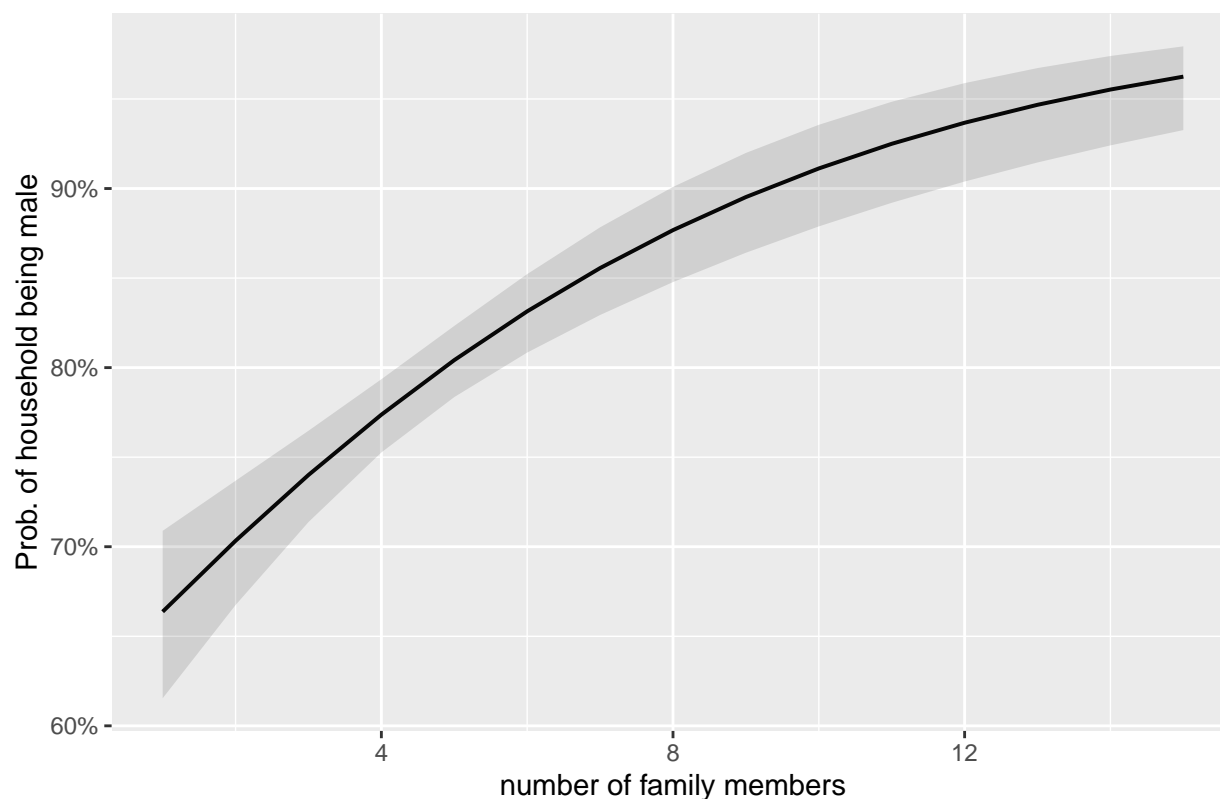
Now, let's add the estimates of the odds to our data set:

Probabilities

Plot the probability of being male



`$Total.Number.of.Family.members`



GLM

Call:

```
glm(formula = Total.Number.of.Family.members ~ Total.Household.Income +
    Total.Food.Expenditure + Household.Head.Sex + Household.Head.Age +
    Type.of.Household + House.Floor.Area + House.Age + Number.of.bedrooms +
    Electricity, family = binomial(link = "logit"), data = data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.2298	0.0000	0.0291	0.1946	1.8561

Coefficients:

	Estimate	Std. Error
(Intercept)	1.944e+01	6.821e+02
Total.Household.Income	-3.557e-06	1.017e-06
Total.Food.Expenditure	1.048e-04	9.895e-06
Household.Head.SexMale	1.143e+00	2.734e-01
Household.Head.Age	-5.534e-03	7.499e-03
Type.of.HouseholdSingle Family	-2.243e+01	6.821e+02
Type.of.HouseholdTwo or More Nonrelated Persons/Members	-7.011e+00	9.230e+03
House.Floor.Area	1.484e-03	1.224e-03
House.Age	-9.422e-04	7.817e-03
Number.of.bedrooms	-1.774e-01	1.034e-01
Electricity	3.235e-01	3.411e-01

	z value	Pr(> z)
(Intercept)	0.028	0.977268
Total.Household.Income	-3.497	0.000471 ***
Total.Food.Expenditure	10.595	< 2e-16 ***
Household.Head.SexMale	4.181	2.91e-05 ***
Household.Head.Age	-0.738	0.460510
Type.of.HouseholdSingle Family	-0.033	0.973771
Type.of.HouseholdTwo or More Nonrelated Persons/Members	-0.001	0.999394
House.Floor.Area	1.212	0.225414
House.Age	-0.121	0.904055
Number.of.bedrooms	-1.715	0.086317 .
Electricity	0.948	0.343015

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 912.10 on 1724 degrees of freedom
Residual deviance: 448.76 on 1714 degrees of freedom
AIC: 470.76

Number of Fisher Scoring iterations: 20

Call:

```
glm(formula = Total.Number.of.Family.members ~ (Total.Household.Income +
  Total.Food.Expenditure + Household.Head.Age + Type.of.Household +
  House.Floor.Area + House.Age + Number.of.bedrooms + Electricity) *
  Household.Head.Sex, family = binomial(link = "logit"), data = data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8031	0.0000	0.0179	0.1650	1.8853

Coefficients:

	Estimate
(Intercept)	2.047e+01
Total.Household.Income	-1.270e-06
Total.Food.Expenditure	7.284e-05
Household.Head.Age	-3.633e-02
Type.of.HouseholdSingle Family	-1.963e+01
Type.of.HouseholdTwo or More Nonrelated Persons/Members	-3.269e+00
House.Floor.Area	6.403e-03
House.Age	-2.041e-03
Number.of.bedrooms	-1.696e-01
Electricity	-1.269e+00
Household.Head.SexMale	2.751e-02
Total.Household.Income:Household.Head.SexMale	-2.883e-06
Total.Food.Expenditure:Household.Head.SexMale	5.175e-05
Household.Head.Age:Household.Head.SexMale	4.213e-02
Type.of.HouseholdSingle Family:Household.Head.SexMale	-4.123e+00
Type.of.HouseholdTwo or More Nonrelated Persons/Members:Household.Head.SexMale	-5.502e+00
House.Floor.Area:Household.Head.SexMale	-6.356e-03
House.Age:Household.Head.SexMale	2.190e-03
Number.of.bedrooms:Household.Head.SexMale	-3.489e-02

Electricity:Household.Head.SexMale	1.999e+00
	Std. Error
(Intercept)	1.884e+03
Total.Household.Income	2.334e-06
Total.Food.Expenditure	1.502e-05
Household.Head.Age	1.423e-02
Type.of.HouseholdSingle Family	1.884e+03
Type.of.HouseholdTwo or More Nonrelated Persons/Members	1.234e+04
House.Floor.Area	3.016e-03
House.Age	1.535e-02
Number.of.bedrooms	2.101e-01
Electricity	7.165e-01
Household.Head.SexMale	2.021e+03
Total.Household.Income:Household.Head.SexMale	2.612e-06
Total.Food.Expenditure:Household.Head.SexMale	2.026e-05
Household.Head.Age:Household.Head.SexMale	1.702e-02
Type.of.HouseholdSingle Family:Household.Head.SexMale	2.021e+03
Type.of.HouseholdTwo or More Nonrelated Persons/Members:Household.Head.SexMale	1.905e+04
House.Floor.Area:Household.Head.SexMale	3.348e-03
House.Age:Household.Head.SexMale	1.807e-02
Number.of.bedrooms:Household.Head.SexMale	2.451e-01
Electricity:Household.Head.SexMale	8.175e-01
	z value
(Intercept)	0.011
Total.Household.Income	-0.544
Total.Food.Expenditure	4.849
Household.Head.Age	-2.552
Type.of.HouseholdSingle Family	-0.010
Type.of.HouseholdTwo or More Nonrelated Persons/Members	0.000
House.Floor.Area	2.123
House.Age	-0.133
Number.of.bedrooms	-0.807
Electricity	-1.772
Household.Head.SexMale	0.000
Total.Household.Income:Household.Head.SexMale	-1.104
Total.Food.Expenditure:Household.Head.SexMale	2.554
Household.Head.Age:Household.Head.SexMale	2.476
Type.of.HouseholdSingle Family:Household.Head.SexMale	-0.002
Type.of.HouseholdTwo or More Nonrelated Persons/Members:Household.Head.SexMale	0.000
House.Floor.Area:Household.Head.SexMale	-1.899
House.Age:Household.Head.SexMale	0.121
Number.of.bedrooms:Household.Head.SexMale	-0.142
Electricity:Household.Head.SexMale	2.445
	Pr(> z)
(Intercept)	0.9913
Total.Household.Income	0.5863
Total.Food.Expenditure	1.24e-06
Household.Head.Age	0.0107
Type.of.HouseholdSingle Family	0.9917
Type.of.HouseholdTwo or More Nonrelated Persons/Members	0.9998
House.Floor.Area	0.0337
House.Age	0.8942
Number.of.bedrooms	0.4197
Electricity	0.0765

```

Household.Head.SexMale                      1.0000
Total.Household.Income:Household.Head.SexMale 0.2698
Total.Food.Expenditure:Household.Head.SexMale 0.0107
Household.Head.Age:Household.Head.SexMale    0.0133
Type.of.HouseholdSingle Family:Household.Head.SexMale 0.9984
Type.of.HouseholdTwo or More Nonrelated Persons/Members:Household.Head.SexMale 0.9998
House.Floor.Area:Household.Head.SexMale      0.0576
House.Age:Household.Head.SexMale             0.9035
Number.of.bedrooms:Household.Head.SexMale    0.8868
Electricity:Household.Head.SexMale           0.0145

```

```

(Intercept)
Total.Household.Income
Total.Food.Expenditure                      ***
Household.Head.Age                          *
Type.of.HouseholdSingle Family
Type.of.HouseholdTwo or More Nonrelated Persons/Members
House.Floor.Area                            *
House.Age
Number.of.bedrooms
Electricity
Household.Head.SexMale
Total.Household.Income:Household.Head.SexMale
Total.Food.Expenditure:Household.Head.SexMale *
Household.Head.Age:Household.Head.SexMale    *
Type.of.HouseholdSingle Family:Household.Head.SexMale
Type.of.HouseholdTwo or More Nonrelated Persons/Members:Household.Head.SexMale
House.Floor.Area:Household.Head.SexMale      .
House.Age:Household.Head.SexMale
Number.of.bedrooms:Household.Head.SexMale
Electricity:Household.Head.SexMale           *

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 912.1 on 1724 degrees of freedom
Residual deviance: 426.7 on 1705 degrees of freedom
AIC: 466.7

```

Number of Fisher Scoring iterations: 20

Model	AIC	BIC
No Interactions	470.76	530.74
Interactions with Head of Household Sex	466.70	575.76