

# R Notebook

*Sanne Meijering  
Hanne Oberman  
Gerbrich Ferdinands*

## Initialization

```
# Load packages  
require(survey)
```

```
## Loading required package: survey  
## Warning: package 'survey' was built under R version 3.4.4  
## Loading required package: grid  
## Loading required package: Matrix  
## Loading required package: survival  
##  
## Attaching package: 'survey'  
## The following object is masked from 'package:graphics':  
##  
##      dotchart
```

```
require(sampling)
```

```
## Loading required package: sampling  
## Warning: package 'sampling' was built under R version 3.4.4  
##  
## Attaching package: 'sampling'  
## The following objects are masked from 'package:survival':  
##  
##      cluster, strata
```

```
require(dplyr)
```

```
## Loading required package: dplyr  
## Warning: package 'dplyr' was built under R version 3.4.4  
##  
## Attaching package: 'dplyr'  
## The following objects are masked from 'package:stats':  
##  
##      filter, lag  
## The following objects are masked from 'package:base':  
##  
##      intersect, setdiff, setequal, union
```

```
require(MASS)
```

```
## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##      select

# Load society dataset
society <- readRDS("Understanding Society innovation pnel wave A.RDS")

# Remove unnecessary columns
# society <- society[,-c(2,3,6,7,15:52, 61, 63, 64, 66:75, 78:80, 82:86, 90, 91, 93)]

# Change variable types
# a_dvage is the average age over all waves. Inspecting the data shows that -6 gives the age during the
society$a_dvage <- as.numeric(society$a_dvage) - 6
```

## Sampling design

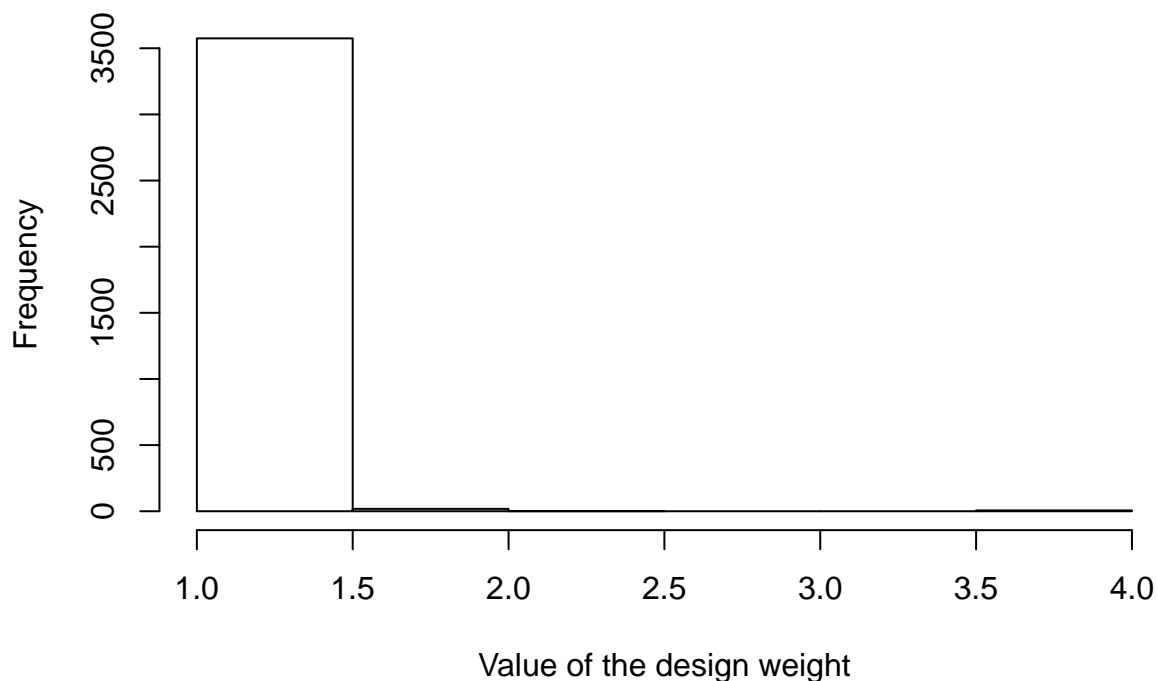
### Question 2

Investigation of design weights

```
# Calculate variance of design weight
Var <- var(society$a_psnenip_xd)
# Combine variance with other descriptive statistics
Descr <- cbind(t(summary(society$a_psnenip_xd)), Var)
# Print with up to 2 decimals
round(Descr, 2)

##      Min. 1st Qu. Median Mean 3rd Qu. Max.  Var
## [1,]    1      1      1 1.01      1     4 0.02

# Create histogram of design weight values
hist(society$a_psnenip_xd,
      breaks = 8,
      main = NULL, #"Histogram of design weight variable 'a_psnenip_xd'",
      xlab = "Value of the design weight")
```



### Question 3

*# Investigate levels of government office region variable*

```
levels(society$a_gor_dv)
```

```
## [1] "missing"           "north east"
## [3] "north west"        "yorkshire and the humber"
## [5] "east midlands"     "west midlands"
## [7] "east of england"   "london"
## [9] "south east"        "south west"
## [11] "wales"             "scotland"
## [13] "northern ireland"
```

```
nrow(society[society$a_gor_dv == "missing",])
```

```
## [1] 0
```

```
nrow(society[society$a_gor_dv == "northern ireland",])
```

```
## [1] 0
```

*# None of the value are missing or northern ireland, so those can be ignored*

*# Run linear regression*

```
coeff <- with(society, lm(a_psnenip_xw ~ a_psnenip_xd + a_sex + a_dvage + a_gor_dv))
```

*# Get coefficients of government office regions*

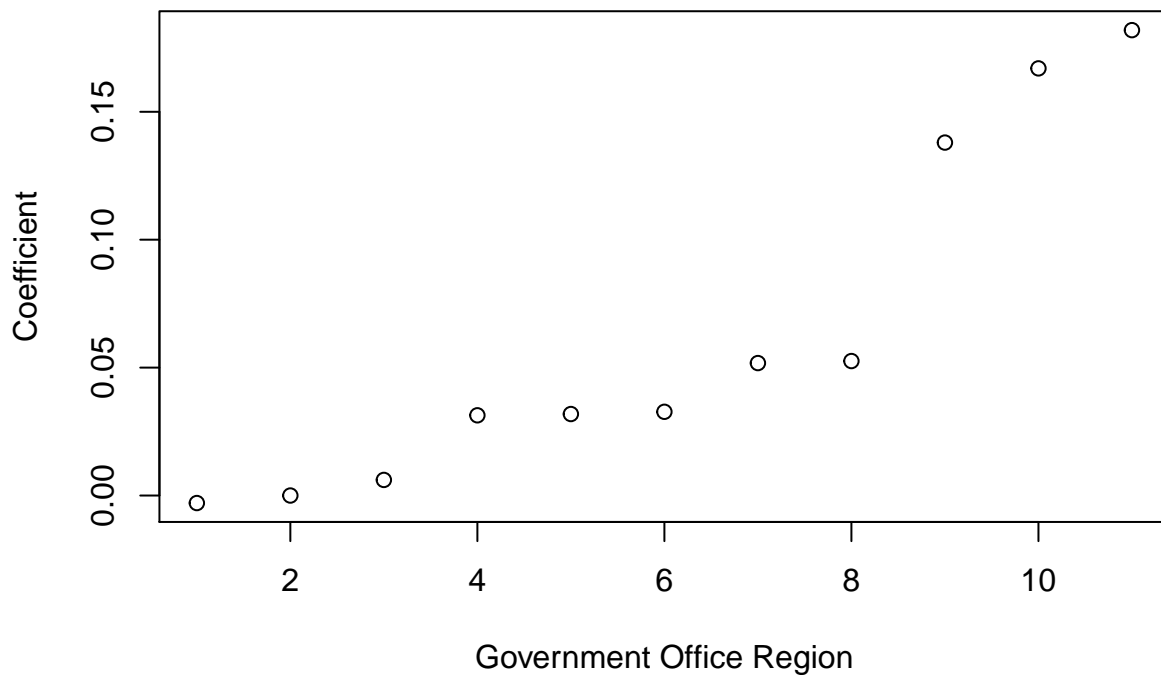
```
coeff <- coeff$coefficients
```

```

coeff.gor <- coeff[5:15]
# Add the base, which has value 0
coeff.gor[11] <- 0
names(coeff.gor)[11] <- "a_gor_dvnorth east"

# Sort and plot coefficients
plot(sort(coeff.gor), xlab = "Government Office Region", ylab = "Coefficient")

```



```

# Show coefficients
coeff.gor

```

```

##          a_gor_dvnorth west a_gor_dvyorkshire and the humber
##          0.052568228          0.031851653
##          a_gor_dveast midlands          a_gor_dvwest midlands
##          0.032723159          0.137962567
##          a_gor_dveast of england          a_gor_dvlondon
##          0.051770584          0.166965124
##          a_gor_dvsouth east          a_gor_dvsouth west
##          -0.002938262          0.006095216
##          a_gor_dvwales          a_gor_dvscotland
##          0.031355723          0.181909406
##          a_gor_dvnorth east
##          0.000000000

```

In accordance with the data, Scotland, London and the West Midlands form one category each, with the fourth category containing all other areas.

```

# Gather grouped region in one variable
society$gor_groups <- "England_Wales"
society$gor_groups[society$a_gor_dv=="london"] <- "London"
society$gor_groups[society$a_gor_dv == "scotland"] <- "Scotland"
society$gor_groups[society$a_gor_dv == "west midlands"] <- "West Midlands"

# Change government office region groups and age into factors
society$gor_groups <- as.factor(society$gor_groups)
society$age_fac <- as.factor(society$a_dvage)

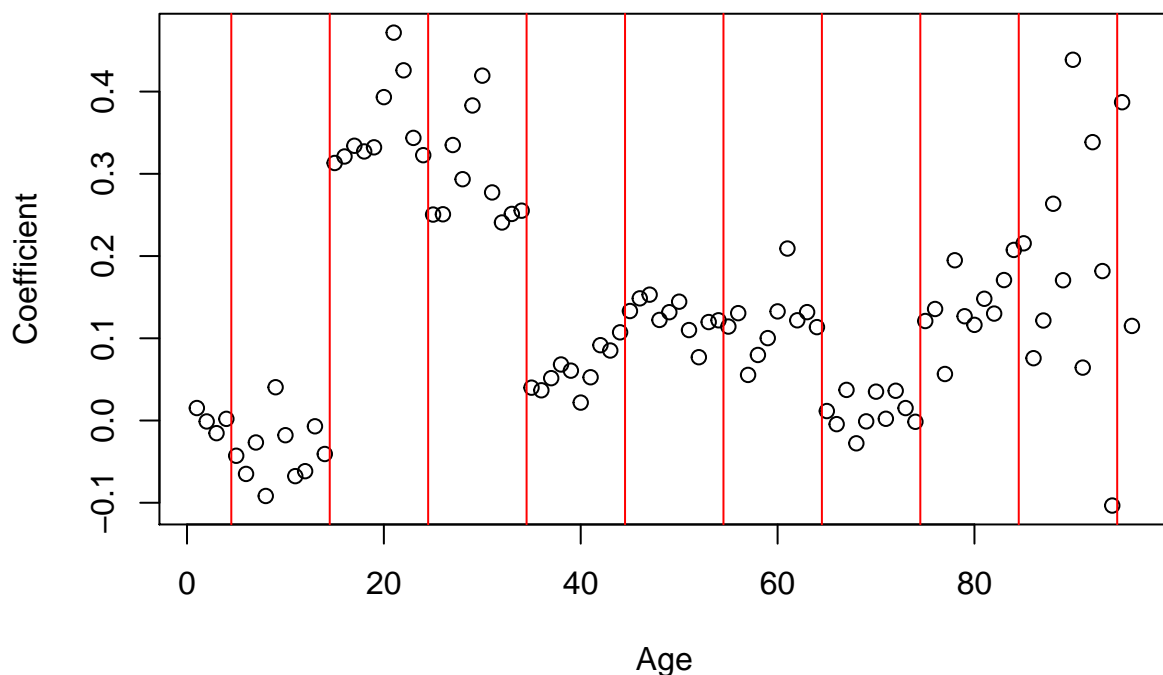
# Rerun linear regression
coeff2 <- with(society, lm(a_psnenip_xw ~ a_psnenip_xd + a_sex + gor_groups + age_fac))

# Get coefficients
coeff2 <- coeff2$coefficients
coeff2.age <- coeff2[7:102]

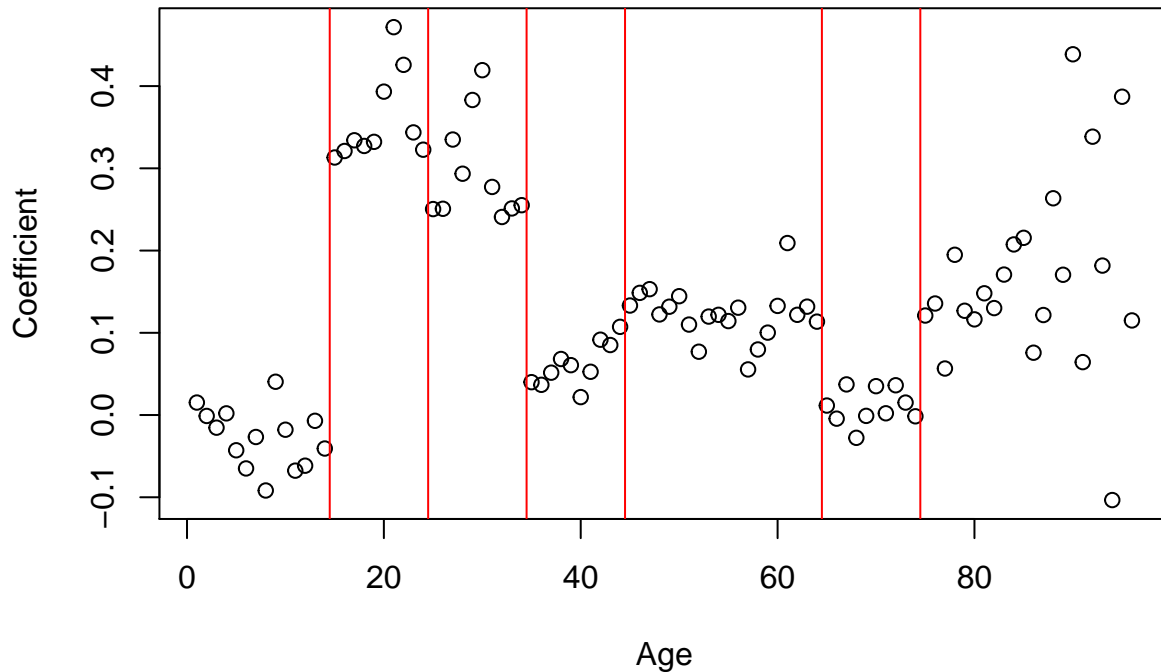
# Change names of coefficients into numbers
x <- unlist(strsplit(names(coeff2.age), "age_fac"))
names(coeff2.age) <- x[x!=""]

# Plot coefficients
plot(as.numeric(names(coeff2.age)), coeff2.age, xlab="Age", ylab="Coefficient")
for (i in seq(4,104,10)+0.5){
  abline(v=i, col="red")
}

```



```
# Plot with group changes
plot(as.numeric(names(coeff2.age)), coeff2.age, xlab="Age", ylab="Coefficient")
for (i in c(14.5,24.5,34.5,44.5,64.5,74.5)){
  abline(v=i, col="red")
}
```



In accordance with the results, age is split up in sections of 10 years, with the 0-10 and 11-20 as well as the 51-60 and 61-79 categories combined and all people of age 81 and older placed into one group.

```
# Assign the levels
levels(society$age_fac)
```

```
## [1] "0" "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11" "12" "13"
## [15] "14" "15" "16" "17" "18" "19" "20" "21" "22" "23" "24" "25" "26" "27"
## [29] "28" "29" "30" "31" "32" "33" "34" "35" "36" "37" "38" "39" "40" "41"
## [43] "42" "43" "44" "45" "46" "47" "48" "49" "50" "51" "52" "53" "54" "55"
## [57] "56" "57" "58" "59" "60" "61" "62" "63" "64" "65" "66" "67" "68" "69"
## [71] "70" "71" "72" "73" "74" "75" "76" "77" "78" "79" "80" "81" "82" "83"
## [85] "84" "85" "86" "87" "88" "89" "90" "91" "92" "93" "94" "95" "96"
```

```
levels(society$age_fac) <- c(rep("group0_14", 15), rep("group15_24", 10),
  rep("group25_34", 10), rep("group35_44", 10),
  rep("group45_64", 20), rep("group65_74", 10),
  rep("group_75", 22))
```

```
# Create table
with(society, table(age_fac, gor_groups))
```

```
##          gor_groups
## age_fac  England_Wales London Scotland West Midlands
## group0_14      511      113      57      53
## group15_24     286       43      21      42
## group25_34     280       56      31      24
## group35_44     404       72      49      49
## group45_64     670       84      70      86
## group65_74     245       26      34      32
## group_75       198       27      20      17

# None of the groups has zero observations

# Run linear regression again
coeff3 <- with(society, lm(a_psnenip_xw ~ a_psnenip_xd + a_sex + gor_groups + age_fac))

#Get coefficients
coeff3 <- coeff3$coefficients
coeff3
```

```
##          (Intercept)          a_psnenip_xd          a_sexfemale
##          -0.15343264           0.98462757          -0.04870161
##      gor_groupsLondon      gor_groupsScotland gor_groupsWest Midlands
##           0.14642803           0.16871017           0.10891317
##      age_facgroup15_24      age_facgroup25_34      age_facgroup35_44
##           0.37993120           0.32321742           0.08846066
##      age_facgroup45_64      age_facgroup65_74      age_facgroup_75
##           0.14922735           0.03522097           0.17247306
```

## Population Estimates

### Question 4

```
# Investigate the a_employ variable.
levels(society$a_employ[society$a_dvage > 15 & society$a_dvage < 64])

## [1] "missing"          "inapplicable"       "proxy respondent"
## [4] "refuse"             "don't know"         "yes"
## [7] "no"
```

```
# The variable a_employ has seven levels.
```

```
summary(society$a_dvage[society$a_employ=="yes"])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      16.00  32.00  42.00  41.55  51.00  96.00
```

```
summary(society$a_dvage[society$a_employ=="no"])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      16.00  39.00  63.00  56.38  73.00  96.00
```

```
# Yes and no contain people of over 21 years of age.
```

```
nrow(society[society$a_employ=="missing",])
```

```
## [1] 0
```

```

nrow(society[society$a_employ=="proxy respondent",])

## [1] 0
# Missing and proxy respondent do not appear in the data

summary(society$a_dvage[society$a_employ=="inapplicable"])

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.000   4.000   8.000   7.787  12.000  15.000
# Inapplicable seems to contain all children and youths of 21 years and younger
# It cannot be assumed that none of them is employed. It can however be assumed that only a
# small part of them is employed, as children under 15 cannot be employed legally
# and most would still be going to a school or university.

nrow(society[society$a_employ=="refuse",])

## [1] 1
nrow(society[society$a_employ=="don't know",])

## [1] 1
# Both refuse and don't know contain one row. These can be treated as missing data.

# Thus, our goal is to compare the proportion of employed people (yes) of working age against
# the number of unemployed people (no, inapplicable), excluding the missing data (refuse, don't know)

with(society, nrow(society[a_employ=="yes" & a_dvage >64,]))

## [1] 41
# There are people older than 64 still working, we should exclude those.
with(society, nrow(society[a_employ=="yes" & a_dvage <15,]))

## [1] 0
# No one younger than 15 years is reported to be working, which is to be expected as it was not
# a question asked to people under 21 years of age.

# Since we wish to know the proportion of employed people of working age, we need 2 groups, one with em
society$employ_dv <- as.numeric(0)
society$employ_dv[society$a_employ=="yes" & society$a_dvage <= 65] <- 1

# Create design
# Don't remove the missing values yet, as the weights are calculated including missing values
Design <- svydesign(ids=~a_hidp, strata=~a_strata, data=society, weights=~a_psnenip_xw)
# Make a subset of non-missing values
Nonmiss <- with(Design, subset(Design, a_employ!="refuse" & a_employ!="don't know"))
svymean(~employ_dv, Nonmiss)

##              mean      SE
## employ_dv 0.46195 0.0098

confint(svymean(~employ_dv, Nonmiss))

##              2.5 %    97.5 %
## employ_dv 0.4427943 0.4810984

```



```
# 46,2% of the population is employed, with a 95% confidence interval of 44.3%-48.1%
```

## Question 5b

```
# Inspect levels of variables
```

```
levels(society$a_livesp_dv)
```

```
## [1] "No" "Yes"
```

```
levels(society$a_cohab_dv)
```

```
## [1] "No" "Yes"
```

```
levels(society$a_single_dv)
```

```
## [1] "No" "Yes"
```

```
levels(society$a_mastat_dv)
```

```
## [1] "Missing"
```

```
## [2] "Inapplicable"
```

```
## [3] "Refusal"
```

```
## [4] "Don't know"
```

```
## [5] "Child under 16"
```

```
## [6] "Single and never married/in civil partnership"
```

```
## [7] "Married"
```

```
## [8] "In a registered same-sex civil partnership"
```

```
## [9] "Separated but legally married"
```

```
## [10] "Divorced"
```

```
## [11] "Widowed"
```

```
## [12] "Separated from civil partner"
```

```
## [13] "A former civil partner"
```

```
## [14] "A surviving civil partner"
```

```
## [15] "Living as couple"
```

```
##### HOUSEHOLD SIZE PER HOUSEHOLD
```

```
# a_hidp - Household identifier
```

```
# Household size: count for each household, how many persons there are in.
```

```
# Get household size per household
```

```
count <- as.matrix(table(society$a_hidp))
```

```
summary(count)
```

```
##          V1
```

```
## Min.      :1.000
```

```
## 1st Qu.:1.000
```

```
## Median :2.000
```

```
## Mean     :2.418
```

```
## 3rd Qu.:3.000
```

```
## Max.     :8.000
```

```
# Turn into dataframe and join to society
```

```
households <- data.frame(a_hidp=as.numeric(rownames(count)), hh_size=count)
```

```
##### NUMBER OF CHILDREN PER HOUSEHOLD
```

```
# Count the number of kids under the age of 16 using the mastat variable
```

```
for(i in households$a_hidp){
```

```
  hh <- society[society$a_hidp==i,]
```

```

households$n_child[households$a_hidp==i]=as.numeric(table(hh$a_mastat_dv)["Child under 16"])
}

##### ANY CHILDREN IN HOUSEHOLD
# Create variable whether the household has children in it (True for households with children)
households$with_child <- households$n_child > 0

##### ANY SINGLE ADULTS IN HOUSEHOLD
# Create variable whether the person is a single adult (true) or not (false)
society$single_adult <- society$a_single_dv == "Yes" & society$a_dvage >= 16
# Create variable whether there is a single adult in the household
for(i in households$a_hidp){
  households$hasSingle[households$a_hidp==i] <- any(society$single_adult[society$a_hidp==i])
}

##### ANY COUPLES IN HOUSEHOLD
# Create variable whether the person is in a couple (true) or not (false)
society$inacouple <- society$a_livesp_dv == "Yes" | society$a_cohab_dv == "Yes"
# Create variable whether there is a couple in the household
for(i in households$a_hidp){
  households$hasCouple[households$a_hidp==i] <- any(society$inacouple[society$a_hidp==i])
}

##### HOUSEHOLD TYPE
# Create matrix containing household states
household_states <- matrix(c("Couple with children", "Couple without children",
                             "Single with children", "Single without children"), nrow=2)
# Create household info variable
households$hh_type <- ""

for (i in seq(1,nrow(households))){
  # Get right row of household state
  if (households$with_child[i]){
    child = 1
  }
  else child = 2

  # Get right column of household state
  if(households$hasCouple[i]){
    state = 1
  }
  else if(households$hasSingle[i]){
    state = 2
  }
  # Select household state and put in households$hh_type
  households$hh_type[i] <- household_states[child, state]
}

# Show how many of each category there are
table(households$hh_type)

```

```

##
##      Couple with children Couple without children      Single with children
##              323              583              112

```

```
## Single without children
## 471
# Join to society dataset
society <- left_join(society, households, by="a_hidp")
```

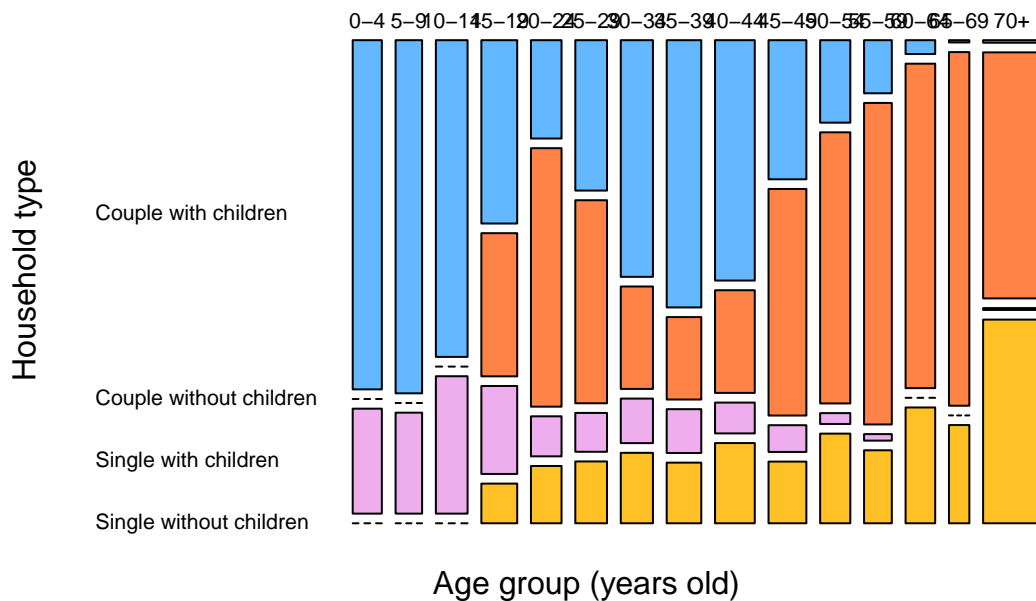
## Question 5d

```
# Create two new age variables: one to plot with the five-year intervals, one with categories to compute
society$a_agegr5a_dv <- factor(society$a_agegr5_dv) #rename levels for readability
levels(society$a_agegr5a_dv) <- c("0-4", "5-9", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39",
  "40-44", "45-49", "50-54", "55-59", "60-64", "65-69", "70+")
society$a_agegr11_dv <- factor(society$a_agegr13_dv) #combine levels so no NAs exist in the contingency
levels(society$a_agegr11_dv) <- c(rep("0-17", 2), "18-19", "20-24", "25-29", "30-34",
  "35-39", "40-44", "45-49", "50-54", "55-59", rep("60+", 2))

# Update sampling design
Design <- svydesign(ids=~pidp, strata=~a_strata, data=society, weights=~a_psnenip_xw)

# Create stacked barplot with 5 year interval variable
tab5 <- svytable(~a_agegr5a_dv + hh_type, Design) # creates a contingency table
plot(tab5, xlab = "Age group (years old)", ylab = "Household type", main = "Household types per age category",
  col=c("steelblue1", "sienna1", "plum2", "goldenrod1"), las = 1, mar = c(3,4,4,2))
```

## Household types per age category



```
# Perform chi square test to evaluate whether age and household type are independent
svychisq(~a_agegr11_dv + hh_type, design=Design, statistic = "Chisq")
```

```
##
## Pearson's X^2: Rao & Scott adjustment
##
## data: svychisq(~a_agegr11_dv + hh_type, design = Design, statistic = "Chisq")
## X-squared = 1763.1, df = 30, p-value < 2.2e-16

#####
# >>>>>> I think this is more interesting with the 5-year variable?
tab <- svytable(~a_agegr11_dv + hh_type, Design) # creates a contingency table
# The Ntotal argument can be either a single number or a data frame whose first column gives the (first
ftable(tab)

##              hh_type Couple with children Couple without children Single with children Single without
## a_agegr11_dv
## 0-17              555.772730              42.555034              196.139619
## 18-19              22.752965              39.572365              15.789405
## 20-24              48.141271              126.500495              19.588616
## 25-29              76.302638              103.007466              19.705917
## 30-34              121.740711              52.679829              22.928404
## 35-39              147.916064              45.618750              24.243592
## 40-44              152.767689              65.253960              19.588328
## 45-49              84.254503              137.213891              16.187806
## 50-54              39.838316              130.933000              5.238851
## 55-59              23.557876              142.619742              2.888893
## 60+                9.654168              496.361554              1.721576

summary(ftable(tab))

##          V1              V2              V3              V4
## Min.   : 9.654   Min.   : 39.57   Min.   : 1.722   Min.   : 6.728
## 1st Qu.: 31.698   1st Qu.: 49.15   1st Qu.: 10.514   1st Qu.: 29.707
## Median : 76.303   Median :103.01   Median : 19.588   Median : 33.559
## Mean   :116.609   Mean   :125.67   Mean   : 31.275   Mean   : 53.724
## 3rd Qu.:134.828   3rd Qu.:134.07   3rd Qu.: 21.317   3rd Qu.: 40.329
## Max.   :555.773   Max.   :496.36   Max.   :196.140   Max.   :274.959

# >>>>>> this is old
#png("age_householdtype", width = 1200, height = 500, res = 72)
#plot(tab, xlab = "Age group (years old)", ylab = "Household type", main = "Household types per age cat
#      col=c("steelblue1","sienna1","plum2","goldenrod1"), las = 1, mar = c(3,4,4,2)
#      )
# BLUE = couple with children
# PINK = single parent with children
# ORANGE = couple without children
# YELLOW = single without children
dev.off() # >>>>>> what is this??

## null device
##          1
#####
```

## Question 6

```
# Summarise a_ivfio
(summary <- summary(society$a_ivfio))
```

##	Missing or wild	Inapplicable
##	0	0
##	Refused	Don't know
##	0	0
##	Full interview	Proxy interview
##	2399	169
##	Telephone intvw	Lost CAPI intvw
##	0	0
##	Refusal	Other non-intvw
##	129	101
##	Moved Ill/away during survey period	
##	0	14
##	Too infirm/elderly	Language difficulties
##	6	6
##	Unknown eligibility	Youth Interview
##	0	257
##	Youth: Refusal	Youth: Oth non-int
##	0	60
##	Child under 10	Youth non-interview
##	459	0
##	Moved/non-int HH	Refusal/non-int HH
##	0	0
##	Lang prob/non-int HH	Age, infirm/non-int HH
##	0	0
##	Non-cont/non-int HH	Out of scope/non-int HH
##	0	0
##	Institutnsd/non-int HH	Child <15 ref/non-int HH
##	0	0
##	Chd <15 lang prob/non-int HH	Chd <15 infirm/non-int HH
##	0	0
##	Chd <15 non-cont/non-int HH	Chd <15 o-o-scope/non-int HH
##	0	0
##	Chd <15 instit/non-int HH	TSM - no OSM/PSM
##	0	0
##	Prev wave adamant refusl	L-t untrace, w-drawn
##	0	0
##	Withdrawn before field	Other ineligible
##	0	0
##	Other Retiring	Dead
##	0	0

```
summary[summary != 0]
```

##	Full interview	Proxy interview
##	2399	169
##	Refusal	Other non-intvw
##	129	101
##	Ill/away during survey period	Too infirm/elderly
##	14	6
##	Language difficulties	Youth Interview
##	6	257
##	Youth: Oth non-int	Child under 10
##	60	459

```

adults <- sum(society$a_ivfio == "Full interview") #the full interviews with adults
youths <- sum(society$a_ivfio == "Youth Interview") #the interviews with children
sum(adults, youths) #all personally completed interviews

## [1] 2656

# The nonresponse indicator is 1 for all (partial) nonresponse and 0 for full (youth) interviews
society$NR <- 1
society$NR[society$a_ivfio == "Youth Interview"] <- 0
society$NR[society$a_ivfio == "Full interview"] <- 0

```

## Question 7

```

# Further investigate the level Language difficulties
society$a_iproxy[society$a_ivfio=="Language difficulties"]

## [1] no interview and no proxy interview - unproductive
## [2] no interview and no proxy interview - unproductive
## [3] no interview and no proxy interview - unproductive
## [4] no interview and no proxy interview - unproductive
## [5] no interview and no proxy interview - unproductive
## [6] no interview and no proxy interview - unproductive
## 7 Levels: missing inapplicable proxy respondent refuse ... no interview and no proxy interview - unp

# It's clearly a complete nonresponse

# Repeat creation of model design from Q4 so NR variable is included
Design <- svydesign(ids=~a_hidp, strata=~a_strata, data=society, weights=~a_psnenip_xw)
Nonmiss <- with(Design, subset(Design, a_employ!="refuse" & a_employ!="don't know"))
# Create model
I_personal <- with(Nonmiss, subset(Nonmiss, NR==0))

# Calculate nonresponse
svymean(~employ_dv, I_personal)

##              mean      SE
## employ_dv 0.50395 0.0113

confint(svymean(~employ_dv, I_personal))

##              2.5 %    97.5 %
## employ_dv 0.4818926 0.5260143

# With NR: 46,2% of the population is employed, with a 95% confidence interval of 44.3%-48.1%
# Exclude NR: 50.4% of the population is employed, with a 95% confidence interval of 48.2%-52.6%

```

## Question 8

```

# Investigate the nonresponse
mean(society$NR) # compute proportion nonresponders

## [1] 0.2622222

# Look at levels of household response
summary(society$a_hhresp_dv)

```

```

##           All present adults interviewed
##                                     2415
## All present adults interview or proxy
##                                     435
##           At least one adult interview
##                                     750
##           No adult interviews
##                                     0

# All households at least filled in the grid, so we only look into person (unit) nonresponse.
# use design weights,
#NRdesign <- svydesign(ids=~a_hidp, strata=~a_strata, data=society, weights=~ )

##### Remove and change problematic variables
# Combine races
society$a_racel_dv <- factor(society$a_racel_dv)
levels(society$a_racel_dv) <- c(rep("missing", 4), "UK Native", rep("White - Nonnative", 2), rep("Mixed", 2))

# Change livewith of same-sex couples to 'yes'
society$a_livewith <- factor(society$a_livewith)
levels(society$a_livewith) <- c("inapplicable", "yes", "no", "yes")

# Combine respm16_dv and respf16_dv
society$a_respm16_dv <- "No"
society$a_respm16_dv[society$a_respf16_dv == "Yes" | society$a_respf16_dv == "Yes"] <- "Yes"

##### MODELS
# model with all predictors
fullmodel <- glm(NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv +
                  a_employ + a_dvage + a_agegr5_dv + a_agegr10_dv + a_agegr13_dv +
                  a_livesp_dv + a_livewith + a_cohab_dv + #a_mastat_dv +
                  a_single_dv + a_depchl_dv + a_rach16_dv + a_respm16_dv +
                  a_nchild_dv + hh_type + hh_size + n_child, family = binomial, data = society)

fullmodel$coefficients

##           (Intercept)
##           17.97002628
## a_gor_dvyorkshire and the humber
##           0.25120269
## a_gor_dvwest midlands
##           0.45737198
## a_gor_dvlondon
##           0.50407740
## a_gor_dvsouth west
##           0.44705180
## a_gor_dvscotland
##           1.13694229
## a_sexfemale
##           -0.22848115
## a_racel_dvWhite - Nonnative
##           -5.84258984
## a_racel_dvAsian or Asian British
##           -5.22295772
## a_gor_dvnorth west
##           0.52313114
## a_gor_dveast midlands
##           -0.23734999
## a_gor_dveast of england
##           -0.18578978
## a_gor_dvsouth east
##           0.85802664
## a_gor_dvwales
##           0.19028687
## a_urban_dvrural area
##           0.27833137
## a_racel_dvUK Native
##           -6.26716363
## a_racel_dvMixed
##           -5.68393705
## a_racel_dvBlack/African/Carribean
##           -6.23981221

```

##	a_race1_dvOther	a_employrefuse
##	-6.16469091	17.31183400
##	a_employdon't know	a_employyes
##	16.00748859	0.73098294
##	a_employno	a_dvage
##	0.54355535	0.01819407
##	a_agegr5_dv5-9 years old	a_agegr5_dv10-14 years old
##	-0.32410764	-21.28686213
##	a_agegr5_dv15-19 years old	a_agegr5_dv20-24 years old
##	-19.68951794	-20.64166629
##	a_agegr5_dv25-29 years old	a_agegr5_dv30-34 years old
##	-20.54152470	-20.83854324
##	a_agegr5_dv35-39 years old	a_agegr5_dv40-44 years old
##	-21.10052997	-21.66147705
##	a_agegr5_dv45-49 years old	a_agegr5_dv50-54 years old
##	-21.20734650	-21.40061765
##	a_agegr5_dv55-59 years old	a_agegr5_dv60-64 years old
##	-21.32896436	-21.40859579
##	a_agegr5_dv65-69 years old	a_agegr5_dv70 years old
##	-22.63678424	-22.11459025
##	a_agegr10_dv10-19 years old	a_agegr10_dv20-29 years old
##	NA	NA
##	a_agegr10_dv30-39 years old	a_agegr10_dv40-49 years old
##	NA	NA
##	a_agegr10_dv50-59 years old	a_agegr10_dv60-69 years old
##	NA	NA
##	a_agegr10_dv70 years or older	a_agegr13_dv16-17 years old
##	NA	-0.71607320
##	a_agegr13_dv18-19 years old	a_agegr13_dv20-24 years old
##	NA	NA
##	a_agegr13_dv25-29 years old	a_agegr13_dv30-34 years old
##	NA	NA
##	a_agegr13_dv35-39 years old	a_agegr13_dv40-44 years old
##	NA	NA
##	a_agegr13_dv45-49 years old	a_agegr13_dv50-54 years old
##	NA	NA
##	a_agegr13_dv55-59 years old	a_agegr13_dv60-64 years old
##	NA	NA
##	a_agegr13_dv65 years or older	a_livesp_dvYes
##	NA	3.40352788
##	a_livewithyes	a_livewithno
##	4.97002155	3.62437432
##	a_cohab_dvYes	a_single_dvYes
##	-1.48576057	NA
##	a_depchl_dvNo	a_rach16_dvYes
##	0.42110865	-0.92275851
##	a_rach16_dvNo	a_resp16_dvYes
##	NA	0.32491384
##	a_nchild_dv	hh_typeCouple without children
##	-0.17867657	0.28108430
##	hh_typeSingle with children	hh_typeSingle without children
##	-0.13038236	0.69136344
##	hh_size	n_child
##	0.26767391	0.12904065



```

# Doesn't produce errors

# Hannes model doesn't:
halffullmodel <- glm(NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_employ + a_dvage +
                      a_agegr5_dv + a_agegr10_dv + a_agegr13_dv + hh_size + n_child +
                      hh_type, family = binomial, data = society)

# model without any predictors
emptymodel <- nothing <- glm(NR ~ 1, family=binomial, data = society )

# backwards selection
backwards <- step(fullmodel)

## Start:  AIC=1357.31
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_employ +
##       a_dvage + a_agegr5_dv + a_agegr10_dv + a_agegr13_dv + a_livesp_dv +
##       a_livewith + a_cohab_dv + a_single_dv + a_depchl_dv + a_rach16_dv +
##       a_resp16_dv + a_nchild_dv + hh_type + hh_size + n_child
##
##
## Step:  AIC=1357.31
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_employ +
##       a_dvage + a_agegr5_dv + a_agegr10_dv + a_agegr13_dv + a_livesp_dv +
##       a_livewith + a_cohab_dv + a_depchl_dv + a_rach16_dv + a_resp16_dv +
##       a_nchild_dv + hh_type + hh_size + n_child
##
##
## Step:  AIC=1357.31
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_employ +
##       a_dvage + a_agegr5_dv + a_agegr13_dv + a_livesp_dv + a_livewith +
##       a_cohab_dv + a_depchl_dv + a_rach16_dv + a_resp16_dv + a_nchild_dv +
##       hh_type + hh_size + n_child
##
##
##           Df Deviance    AIC
## - a_employ      3   1254.0 1352.0
## - hh_type       3   1255.5 1353.5
## - a_dvage        1   1253.5 1355.5
## - a_cohab_dv     1   1253.6 1355.6
## - a_resp16_dv    1   1253.7 1355.7
## - a_depchl_dv    1   1253.7 1355.7
## - n_child        1   1254.0 1356.0
## - a_nchild_dv    1   1254.0 1356.0
## - a_agegr13_dv   1   1254.7 1356.7
## - a_sex          1   1255.0 1357.0
## <none>           1253.3 1357.3
## - a_urban_dv     1   1255.4 1357.4
## - a_rach16_dv     1   1256.1 1358.1
## - a_gor_dv       10  1276.6 1360.6
## - hh_size        1   1259.4 1361.4
## - a_livesp_dv     1   1272.2 1374.2
## - a_livewith      2   1279.7 1379.7
## - a_agegr5_dv     4   1586.3 1682.3
## - a_racel_dv      6   2282.9 2374.9
##

```

```

## Step: AIC=1352.02
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_dvage + a_agegr5_dv +
##       a_agegr13_dv + a_livesp_dv + a_livewith + a_cohab_dv + a_depchl_dv +
##       a_rach16_dv + a_respl6_dv + a_nchild_dv + hh_type + hh_size +
##       n_child
##
##           Df Deviance    AIC
## - hh_type      3   1256.2 1348.2
## - a_dvage       1   1254.3 1350.3
## - a_cohab_dv    1   1254.3 1350.3
## - a_respl6_dv   1   1254.4 1350.4
## - n_child       1   1254.7 1350.7
## - a_nchild_dv   1   1254.8 1350.8
## - a_depchl_dv   1   1254.9 1350.9
## - a_agegr13_dv  1   1255.3 1351.3
## - a_sex         1   1255.8 1351.8
## <none>          1254.0 1352.0
## - a_urban_dv    1   1256.1 1352.1
## - a_rach16_dv   1   1257.0 1353.0
## - a_gor_dv     10   1277.4 1355.4
## - hh_size       1   1260.0 1356.0
## - a_livesp_dv   1   1273.2 1369.2
## - a_livewith    2   1280.5 1374.5
## - a_agegr5_dv   4   1586.4 1676.4
## - a_racel_dv    6   2290.2 2376.2
##
## Step: AIC=1348.25
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_dvage + a_agegr5_dv +
##       a_agegr13_dv + a_livesp_dv + a_livewith + a_cohab_dv + a_depchl_dv +
##       a_rach16_dv + a_respl6_dv + a_nchild_dv + hh_size + n_child
##
##           Df Deviance    AIC
## - a_respl6_dv   1   1256.3 1346.3
## - a_dvage       1   1256.5 1346.5
## - n_child       1   1256.6 1346.6
## - a_cohab_dv    1   1256.6 1346.6
## - a_nchild_dv   1   1256.7 1346.7
## - a_depchl_dv   1   1257.2 1347.2
## - a_agegr13_dv  1   1257.8 1347.8
## - a_sex         1   1257.9 1347.9
## <none>          1256.2 1348.2
## - a_urban_dv    1   1258.5 1348.5
## - a_gor_dv     10   1279.6 1351.6
## - hh_size       1   1263.3 1353.3
## - a_rach16_dv   1   1264.3 1354.3
## - a_livesp_dv   1   1279.3 1369.3
## - a_livewith    2   1281.9 1369.9
## - a_agegr5_dv   4   1590.5 1674.5
## - a_racel_dv    6   2295.1 2375.1
##
## Step: AIC=1346.3
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_dvage + a_agegr5_dv +
##       a_agegr13_dv + a_livesp_dv + a_livewith + a_cohab_dv + a_depchl_dv +
##       a_rach16_dv + a_nchild_dv + hh_size + n_child

```

```

##
##           Df Deviance    AIC
## - a_dvage      1  1256.6 1344.6
## - n_child      1  1256.6 1344.6
## - a_cohab_dv   1  1256.7 1344.7
## - a_nchild_dv  1  1256.7 1344.7
## - a_depchl_dv  1  1257.2 1345.2
## - a_agegr13_dv 1  1257.8 1345.8
## - a_sex        1  1258.2 1346.2
## <none>         1256.3 1346.3
## - a_urban_dv   1  1258.5 1346.5
## - a_gor_dv     10 1279.8 1349.8
## - hh_size      1  1263.3 1351.3
## - a_rach16_dv  1  1269.4 1357.4
## - a_livesp_dv  1  1279.7 1367.7
## - a_livewith   2  1282.0 1368.0
## - a_agegr5_dv  4  1590.6 1672.6
## - a_racel_dv   6  2295.3 2373.3
##
## Step:  AIC=1344.6
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_agegr5_dv +
##       a_agegr13_dv + a_livesp_dv + a_livewith + a_cohab_dv + a_depchl_dv +
##       a_rach16_dv + a_nchild_dv + hh_size + n_child
##
##           Df Deviance    AIC
## - n_child      1  1256.9 1342.9
## - a_cohab_dv   1  1257.0 1343.0
## - a_nchild_dv  1  1257.0 1343.0
## - a_depchl_dv  1  1257.6 1343.6
## - a_agegr13_dv 1  1258.3 1344.3
## - a_sex        1  1258.5 1344.5
## <none>         1256.6 1344.6
## - a_urban_dv   1  1258.8 1344.8
## - a_gor_dv     10 1280.0 1348.0
## - hh_size      1  1263.6 1349.6
## - a_rach16_dv  1  1269.7 1355.7
## - a_livesp_dv  1  1280.0 1366.0
## - a_livewith   2  1282.4 1366.4
## - a_agegr5_dv  4  1794.8 1874.8
## - a_racel_dv   6  2301.0 2377.0
##
## Step:  AIC=1342.93
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_agegr5_dv +
##       a_agegr13_dv + a_livesp_dv + a_livewith + a_cohab_dv + a_depchl_dv +
##       a_rach16_dv + a_nchild_dv + hh_size
##
##           Df Deviance    AIC
## - a_nchild_dv  1  1257.1 1341.1
## - a_cohab_dv   1  1257.3 1341.3
## - a_depchl_dv  1  1257.9 1341.9
## - a_agegr13_dv 1  1258.5 1342.5
## - a_sex        1  1258.8 1342.8
## <none>         1256.9 1342.9
## - a_urban_dv   1  1259.1 1343.1

```

```

## - a_gor_dv      10    1280.8 1346.8
## - a_rach16_dv   1     1269.9 1353.9
## - hh_size       1     1271.6 1355.6
## - a_livesp_dv   1     1280.0 1364.0
## - a_livewith    2     1282.4 1364.4
## - a_agegr5_dv   4     1797.1 1875.1
## - a_racel_dv    6     2301.0 2375.0
##
## Step:  AIC=1341.09
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_agegr5_dv +
##       a_agegr13_dv + a_livesp_dv + a_livewith + a_cohab_dv + a_depchl_dv +
##       a_rach16_dv + hh_size
##
##           Df Deviance    AIC
## - a_cohab_dv   1    1257.5 1339.5
## - a_depchl_dv  1    1258.1 1340.1
## - a_agegr13_dv 1    1258.7 1340.7
## - a_sex        1    1258.8 1340.8
## <none>          1    1257.1 1341.1
## - a_urban_dv   1    1259.2 1341.2
## - a_gor_dv     10    1280.8 1344.8
## - hh_size      1    1273.8 1355.8
## - a_rach16_dv  1    1275.0 1357.0
## - a_livesp_dv  1    1280.2 1362.2
## - a_livewith   2    1283.5 1363.5
## - a_agegr5_dv  4    1797.3 1873.3
## - a_racel_dv   6    2302.9 2374.9
##
## Step:  AIC=1339.46
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_agegr5_dv +
##       a_agegr13_dv + a_livesp_dv + a_livewith + a_depchl_dv + a_rach16_dv +
##       hh_size
##
##           Df Deviance    AIC
## - a_depchl_dv   1    1258.4 1338.4
## - a_agegr13_dv  1    1259.1 1339.1
## - a_sex         1    1259.2 1339.2
## <none>           1    1257.5 1339.5
## - a_urban_dv    1    1259.6 1339.6
## - a_gor_dv     10    1281.0 1343.0
## - hh_size       1    1274.4 1354.4
## - a_rach16_dv   1    1275.3 1355.3
## - a_livesp_dv   1    1280.5 1360.5
## - a_livewith    2    1284.0 1362.0
## - a_agegr5_dv   4    1797.7 1871.7
## - a_racel_dv    6    2303.1 2373.1
##
## Step:  AIC=1338.42
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_agegr5_dv +
##       a_agegr13_dv + a_livesp_dv + a_livewith + a_rach16_dv + hh_size
##
##           Df Deviance    AIC
## - a_sex         1    1260.1 1338.1
## <none>           1    1258.4 1338.4

```

```

## - a_urban_dv      1    1260.4 1338.4
## - a_agegr13_dv    1    1263.3 1341.3
## - a_gor_dv       10    1281.8 1341.8
## - hh_size        1    1275.1 1353.1
## - a_rach16_dv     1    1276.2 1354.2
## - a_livesp_dv     1    1281.6 1359.6
## - a_livewith      2    1285.0 1361.0
## - a_agegr5_dv     4    1798.5 1870.5
## - a_racel_dv      6    2304.6 2372.6
##
## Step:  AIC=1338.12
## NR ~ a_gor_dv + a_urban_dv + a_racel_dv + a_agegr5_dv + a_agegr13_dv +
##       a_livesp_dv + a_livewith + a_rach16_dv + hh_size
##
##           Df Deviance    AIC
## - a_urban_dv      1    1261.9 1337.9
## <none>                1260.1 1338.1
## - a_agegr13_dv    1    1264.7 1340.7
## - a_gor_dv       10    1283.8 1341.8
## - hh_size        1    1278.4 1354.4
## - a_livesp_dv     1    1283.9 1359.9
## - a_livewith      2    1286.8 1360.8
## - a_rach16_dv     1    1287.0 1363.0
## - a_agegr5_dv     4    1800.7 1870.7
## - a_racel_dv      6    2312.0 2378.0
##
## Step:  AIC=1337.89
## NR ~ a_gor_dv + a_racel_dv + a_agegr5_dv + a_agegr13_dv + a_livesp_dv +
##       a_livewith + a_rach16_dv + hh_size
##
##           Df Deviance    AIC
## <none>                1261.9 1337.9
## - a_agegr13_dv    1    1266.3 1340.3
## - a_gor_dv       10    1285.5 1341.5
## - hh_size        1    1280.7 1354.7
## - a_livesp_dv     1    1286.0 1360.0
## - a_livewith      2    1288.6 1360.6
## - a_rach16_dv     1    1288.9 1362.9
## - a_agegr5_dv     4    1801.8 1869.8
## - a_racel_dv      6    2314.1 2378.1

```

```

halfbackwards <- step(halffullmodel)

```

```

## Start:  AIC=1379.35
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_employ +
##       a_dvage + a_agegr5_dv + a_agegr10_dv + a_agegr13_dv + hh_size +
##       n_child + hh_type
##
##
## Step:  AIC=1379.35
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_employ +
##       a_dvage + a_agegr5_dv + a_agegr13_dv + hh_size + n_child +
##       hh_type
##
##           Df Deviance    AIC

```

```

## - a_employ      3    1294.0 1376.0
## - n_child       1    1291.4 1377.4
## - hh_type       3    1295.9 1377.9
## - a_dvage       1    1291.9 1377.9
## <none>          1291.3 1379.3
## - a_urban_dv    1    1293.4 1379.4
## - a_agegr13_dv  1    1293.9 1379.9
## - a_gor_dv     10    1313.8 1381.8
## - a_sex         1    1300.2 1386.2
## - hh_size       1    1304.0 1390.0
## - a_agegr5_dv   4    1634.1 1714.1
## - a_racel_dv    6    2399.7 2475.7
##
## Step: AIC=1376
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_dvage + a_agegr5_dv +
##       a_agegr13_dv + hh_size + n_child + hh_type
##
##           Df Deviance    AIC
## - n_child      1    1294.1 1374.1
## - a_dvage       1    1294.6 1374.6
## - hh_type       3    1299.1 1375.1
## <none>          1294.0 1376.0
## - a_urban_dv    1    1296.1 1376.1
## - a_gor_dv     10    1316.4 1378.4
## - a_sex         1    1304.2 1384.2
## - hh_size       1    1306.0 1386.0
## - a_agegr13_dv  2    1324.6 1402.6
## - a_agegr5_dv   4    1635.1 1709.1
## - a_racel_dv    6    2411.8 2481.8
##
## Step: AIC=1374.06
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_dvage + a_agegr5_dv +
##       a_agegr13_dv + hh_size + hh_type
##
##           Df Deviance    AIC
## - a_dvage       1    1294.6 1372.6
## <none>          1294.1 1374.1
## - a_urban_dv    1    1296.2 1374.2
## - hh_type       3    1301.9 1375.9
## - a_gor_dv     10    1316.4 1376.4
## - a_sex         1    1304.4 1382.4
## - hh_size       1    1312.6 1390.6
## - a_agegr13_dv  2    1325.8 1401.8
## - a_agegr5_dv   4    1635.3 1707.3
## - a_racel_dv    6    2416.2 2484.2
##
## Step: AIC=1372.63
## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_agegr5_dv +
##       a_agegr13_dv + hh_size + hh_type
##
##           Df Deviance    AIC
## <none>          1294.6 1372.6
## - a_urban_dv    1    1296.8 1372.8
## - hh_type       3    1302.2 1374.2

```

```

## - a_gor_dv      10   1317.0 1375.0
## - a_sex         1   1305.0 1381.0
## - hh_size       1   1313.2 1389.2
## - a_agegr13_dv  2   1328.7 1402.7
## - a_agegr5_dv   4   1831.2 1901.2
## - a_racel_dv    6   2421.8 2487.8

# forwards selection to compare backwards model with it.
forwards <- step(emptymodel,
scope=list(lower=formula(emptymodel),upper=formula(fullmodel)), direction="forward")

## Start:  AIC=4144.65
## NR ~ 1
##
##           Df Deviance    AIC
## + a_racel_dv    6   2013.6 2027.6
## + a_agegr5_dv  14   2619.3 2649.3
## + a_agegr10_dv  7   2634.2 2650.2
## + a_agegr13_dv 12   3291.2 3317.2
## + a_rach16_dv   2   3329.0 3335.0
## + a_dvage       1   3336.2 3340.2
## + a_employ      4   3348.4 3358.4
## + a_depchl_dv   1   3444.8 3448.8
## + hh_size       1   3752.8 3756.8
## + n_child       1   3785.1 3789.1
## + hh_type       3   3817.7 3825.7
## + a_single_dv   1   3850.8 3854.8
## + a_livesp_dv   1   3914.0 3918.0
## + a_nchild_dv   1   4080.7 4084.7
## + a_sex         1   4097.3 4101.3
## + a_gor_dv      10   4101.5 4123.5
## + a_cohab_dv    1   4123.2 4127.2
## + a_livewith    2   4122.2 4128.2
## <none>          4142.7 4144.7
## + a_respl6_dv   1   4141.5 4145.5
## + a_urban_dv    1   4142.4 4146.4
##
## Step:  AIC=2027.56
## NR ~ a_racel_dv
##
##           Df Deviance    AIC
## + a_agegr5_dv  14   1404.0 1446.0
## + a_agegr10_dv  7   1575.3 1603.3
## + a_agegr13_dv 12   1881.4 1919.4
## + a_rach16_dv   2   1908.1 1926.1
## + a_employ      4   1912.2 1934.2
## + a_depchl_dv   1   1953.6 1969.6
## + a_livewith    2   1956.9 1974.9
## + hh_type       3   1967.0 1987.0
## + a_respl6_dv   1   1994.1 2010.1
## + a_single_dv   1   1996.0 2012.0
## + a_sex         1   2000.6 2016.6
## + hh_size       1   2003.5 2019.5
## + a_livesp_dv   1   2004.0 2020.0
## + a_cohab_dv    1   2010.2 2026.2

```

```

## + a_nchild_dv    1    2010.5 2026.5
## + n_child        1    2010.9 2026.9
## <none>           2013.6 2027.6
## + a_urban_dv     1    2012.5 2028.5
## + a_dvage        1    2013.5 2029.5
## + a_gor_dv       10    2006.3 2040.3
##
## Step:  AIC=1446.03
## NR ~ a_racel_dv + a_agegr5_dv
##
##           Df Deviance    AIC
## + a_rach16_dv  2    1351.1 1397.1
## + a_agegr13_dv 2    1360.5 1406.5
## + a_employ     4    1359.9 1409.9
## + hh_size      1    1379.0 1423.0
## + hh_type      3    1385.8 1433.8
## + a_depchl_dv  1    1390.9 1434.9
## + a_livewith   2    1389.0 1435.0
## + a_resp16_dv  1    1391.9 1435.9
## + a_sex        1    1393.0 1437.0
## + a_single_dv  1    1399.7 1443.7
## + a_livesp_dv  1    1400.5 1444.5
## + a_urban_dv   1    1401.6 1445.6
## + a_dvage      1    1401.8 1445.8
## <none>         1404.0 1446.0
## + a_gor_dv     10    1384.1 1446.1
## + n_child      1    1402.5 1446.5
## + a_nchild_dv  1    1403.9 1447.9
## + a_cohab_dv   1    1404.0 1448.0
##
## Step:  AIC=1397.13
## NR ~ a_racel_dv + a_agegr5_dv + a_rach16_dv
##
##           Df Deviance    AIC
## + hh_size      1    1316.8 1364.8
## + n_child      1    1336.2 1384.2
## + hh_type      3    1333.8 1385.8
## + a_resp16_dv  1    1344.4 1392.4
## + a_nchild_dv  1    1344.8 1392.8
## + a_single_dv  1    1346.1 1394.1
## + a_livesp_dv  1    1346.3 1394.3
## + a_agegr13_dv 1    1346.6 1394.6
## + a_sex        1    1346.9 1394.9
## + a_gor_dv     10    1329.6 1395.6
## + a_livewith   2    1346.1 1396.1
## + a_depchl_dv  1    1348.1 1396.1
## <none>         1351.1 1397.1
## + a_urban_dv   1    1349.2 1397.2
## + a_dvage      1    1350.6 1398.6
## + a_cohab_dv   1    1351.1 1399.1
## + a_employ     3    1348.6 1400.6
##
## Step:  AIC=1364.81
## NR ~ a_racel_dv + a_agegr5_dv + a_rach16_dv + hh_size

```



```

##
##           Df Deviance    AIC
## + a_gor_dv    10   1293.9 1361.9
## + a_agegr13_dv 1   1312.3 1362.3
## + a_depchl_dv  1   1313.3 1363.3
## <none>                1316.8 1364.8
## + a_sex        1   1314.9 1364.9
## + a_urban_dv   1   1315.1 1365.1
## + a_livewith   2   1313.2 1365.2
## + a_single_dv  1   1315.3 1365.3
## + a_livesp_dv  1   1315.8 1365.8
## + a_dvage      1   1316.0 1366.0
## + a_resp16_dv  1   1316.5 1366.5
## + n_child      1   1316.7 1366.7
## + a_cohab_dv   1   1316.8 1366.8
## + a_nchild_dv  1   1316.8 1366.8
## + a_employ     3   1313.6 1367.6
## + hh_type      3   1314.3 1368.3
##
## Step:  AIC=1361.94
## NR ~ a_racel_dv + a_agegr5_dv + a_rach16_dv + hh_size + a_gor_dv
##
##           Df Deviance    AIC
## + a_agegr13_dv 1   1289.6 1359.6
## + a_depchl_dv  1   1290.3 1360.3
## <none>                1293.9 1361.9
## + a_urban_dv   1   1292.2 1362.2
## + a_single_dv  1   1292.3 1362.3
## + a_sex        1   1292.4 1362.4
## + a_livewith   2   1290.5 1362.5
## + a_livesp_dv  1   1293.0 1363.0
## + a_dvage      1   1293.1 1363.1
## + n_child      1   1293.5 1363.5
## + a_nchild_dv  1   1293.7 1363.7
## + a_cohab_dv   1   1293.8 1363.8
## + a_resp16_dv  1   1293.9 1363.9
## + a_employ     3   1291.0 1365.0
## + hh_type      3   1291.0 1365.0
##
## Step:  AIC=1359.56
## NR ~ a_racel_dv + a_agegr5_dv + a_rach16_dv + hh_size + a_gor_dv +
##       a_agegr13_dv
##
##           Df Deviance    AIC
## <none>                1289.6 1359.6
## + a_urban_dv   1   1287.7 1359.7
## + a_sex        1   1287.8 1359.8
## + a_livewith   2   1286.0 1360.0
## + a_single_dv  1   1288.1 1360.1
## + a_livesp_dv  1   1288.6 1360.6
## + a_depchl_dv  1   1288.8 1360.8
## + a_dvage      1   1289.2 1361.2
## + n_child      1   1289.2 1361.2
## + a_nchild_dv  1   1289.3 1361.3

```

```
## + a_cohab_dv    1    1289.5 1361.5
## + a_respl6_dv   1    1289.5 1361.5
## + hh_type       3    1286.8 1362.8
## + a_employ      3    1287.6 1363.6
```

```
halfforwards <- step(emptymodel,scope=list(lower=formula(emptymodel),
                                             upper=formula(halffullmodel)), direction="forward")
```

```
## Start:  AIC=4144.65
```

```
## NR ~ 1
```

```
##
```

	Df	Deviance	AIC
## + a_racel_dv	6	2013.6	2027.6
## + a_agegr5_dv	14	2619.3	2649.3
## + a_agegr10_dv	7	2634.2	2650.2
## + a_agegr13_dv	12	3291.2	3317.2
## + a_dvage	1	3336.2	3340.2
## + a_employ	4	3348.4	3358.4
## + hh_size	1	3752.8	3756.8
## + n_child	1	3785.1	3789.1
## + hh_type	3	3817.7	3825.7
## + a_sex	1	4097.3	4101.3
## + a_gor_dv	10	4101.5	4123.5
## <none>		4142.7	4144.7
## + a_urban_dv	1	4142.4	4146.4

```
##
```

```
## Step:  AIC=2027.56
```

```
## NR ~ a_racel_dv
```

```
##
```

	Df	Deviance	AIC
## + a_agegr5_dv	14	1404.0	1446.0
## + a_agegr10_dv	7	1575.3	1603.3
## + a_agegr13_dv	12	1881.4	1919.4
## + a_employ	4	1912.2	1934.2
## + hh_type	3	1967.0	1987.0
## + a_sex	1	2000.6	2016.6
## + hh_size	1	2003.5	2019.5
## + n_child	1	2010.9	2026.9
## <none>		2013.6	2027.6
## + a_urban_dv	1	2012.5	2028.5
## + a_dvage	1	2013.5	2029.5
## + a_gor_dv	10	2006.3	2040.3

```
##
```

```
## Step:  AIC=1446.03
```

```
## NR ~ a_racel_dv + a_agegr5_dv
```

```
##
```

	Df	Deviance	AIC
## + a_agegr13_dv	2	1360.5	1406.5
## + a_employ	4	1359.9	1409.9
## + hh_size	1	1379.0	1423.0
## + hh_type	3	1385.8	1433.8
## + a_sex	1	1393.0	1437.0
## + a_urban_dv	1	1401.6	1445.6
## + a_dvage	1	1401.8	1445.8
## <none>		1404.0	1446.0

```

## + a_gor_dv      10   1384.1 1446.1
## + n_child       1   1402.5 1446.5
##
## Step:  AIC=1406.48
## NR ~ a_racel_dv + a_agegr5_dv + a_agegr13_dv
##
##           Df Deviance    AIC
## + hh_size    1   1336.1 1384.1
## + a_sex       1   1349.4 1397.4
## + hh_type     3   1346.7 1398.7
## + n_child     1   1354.9 1402.9
## + a_urban_dv  1   1358.4 1406.4
## <none>        1360.5 1406.5
## + a_gor_dv   10   1340.6 1406.6
## + a_dvage     1   1360.2 1408.2
## + a_employ    3   1356.9 1408.9
##
## Step:  AIC=1384.09
## NR ~ a_racel_dv + a_agegr5_dv + a_agegr13_dv + hh_size
##
##           Df Deviance    AIC
## + a_sex       1   1325.2 1375.2
## + hh_type     3   1328.8 1382.8
## + n_child     1   1333.6 1383.6
## + a_gor_dv   10   1315.7 1383.7
## <none>        1336.1 1384.1
## + a_urban_dv  1   1334.2 1384.2
## + a_employ    3   1330.6 1384.6
## + a_dvage     1   1335.7 1385.7
##
## Step:  AIC=1375.25
## NR ~ a_racel_dv + a_agegr5_dv + a_agegr13_dv + hh_size + a_sex
##
##           Df Deviance    AIC
## + a_gor_dv   10   1304.9 1374.9
## + hh_type     3   1318.9 1374.9
## + a_urban_dv  1   1323.0 1375.0
## <none>        1325.2 1375.2
## + n_child     1   1323.4 1375.4
## + a_dvage     1   1324.8 1376.8
## + a_employ    3   1321.5 1377.5
##
## Step:  AIC=1374.86
## NR ~ a_racel_dv + a_agegr5_dv + a_agegr13_dv + hh_size + a_sex +
##       a_gor_dv
##
##           Df Deviance    AIC
## + hh_type     3   1296.8 1372.8
## + n_child     1   1301.7 1373.7
## + a_urban_dv  1   1302.2 1374.2
## <none>        1304.9 1374.9
## + a_dvage     1   1304.5 1376.5
## + a_employ    3   1301.1 1377.1
##

```

```

## Step: AIC=1372.79
## NR ~ a_racel_dv + a_agegr5_dv + a_agegr13_dv + hh_size + a_sex +
##       a_gor_dv + hh_type
##
##           Df Deviance    AIC
## + a_urban_dv 1    1294.6 1372.6
## <none>          1296.8 1372.8
## + a_dvage     1    1296.2 1374.2
## + n_child     1    1296.7 1374.7
## + a_employ    3    1294.1 1376.1
##
## Step: AIC=1372.63
## NR ~ a_racel_dv + a_agegr5_dv + a_agegr13_dv + hh_size + a_sex +
##       a_gor_dv + hh_type + a_urban_dv
##
##           Df Deviance    AIC
## <none>          1294.6 1372.6
## + a_dvage     1    1294.1 1374.1
## + n_child     1    1294.6 1374.6
## + a_employ    3    1292.0 1376.0
# show different models:
formula(fullmodel)

## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_employ +
##       a_dvage + a_agegr5_dv + a_agegr10_dv + a_agegr13_dv + a_livesp_dv +
##       a_livewith + a_cohab_dv + a_single_dv + a_depchl_dv + a_rach16_dv +
##       a_resp16_dv + a_nchild_dv + hh_type + hh_size + n_child
formula(backwards)

## NR ~ a_gor_dv + a_racel_dv + a_agegr5_dv + a_agegr13_dv + a_livesp_dv +
##       a_livewith + a_rach16_dv + hh_size
formula(forwards)

## NR ~ a_racel_dv + a_agegr5_dv + a_rach16_dv + hh_size + a_gor_dv +
##       a_agegr13_dv
formula(halfforwards)

## NR ~ a_racel_dv + a_agegr5_dv + a_agegr13_dv + hh_size + a_sex +
##       a_gor_dv + hh_type + a_urban_dv
formula(halffullmodel)

## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_employ +
##       a_dvage + a_agegr5_dv + a_agegr10_dv + a_agegr13_dv + hh_size +
##       n_child + hh_type
formula(halfbackwards)

## NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv + a_agegr5_dv +
##       a_agegr13_dv + hh_size + hh_type
# dive deeper into halfbackwards since this is our model of interest
summary(halfbackwards)

##
## Call:

```

```

## glm(formula = NR ~ a_gor_dv + a_urban_dv + a_sex + a_racel_dv +
##       a_agegr5_dv + a_agegr13_dv + hh_size + hh_type, family = binomial,
##       data = society)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4932  -0.3487  -0.2330   0.0001   3.9325
##
## Coefficients: (10 not defined because of singularities)
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      18.15987    689.32392   0.026  0.97898
## a_gor_dvnorth west      0.54928     0.50490   1.088  0.27665
## a_gor_dvyorkshire and the humber  0.23904     0.52235   0.458  0.64722
## a_gor_dveast midlands    -0.14401     0.57453  -0.251  0.80207
## a_gor_dvwest midlands     0.47888     0.52777   0.907  0.36421
## a_gor_dveast of england   -0.14253     0.54241  -0.263  0.79273
## a_gor_dvlondon           0.58688     0.51243   1.145  0.25209
## a_gor_dvsouth east        0.84116     0.49214   1.709  0.08742
## a_gor_dvsouth west        0.49277     0.51426   0.958  0.33796
## a_gor_dvwales            0.18818     0.60303   0.312  0.75499
## a_gor_dvscotland          1.16013     0.51151   2.268  0.02333
## a_urban_dvrural area      0.28284     0.19050   1.485  0.13761
## a_sexfemale             -0.49673     0.15540  -3.196  0.00139
## a_racel_dvUK Native      -6.40958     0.36410 -17.604 < 2e-16
## a_racel_dvWhite - Nonnative -6.02550     0.63409  -9.503 < 2e-16
## a_racel_dvMixed          -5.83151     0.84204  -6.925 4.35e-12
## a_racel_dvAsian or Asian British -5.45720     0.46950 -11.623 < 2e-16
## a_racel_dvBlack/African/Carribean -6.47857     0.82457  -7.857 3.94e-15
## a_racel_dvOther          -6.46329     0.72446  -8.922 < 2e-16
## a_agegr5_dv5-9 years old   -0.23333    981.11604   0.000  0.99981
## a_agegr5_dv10-14 years old -21.12668    689.32372  -0.031  0.97555
## a_agegr5_dv15-19 years old -19.44553    689.32392  -0.028  0.97750
## a_agegr5_dv20-24 years old -15.91029    689.32388  -0.023  0.98159
## a_agegr5_dv25-29 years old -15.86298    689.32386  -0.023  0.98164
## a_agegr5_dv30-34 years old -16.16052    689.32386  -0.023  0.98130
## a_agegr5_dv35-39 years old -16.35101    689.32384  -0.024  0.98108
## a_agegr5_dv40-44 years old -16.68888    689.32385  -0.024  0.98068
## a_agegr5_dv45-49 years old -16.13855    689.32385  -0.023  0.98132
## a_agegr5_dv50-54 years old -16.26146    689.32388  -0.024  0.98118
## a_agegr5_dv55-59 years old -16.13181    689.32388  -0.023  0.98133
## a_agegr5_dv60-64 years old -16.21021    689.32389  -0.024  0.98124
## a_agegr5_dv65-69 years old -17.43135    689.32404  -0.025  0.97983
## a_agegr5_dv70 years old   -16.75895    689.32387  -0.024  0.98060
## a_agegr13_dv16-17 years old  3.46894     0.75324   4.605 4.12e-06
## a_agegr13_dv18-19 years old  4.48115     0.75431   5.941 2.84e-09
## a_agegr13_dv20-24 years old      NA          NA          NA      NA
## a_agegr13_dv25-29 years old      NA          NA          NA      NA
## a_agegr13_dv30-34 years old      NA          NA          NA      NA
## a_agegr13_dv35-39 years old      NA          NA          NA      NA
## a_agegr13_dv40-44 years old      NA          NA          NA      NA
## a_agegr13_dv45-49 years old      NA          NA          NA      NA
## a_agegr13_dv50-54 years old      NA          NA          NA      NA
## a_agegr13_dv55-59 years old      NA          NA          NA      NA
## a_agegr13_dv60-64 years old      NA          NA          NA      NA

```

## a_agegr13_dv65 years or older	NA	NA	NA	NA
## hh_size	0.32727	0.07485	4.372	1.23e-05
## hh_typeCouple without children	0.53579	0.24882	2.153	0.03129
## hh_typeSingle with children	-0.18336	0.29446	-0.623	0.53348
## hh_typeSingle without children	0.08884	0.39247	0.226	0.82093
##				
## (Intercept)				
## a_gor_dvnorth west				
## a_gor_dvyorkshire and the humber				
## a_gor_dveast midlands				
## a_gor_dvwest midlands				
## a_gor_dveast of england				
## a_gor_dvlondon				
## a_gor_dvsouth east	.			
## a_gor_dvsouth west				
## a_gor_dvwales				
## a_gor_dvscotland	*			
## a_urban_dvrural area				
## a_sexfemale	**			
## a_racel_dvUK Native	***			
## a_racel_dvWhite - Nonnative	***			
## a_racel_dvMixed	***			
## a_racel_dvAsian or Asian British	***			
## a_racel_dvBlack/African/Carribean	***			
## a_racel_dvOther	***			
## a_agegr5_dv5-9 years old				
## a_agegr5_dv10-14 years old				
## a_agegr5_dv15-19 years old				
## a_agegr5_dv20-24 years old				
## a_agegr5_dv25-29 years old				
## a_agegr5_dv30-34 years old				
## a_agegr5_dv35-39 years old				
## a_agegr5_dv40-44 years old				
## a_agegr5_dv45-49 years old				
## a_agegr5_dv50-54 years old				
## a_agegr5_dv55-59 years old				
## a_agegr5_dv60-64 years old				
## a_agegr5_dv65-69 years old				
## a_agegr5_dv70 years old				
## a_agegr13_dv16-17 years old	***			
## a_agegr13_dv18-19 years old	***			
## a_agegr13_dv20-24 years old				
## a_agegr13_dv25-29 years old				
## a_agegr13_dv30-34 years old				
## a_agegr13_dv35-39 years old				
## a_agegr13_dv40-44 years old				
## a_agegr13_dv45-49 years old				
## a_agegr13_dv50-54 years old				
## a_agegr13_dv55-59 years old				
## a_agegr13_dv60-64 years old				
## a_agegr13_dv65 years or older				
## hh_size	***			
## hh_typeCouple without children	*			
## hh_typeSingle with children				

```

## hh_typeSingle without children
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 4142.7  on 3599  degrees of freedom
## Residual deviance: 1294.6  on 3561  degrees of freedom
## AIC: 1372.6
##
## Number of Fisher Scoring iterations: 18

# significant variables are a_agegr13, hh_size, hh_type, sex, urban and a_gor
# significant with p < .001 are sex, a_agegr13_dv and hh_size. plot those (see below)

# evaluate fit
fit <- anova(halfbackwards, halffullmodel)
library(lmtest)

## Warning: package 'lmtest' was built under R version 3.4.4
## Loading required package: zoo
## Warning: package 'zoo' was built under R version 3.4.4
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

test <- lrtest(halfbackwards, halffullmodel)

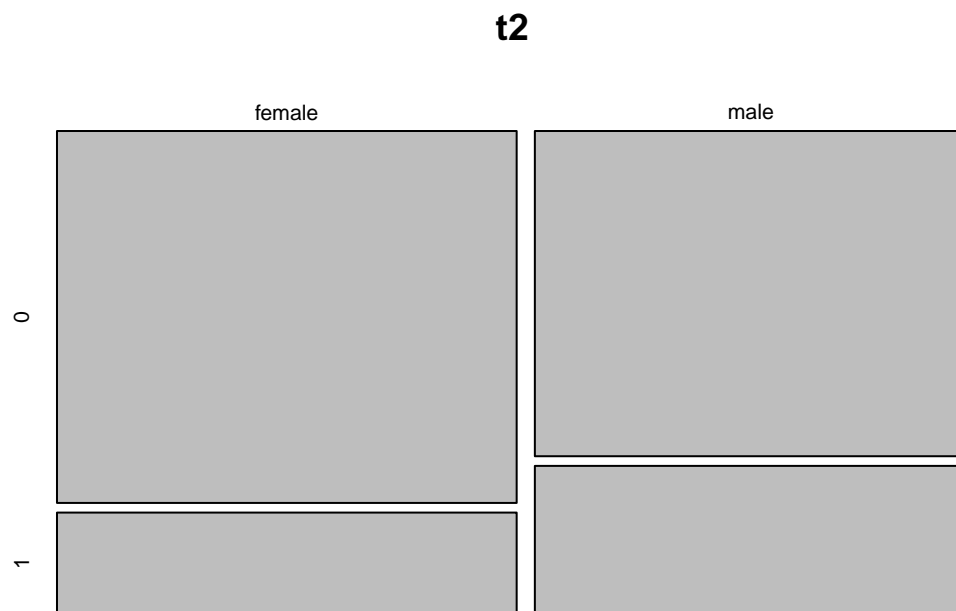
# interesting: a_agegr13_dv produces NA's for every category except 16-17 and 18-19 years old..?
# i tried to load all kinds of packages to compute a pseudo r squared but they didn't work..

# plot nonresponse
tbl <- table(society$NR)
t1 <- tbl/3600*100

# plot relationship sex / nonresponse
t2 <- table(as.matrix(society$a_sex), as.matrix(society$NR)) # females respond more often then males do

plot(t2)

```

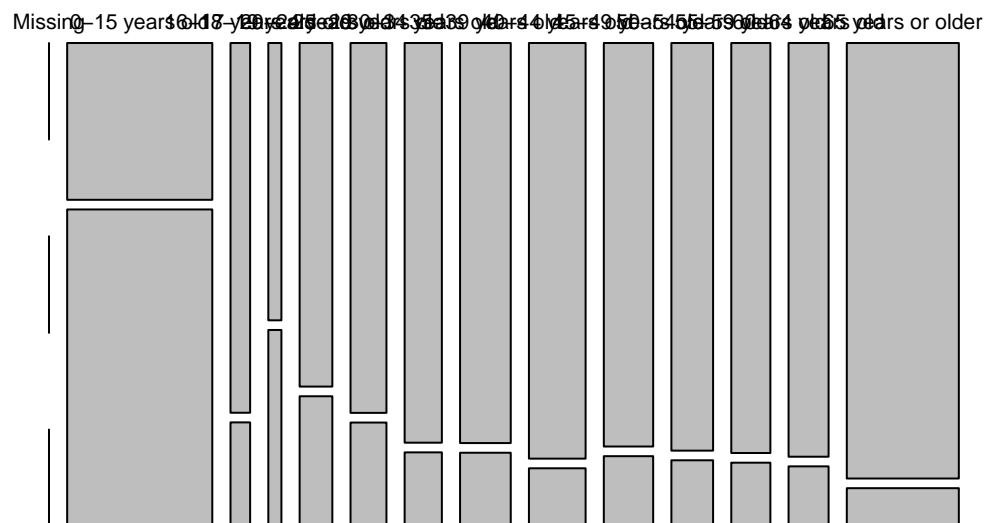


```
# plot relationship age / nonresponse
t3 <- table(society$a_agegr13_dv, society$NR)

plot(t3) # the higher the age category, the less nonresponse.
```



t3



```
# plot relationship hh_size / non response
t4 <- table(society$hh_size, society$NR) # the bigger the household, the more nonresponders
plot(t4)
```

**t4**

