Code

Ghana Agricultural Profits Analysis Steps and Explanations

Group 07 - Siqi Zhang, Huibo Jia, Qiyu Ye, Mark Russeff

12-06-2018

Read data

After reading through the documentation, we determined these data files contain the data we want to use to analyze the influencing factors of Ghana agricultural profits

Hide

```
# list of data file pairs to read, first is subfolder, second is file name
datafiles <- list(c("aggregates/", "agg2"),
                  c("", "sec8b"),
                  c("", "sec0a"),
                  c("", "sec1"),
                  c("", "sec2a"),
                  c("", "sec8a2"),
                  c("", "sec8a3"),
                  c("", "sec8c1"),
                  c("", "sec8c2"),
                  c("community/", "cs2"))
invisible(sapply(datafiles, readRawData))
```

Calculate agricultural profit per area unit

First we need to convert different area unit into a same one. Looking at count of each unit, acre is the major one, so we decide to map other units to acre. According to G4USERSG.pdf, we create a map between each unit to a multiplier of it when it's converted to acre (the "Other" unit is treated as NA). Then join the map onto `sec8b` and multiply with original land size to get unified land size (values greater than 1e+100 are treated as NA). We noticed that some households have multiple entries of land size info and among them there is NA value. So we convert NA to 0 so we can sum up land size for each household. Eventually, household with land size less than 0.5 acre are filtered out because that's probably not a real farm.

Area unit	To Acre
1 Acre	1 Acre
1 Poles	1 Acre

1 Ropes	1/9 Acre
Other	NA

Hide

```
# map land size units to acre with a multiplier (s8bq4bm)
# the "Other" unit is treated as NA
table(sec8b$s8bq4b)
```

```
                   1                    2                    3
                7725                 1477                 1354
                   4 1.74980057982641e+100
                  13                    1
```

Hide

```
land_size_unit_map <- data.frame("s8bq4b" = as.double(c(1:4)),
                                 "s8bq4bm" = c(1, 1, 1/9, NA))
# convert all land size to acre
land_size_info <- sec8b %>%
  select(clust, nh, s8bq4a, s8bq4b) %>%
  mutate(s8bq4a = ifelse(s8bq4a > 1e+100, NA, s8bq4a),
         s8bq4b = ifelse(s8bq4b > 1e+100, NA, s8bq4b)) %>%
  left_join(land_size_unit_map) %>%
  # s8bq4ac is corrected s8bq4a, which is land size in acre
  mutate(s8bq4ac = s8bq4a * s8bq4bm) %>%
  # convert NA to 0 to calculate sum of land size for each household
  replace(., is.na(.), 0) %>%
  group_by(clust, nh) %>%
  summarise(landSize = sum(s8bq4ac)) %>%
  # remove household with small land size
  filter(landSize >= 0.5)
```

```
Joining, by = "s8bq4b"
```

Then we use agricultural income minus depreciation of farm equipment (`agri1 - hhagdepn`) as household profit, and devide it by household land size to get profit per area unit. `agg2` does not have NA value so no need to clean up.

Hide

```
# identify count of NAs in agg2
colSums(is.na(agg2))
```

```
   clust       nh    agri1   agri1c    agri2   agri2c hhagdepn
       0        0        0        0        0        0        0
```

Hide

```
# check correlation between agricultural income and corrected agricultural income
cor(agg2$agri1, agg2$agri1c)
```

```
[1] 1
```

Hide

```
# resolve warnings during join
attr(agg2$clust, "label") <- "Enumeration Area number"
attr(agg2$nh, "label") <- "Household ID"
# calculate household agri profit per acre
# because the above correlations are both 1,
# use only "agri1" to calculate agricultural profit
hh_profit_info <- agg2 %>%
  select(clust, nh, agri1, hhagdepn) %>%
  filter(agri1 != 0) %>%
  inner_join(land_size_info, by = c("clust", "nh")) %>%
  mutate(profit = (agri1 - hhagdepn) / landSize) %>%
  select(-agri1, -hhagdepn, -landSize)
attr(hh_profit_info$profit, "label") <- "HH agri profit"
```

Tidy household basic information

For basic household information, we are interested in language of respondent (values greater than 1e+100 are treated as “Unknown” language), ecological zone number and 3 different locality classifications. All variables are converted to factor type so they could be used as dummy variables in linear model.

Hide

```
# household basic information
hh_basic_info <- sec0a %>%
  select(region, district, eanum, clust, nh, reslan, ez:loc3) %>%
  mutate(reslan = cleanAndFactorize(reslan, 1e+100, 99,
                                    levels = as.character(c(1:8, 99)),
                                    labels = c("English", "Akan", "Ewe", "GaAdangbe",
                                               "Dagbani", "Hausa", "Nzema", "Other",
"Unknown")),
         ez = factor(ez,
                     levels = as.character(c(1:3)),
                     labels = c("Coastal", "Forest", "Savannah")),
         loc2 = factor(loc2,
                       levels = as.character(c(1,2)),
                       labels = c("Urban", "Rural")),
         loc5 = factor(loc5,
                       levels = as.character(c(1:5)),
                       labels = c("Accra", "OtherUrban", "RuralCoastal",
                                  "RuralForest", "RuralSavannah")),
         loc3 = factor(loc3,
                       levels = as.character(c(1:3)),
                       labels = c("Accra", "OtherUrban", "Rural"))
  )
```

Tidy household member information

We picked some data about household members such as sex, age and education and aggregate them on household level to analyze the effect of variety of household members on agricultural profit.

Select household members sex and age information and calculate mean, max and min age of household. The relation (`rel`) data is used to filter by head of household later.

Hide

```
# household member information
hhm_info <- sec1 %>%
  select(clust, nh, pid, sex, agey, rel) %>%
  group_by(clust, nh) %>%
  mutate(female = sex == 2,
         age = agey,
         avgAge = mean(agey),
         maxAge = max(agey),
         minAge = min(agey)
  ) %>%
  select(-sex, -agey)
```

For education data, we categorize educational qualification (`s2aq3`) into 5 levels: “None”, “BasicEducation”, “SecondaryEducation”, “TertiaryEducation” and “Other”, based on information from the two sources below. For missing value in `s2aq3` , it’s because these people never attended school (`s2aq1 == 2`). In this case we set education to None.

Education System in Ghana (https://www.scholaro.com/pro/Countries/Ghana/Education-System)

Education in Ghana Wikipedia (https://en.wikipedia.org/wiki/Education_in_Ghana)

Educational Qualification	**Education Level**
01 None	None
02 MSLC?BECE	BasicEducation
03 Voc/Comm	SecondaryEducation
04 ‘O’ Level	SecondaryEducation
05 SSS	SecondaryEducation
06 ‘A’ Level	SecondaryEducation
07 T/T Cert. B	TertiaryEducation
08 T /T Cert. A	TertiaryEducation
09 Nursing	TertiaryEducation
10 Tech/Prof Cert.	TertiaryEducation
11 Tech/Prof Dip	TertiaryEducation
12 Bachelor	TertiaryEducation
13 Masters	TertiaryEducation
14 Doctorate	TertiaryEducation
96 Other	Other

Hide

```
# household member education information
# map education qualification to education level
educ_level_map <- data.frame("s2aq3" = c(1:14, 96),
                             "s2aq3l" = c(1, 2, rep(3, 4), rep(4, 8), 5))
# resolve warnings
attr(educ_level_map$s2aq3, "label") <- "Highest educ qualification"
attr(educ_level_map$s2aq3, "format.stata") <- "%10.0g"
# Check NAs
colSums(is.na(sec2a))
```

```
     nh     pid   s1q23   clust   s2aq1   s2aq2   s2aq3   s2aq4  s2aq5a
      0       0       0       0       0       0       0       0       0
 s2aq5b   s2aq6   s2aq7   s2aq8   s2aq9  s2aq10  s2aq11  s2aq12  s2aq13
      0       0       0       0       0       0       0       0       0
s2aq13a  s2aq14  s2aq15  s2aq16  s2aq17
      0       0       0       0       0
```

Hide

```
hhm_educ <- sec2a %>%
  select(clust, nh, pid, s2aq1, s2aq3) %>%
  # if never attended school (s2aq1 == 2), set education to None (s2aq3 = 1)
  mutate(s2aq3 = ifelse(s2aq1 == 2, 1, s2aq3)) %>%
  left_join(educ_level_map) %>%
  mutate(educ = factor(s2aq3l,
                       levels = as.character(c(1:5)),
                       labels = c("None", "BasicEducation", "SecondaryEducation",
                                  "TertiaryEducation", "Other"))) %>%
  select(-s2aq1, -s2aq3, -s2aq3l)
```

```
Joining, by = "s2aq3"
```

Then we join household member info and education info, and use data from head of household (`rel == 1`) as a representative of the household.

Hide

```
hh_head_info <- hhm_info %>%
  inner_join(hhm_educ, by=c("clust", "nh", "pid")) %>%
  filter(rel == 1) %>% # head of household
  select(-pid, -rel)
```

Tidy agricultural characteristics information

We included agricultural characteristics information in our analysis as well, includign Livestock and Fishing (`sec8a2`), Agricultural equipment (`sec8a3`) and Crops (`sec8c1` and `sec8c2`). We would like to see if any certain types of livestock, equipment and/or crop would contribute a lot to profit.

For livestock data, we found that there are too many missing values (value greater than 1e+100) in its unit of messaure (`s8aq22b`). So we decided to spread livestock count by livestock type so that each livestock type is a column with the count as its value, and assume same livestock type has same unit of messaure so we can compare accros them. We also created a column for the count of different livestock types each household owns.

Hide

```
table(sec8a2$s8aq22b)
```

```
                    1                     2                     3 
                   32                    41                    22 
                    4 1.74980057982641e+100                       
                   30                  4716                       
```

Hide

```
# spread livestock count and count livestock type
# since there are too many missing value in livestock unit of messaure (s8aq22b),
# we assume same livestock type has same unit,
# and we only take livestock count (s8aq22a) into consideration
hh_livestock_info <- sec8a2 %>%
  select(clust, nh, livstcd, s8aq22a) %>%
  group_by(clust, nh) %>%
  mutate(livstcdTypeCount = n()) %>%
  spread(key = livstcd,
         value = s8aq22a,
         fill = 0,
         sep = "")
summary(hh_livestock_info)
```

```
     clust            nh         livstcdTypeCount    livstcd1      
 Min.   :4002   Min.   : 1.00   Min.   :1.00     Min.   :0.0000  
 1st Qu.:4395   1st Qu.: 6.00   1st Qu.:1.00     1st Qu.:0.0000  
 Median :4589   Median :11.00   Median :2.00     Median :0.0000  
 Mean   :4580   Mean   :10.79   Mean   :2.13     Mean   :0.1205  
 3rd Qu.:4855   3rd Qu.:16.00   3rd Qu.:3.00     3rd Qu.:0.0000  
 Max.   :4999   Max.   :25.00   Max.   :7.00     Max.   :9.0000  
    livstcd2          livstcd3          livstcd4          livstcd5      
 Min.   : 0.0000   Min.   :      0   Min.   :      0   Min.   : 0.0000  
 1st Qu.: 0.0000   1st Qu.:      0   1st Qu.:      0   1st Qu.: 0.0000  
 Median : 0.0000   Median :      0   Median :      1   Median : 0.0000  
 Mean   : 0.8517   Mean   :    464   Mean   :   1178   Mean   : 0.6881  
 3rd Qu.: 0.0000   3rd Qu.:      3   3rd Qu.:      5   3rd Qu.: 0.0000  
 Max.   :90.0000   Max.   :1050000   Max.   :2670000   Max.   :36.0000  
    livstcd6        livstcd7          livstcd8          livstcd9      
 Min.   :0.000   Min.   :     0.0   Min.   :  0.000   Min.   : 0.00000  
 1st Qu.:0.000   1st Qu.:     1.0   1st Qu.:  0.000   1st Qu.: 0.00000  
 Median :0.000   Median :     6.0   Median :  0.000   Median : 0.00000  
 Mean   :0.022   Mean   :   146.1   Mean   :  2.227   Mean   : 0.07787  
 3rd Qu.:0.000   3rd Qu.:    12.0   3rd Qu.:  0.000   3rd Qu.: 0.00000  
 Max.   :8.000   Max.   :130000.0   Max.   :150.000   Max.   :40.00000  
   livstcd10          livstcd11          livstcd12      
 Min.   :    0.000   Min.   :  0.00000   Min.   : 0.00000  
 1st Qu.:    0.000   1st Qu.:  0.00000   1st Qu.: 0.00000  
 Median :    0.000   Median :  0.00000   Median : 0.00000  
 Mean   :    9.321   Mean   :  0.06599   Mean   : 0.06247  
 3rd Qu.:    0.000   3rd Qu.:  0.00000   3rd Qu.: 0.00000  
 Max.   :13000.000   Max.   :100.00000   Max.   :50.00000  
```

Similarly we process the agricultural equipment data.

Hide

```
# spread agricultural equipment count and count agricultural equipment type
hh_equip_info <- sec8a3 %>%
  select(clust, nh, eqcdown, s8aq34) %>%
  replace(., . > 1e+100, 0) %>%
  filter(s8aq34 != 0) %>%
  group_by(clust, nh) %>%
  mutate(equipTypeCount = n()) %>%
  spread(key = eqcdown,
         value = s8aq34,
         fill = 0,
         sep = "")
summary(hh_equip_info)
```

```
     clust            nh         equipTypeCount    eqcdown21
 Min.   :4002   Min.   : 1.00   Min.   :1.000   Min.   :0.000000
 1st Qu.:4125   1st Qu.: 6.00   1st Qu.:1.000   1st Qu.:0.000000
 Median :4539   Median :11.00   Median :1.000   Median :0.000000
 Mean   :4481   Mean   :10.71   Mean   :1.142   Mean   :0.003823
 3rd Qu.:4795   3rd Qu.:16.00   3rd Qu.:1.000   3rd Qu.:0.000000
 Max.   :4992   Max.   :25.00   Max.   :4.000   Max.   :1.000000
   eqcdown22          eqcdown31       eqcdown51          eqcdown61
 Min.   :0.00000   Min.   :0.000   Min.   :  0.0000   Min.   :0.0000
 1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:  0.0000   1st Qu.:0.0000
 Median :0.00000   Median :0.000   Median :  0.0000   Median :0.0000
 Mean   :0.03287   Mean   :0.013   Mean   :  0.2378   Mean   :0.0107
 3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:  0.0000   3rd Qu.:0.0000
 Max.   :2.00000   Max.   :2.000   Max.   :160.0000   Max.   :1.0000
   eqcdown62          eqcdown63          eqcdown64        eqcdown65
 Min.   :  0.0000   Min.   :  0.0000   Min.   :0.0000   Min.   :  0.000
 1st Qu.:  0.0000   1st Qu.:  0.0000   1st Qu.:0.0000   1st Qu.:  1.000
 Median :  0.0000   Median :  0.0000   Median :0.0000   Median :  3.000
 Mean   :  0.2791   Mean   :  0.6017   Mean   :0.0107   Mean   :  3.993
 3rd Qu.:  0.0000   3rd Qu.:  0.0000   3rd Qu.:0.0000   3rd Qu.:  4.000
 Max.   :130.0000   Max.   :236.0000   Max.   :7.0000   Max.   :800.000
```

For crop data, we noticed that some households have multiple entries of same crop type. So we need to group by household and crop type and sum up across same crop type.

Hide

```
# spread household havested crop count and count havested crop type
# some households have multiple entries of same crop type, for example:
(sec8c1 %>% select(clust, nh, cropcd, s8cq3a))[c(377, 378),]
```

clust	nh	cropcd	s8cq3a
<dbl>	<dbl>	<dbl>	<dbl>
4049	18	1	10.0
4049	18	1	0.5

2 rows

Hide

```
# so need to sum up quantity of same crop type
hh_crop_info <- sec8c1 %>%
  select(clust, nh, cropcd, s8cq3a) %>%
  replace(., . > 1e+100, 0) %>%
  group_by(clust, nh, cropcd) %>%
  # sum up quantity of same crop type
  summarise(s8cq3ac = sum(s8cq3a)) %>%
  group_by(clust, nh) %>%
  mutate(cropTypeCount = n()) %>%
  spread(key = cropcd,
         value = s8cq3ac,
         fill = 0,
         sep = "")
summary(hh_crop_info)
```

```
     clust            nh        cropTypeCount      cropcd0        
 Min.   :4002   Min.   : 1.00   Min.   :1.000   Min.   : 0.000000  
 1st Qu.:4392   1st Qu.: 6.00   1st Qu.:1.000   1st Qu.: 0.000000  
 Median :4572   Median :11.00   Median :2.000   Median : 0.000000  
 Mean   :4560   Mean   :10.83   Mean   :2.012   Mean   : 0.003218  
 3rd Qu.:4819   3rd Qu.:16.00   3rd Qu.:3.000   3rd Qu.: 0.000000  
 Max.   :4999   Max.   :25.00   Max.   :7.000   Max.   :12.000000  
    cropcd1           cropcd2             cropcd3         
 Min.   :  0.000   Min.   :  0.00000   Min.   :0.0000000  
 1st Qu.:  0.000   1st Qu.:  0.00000   1st Qu.:0.0000000  
 Median :  0.000   Median :  0.00000   Median :0.0000000  
 Mean   :  3.781   Mean   :  0.06846   Mean   :0.0008045  
 3rd Qu.:  0.000   3rd Qu.:  0.00000   3rd Qu.:0.0000000  
 Max.   :800.000   Max.   :227.80000   Max.   :3.0000000  
    cropcd4            cropcd5            cropcd6         
 Min.   :   0.000   Min.   :0.000000   Min.   :  0.00000  
 1st Qu.:   0.000   1st Qu.:0.000000   1st Qu.:  0.00000  
 Median :   0.000   Median :0.000000   Median :  0.00000  
 Mean   :   9.701   Mean   :0.002145   Mean   :  0.03433  
 3rd Qu.:   0.000   3rd Qu.:0.000000   3rd Qu.:  0.00000  
 Max.   : 999.990   Max.   :6.000000   Max.   :125.00000  
    cropcd8            cropcd9             cropcd10        
 Min.   :  0.0000   Min.   : 0.000000   Min.   :  0.000  
 1st Qu.:  0.0000   1st Qu.: 0.000000   1st Qu.:  0.000  
 Median :  0.0000   Median : 0.000000   Median :  0.000  
 Mean   :  0.5093   Mean   : 0.004023   Mean   :  6.393  
 3rd Qu.:  0.0000   3rd Qu.: 0.000000   3rd Qu.:  0.000  
 Max.   :999.0000   Max.   :15.000000   Max.   :300.000  
    cropcd11           cropcd12           cropcd13           cropcd14      
 Min.   : 0.00000   Min.   : 0.00000   Min.   :  0.0000   Min.   :  0.000  
 1st Qu.: 0.00000   1st Qu.: 0.00000   1st Qu.:  0.0000   1st Qu.:  0.000  
```

```
Median : 0.00000   Median : 0.00000   Median :  0.0000   Median :  0.000
Mean   : 0.01341   Mean   : 0.02105   Mean   :  0.1077   Mean   :  1.671
3rd Qu.: 0.00000   3rd Qu.: 0.00000   3rd Qu.:  0.0000   3rd Qu.:  0.000
Max.   :50.00000   Max.   :61.00000   Max.   :302.0000   Max.   :132.000
   cropcd15           cropcd16           cropcd17          cropcd18
Min.   :  0.0000   Min.   : 0.00000   Min.   :  0.000   Min.   :  0.000
1st Qu.:  0.0000   1st Qu.: 0.00000   1st Qu.:  0.000   1st Qu.:  0.000
Median :  0.0000   Median : 0.00000   Median :  0.000   Median :  0.000
Mean   :  0.1483   Mean   : 0.01341   Mean   :  2.198   Mean   :  0.341
3rd Qu.:  0.0000   3rd Qu.: 0.00000   3rd Qu.:  0.000   3rd Qu.:  0.000
Max.   :200.0000   Max.   :50.00000   Max.   :999.000   Max.   :375.000
   cropcd19          cropcd20           cropcd21
Min.   :  0.000   Min.   : 0.000000   Min.   : 0.000000
1st Qu.:  0.000   1st Qu.: 0.000000   1st Qu.: 0.000000
Median :  0.000   Median : 0.000000   Median : 0.000000
Mean   :  0.458   Mean   : 0.005363   Mean   : 0.009118
3rd Qu.:  0.000   3rd Qu.: 0.000000   3rd Qu.: 0.000000
Max.   :990.000   Max.   :20.000000   Max.   :10.000000
   cropcd22          cropcd23        cropcd24          cropcd25
Min.   :   0.00   Min.   :  0.0   Min.   : 0.0000   Min.   :  0.0000
1st Qu.:   1.00   1st Qu.:  0.0   1st Qu.: 0.0000   1st Qu.:  0.0000
Median :   4.00   Median :  0.0   Median : 0.0000   Median :  0.0000
Mean   :  21.69   Mean   :  1.2   Mean   : 0.7434   Mean   :  0.1372
3rd Qu.:  11.00   3rd Qu.:  0.0   3rd Qu.: 0.0000   3rd Qu.:  0.0000
Max.   :1029.00   Max.   :180.0   Max.   :40.0000   Max.   :102.0000
   cropcd26           cropcd27          cropcd28
Min.   : 0.000000   Min.   : 0.00000   Min.   :  0.0000
1st Qu.: 0.000000   1st Qu.: 0.00000   1st Qu.:  0.0000
Median : 0.000000   Median : 0.00000   Median :  0.0000
Mean   : 0.007375   Mean   : 0.01555   Mean   :  0.9546
3rd Qu.: 0.000000   3rd Qu.: 0.00000   3rd Qu.:  0.0000
Max.   :13.000000   Max.   :35.00000   Max.   :200.0000
   cropcd29           cropcd31          cropcd32
Min.   :  0.0000   Min.   :0.000000   Min.   :  0.0000
1st Qu.:  0.0000   1st Qu.:0.000000   1st Qu.:  0.0000
Median :  0.0000   Median :0.000000   Median :  0.0000
Mean   :  0.2605   Mean   :0.001073   Mean   :  0.4868
3rd Qu.:  0.0000   3rd Qu.:0.000000   3rd Qu.:  0.0000
Max.   :343.7500   Max.   :4.000000   Max.   :700.0000
   cropcd33           cropcd34          cropcd35
Min.   : 0.00000   Min.   :  0.00000   Min.   : 0.00000
1st Qu.: 0.00000   1st Qu.:  0.00000   1st Qu.: 0.00000
Median : 0.00000   Median :  0.00000   Median : 0.00000
Mean   : 0.03252   Mean   :  0.06704   Mean   : 0.04545
3rd Qu.: 0.00000   3rd Qu.:  0.00000   3rd Qu.: 0.00000
Max.   :20.00000   Max.   :250.00000   Max.   :50.00000
```

Hide

```
# spread household havested root count and count havested root type
hh_root_info <- sec8c2 %>%
  select(clust, nh, rootcd, s8cq21a) %>%
  replace(., . > 1e+100, 0) %>%
  group_by(clust, nh, rootcd) %>%
  # sum up quantity of same root type
  summarise(s8cq21ac = sum(s8cq21a)) %>%
  group_by(clust, nh) %>%
  mutate(rootTypeCount = n()) %>%
  spread(key = rootcd,
         value = s8cq21ac,
         fill = 0,
         sep = "")
summary(hh_root_info)
```

```
     clust            nh        rootTypeCount       rootcd0      
 Min.   :4002   Min.   : 1.00   Min.   : 1.000   Min.   :0  
 1st Qu.:4395   1st Qu.: 6.00   1st Qu.: 3.000   1st Qu.:0  
 Median :4572   Median :11.00   Median : 6.000   Median :0  
 Mean   :4562   Mean   :10.86   Mean   : 6.511   Mean   :0  
 3rd Qu.:4819   3rd Qu.:16.00   3rd Qu.: 9.000   3rd Qu.:0  
 Max.   :4999   Max.   :25.00   Max.   :18.000   Max.   :0  
    rootcd5           rootcd6          rootcd7           rootcd8       
 Min.   :  0.000   Min.   :  0.0   Min.   : 0.0000   Min.   :  0.000  
 1st Qu.:  0.000   1st Qu.:  0.0   1st Qu.: 0.0000   1st Qu.:  0.000  
 Median :  0.000   Median :  0.0   Median : 0.0000   Median :  0.000  
 Mean   :  1.747   Mean   :  2.8   Mean   : 0.3084   Mean   :  3.587  
 3rd Qu.:  0.000   3rd Qu.:  2.0   3rd Qu.: 0.0000   3rd Qu.:  0.000  
 Max.   :225.000   Max.   :143.0   Max.   :80.0000   Max.   :550.000  
    rootcd9            rootcd11           rootcd14    rootcd16      
 Min.   :  0.0000   Min.   :  0.0000   Min.   :0   Min.   :  0.0000  
 1st Qu.:  0.0000   1st Qu.:  0.0000   1st Qu.:0   1st Qu.:  0.0000  
 Median :  0.0000   Median :  0.0000   Median :0   Median :  0.0000  
 Mean   :  0.1723   Mean   :  0.6951   Mean   :0   Mean   :  0.4183  
 3rd Qu.:  0.0000   3rd Qu.:  0.0000   3rd Qu.:0   3rd Qu.:  0.0000  
 Max.   :210.0000   Max.   :700.0000   Max.   :0   Max.   :280.0000  
    rootcd18          rootcd19          rootcd20          rootcd21       
 Min.   : 0.000   Min.   :  0.000   Min.   : 0.0000   Min.   :  0.0000  
 1st Qu.: 0.000   1st Qu.:  0.000   1st Qu.: 0.0000   1st Qu.:  0.0000  
 Median : 1.000   Median :  0.000   Median : 0.0000   Median :  0.0000  
 Mean   : 2.565   Mean   :  3.729   Mean   : 0.4482   Mean   :  0.2046  
 3rd Qu.: 3.000   3rd Qu.:  0.000   3rd Qu.: 0.2500   3rd Qu.:  0.0000  
 Max.   :83.000   Max.   :800.000   Max.   :30.0000   Max.   :100.0000  
    rootcd22           rootcd25           rootcd26          rootcd27       
 Min.   :  0.00000   Min.   :  0.0000   Min.   :  0.000   Min.   : 0.0000  
```

```
 1st Qu.:  0.00000   1st Qu.:  0.0000   1st Qu.:  0.000   1st Qu.: 0.0000
 Median :  0.00000   Median :  0.0000   Median :  0.000   Median : 0.0000
 Mean   :  0.03964   Mean   :  0.7617   Mean   :  1.626   Mean   : 0.6471
 3rd Qu.:  0.00000   3rd Qu.:  0.0000   3rd Qu.:  0.000   3rd Qu.: 0.0000
 Max.   :100.00000   Max.   :120.0000   Max.   :375.000   Max.   :80.0000
    rootcd29           rootcd30           rootcd31           rootcd33
 Min.   :  0.000   Min.   :  0.000   Min.   : 0.0000   Min.   : 0.0000
 1st Qu.:  0.000   1st Qu.:  0.000   1st Qu.: 0.0000   1st Qu.: 0.0000
 Median :  0.000   Median :  0.000   Median : 0.0000   Median : 0.0000
 Mean   :  1.897   Mean   :  2.534   Mean   : 0.4419   Mean   : 0.1937
 3rd Qu.:  2.000   3rd Qu.:  3.000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
 Max.   :108.000   Max.   :120.000   Max.   :21.0000   Max.   :63.0000
    rootcd34           rootcd35           rootcd36
 Min.   :  0.000   Min.   :  0.0000   Min.   : 0.0000
 1st Qu.:  0.000   1st Qu.:  0.0000   1st Qu.: 0.0000
 Median :  0.000   Median :  0.0000   Median : 0.0000
 Mean   :  3.025   Mean   :  0.5926   Mean   : 0.9104
 3rd Qu.:  0.000   3rd Qu.:  0.0000   3rd Qu.: 0.0000
 Max.   :550.000   Max.   :150.0000   Max.   :50.0000
```

Do a full join to combine all 4 trunk of data together. Since not every household have entries in all 4 type of agricultural characteristics, we assume the missing value simply means they just don't have that type and replace it with 0.

Hide

```
hh_agri_info <- hh_livestock_info %>%
  full_join(hh_equip_info, by=c("clust", "nh")) %>%
  full_join(hh_crop_info, by=c("clust", "nh")) %>%
  full_join(hh_root_info, by=c("clust", "nh")) %>%
  replace(., is.na(.), 0)
```

Tidy community information

We chose market and transport information from the community data and convert them to logical value.

Hide

```
hh_comm_info <- cs2 %>%
  select(region, district, eanum, s2q19, s2q23) %>%
  mutate(market = ifelse(s2q19 == 1, TRUE, FALSE),
         transport = ifelse(s2q23 == 1, TRUE, FALSE)) %>%
  select(-s2q19, -s2q23)
```

Combine all information

Combine household basic info, household member info (using head as representitive), community info and household agricultural characteristics info together by inner join, because if a household does not have all 4 info, we don't want to use its data in our model fitting. From the combined data, we also created two subset for rural and urban areas. We are interested to see if the two ares have different influencing factors on profit.

Hide

```
hh_all_info <- hh_basic_info %>%
  inner_join(hh_head_info, by=c("clust", "nh")) %>%
  inner_join(hh_comm_info, by=c("region", "district", "eanum")) %>%
  inner_join(hh_agri_info, by=c("clust", "nh"))
hh_profit <- hh_profit_info %>%
  inner_join(hh_all_info, by=c("clust", "nh")) %>%
  select(-region, -district, -eanum, -clust, -nh)
# filter out rural and urban data from hh_profit
hh_profit_rural <- hh_profit %>%
  filter(loc2 == "Rural")
# remove factor type column with only one value
hh_profit_rural <- Filter(function(x) !isSingleValueFactorColumn(x), hh_profit_rural)
hh_profit_urban <- hh_profit %>%
  filter(loc2 == "Urban")
# remove factor type column with only one value
hh_profit_urban <- Filter(function(x) !isSingleValueFactorColumn(x), hh_profit_urban)
```

Fit model and regression diagnostics

Now we fit linear regression models with different variables and do regression diagnostics.

Unrestricted model

First we fit an unrestricted model. Looking at its summary, there are some significant variables. Looking at null hypothesis test result, there are correlated variables in this model. So we should fit a restriced model. Following the model summary is the histogram of standardized residuals and plot of residuals vs fitted values.

Hide

```
hh_profit_model_ur <- lm(profit ~ .,
                         data = hh_profit)
checkCorrVarAndTestHnull(hh_profit_model_ur)
```

```
====================== model summary ======================

Call:
lm(formula = profit ~ ., data = hh_profit)

Residuals:
```

```
     Min       1Q   Median       3Q      Max 
-4146784  -238140  -117148    47055 20645824 

Coefficients: (6 not defined because of singularities)
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)             2.397e+05  2.756e+05   0.870 0.384553    
reslanAkan             -3.865e+04  6.138e+04  -0.630 0.528985    
reslanEwe              -1.034e+05  8.011e+04  -1.290 0.197057    
reslanGaAdangbe         2.459e+05  1.149e+05   2.139 0.032504 *  
reslanDagbani          -1.891e+05  1.084e+05  -1.745 0.081089 .  
reslanHausa            -1.629e+05  1.280e+05  -1.273 0.203271    
reslanOther            -2.358e+05  8.377e+04  -2.815 0.004908 ** 
reslanUnknown          -1.120e+05  7.880e+04  -1.421 0.155413    
ezForest                2.186e+05  2.824e+05   0.774 0.438984    
ezSavannah              1.021e+05  5.465e+04   1.867 0.061931 .  
loc2Rural               7.353e+04  2.560e+05   0.287 0.773985    
loc5RuralCoastal               NA         NA      NA       NA    
loc5RuralForest        -1.728e+05  2.799e+05  -0.617 0.537106    
loc5RuralSavannah              NA         NA      NA       NA    
loc3Rural                      NA         NA      NA       NA    
femaleTRUE              1.523e+04  3.363e+04   0.453 0.650717    
age                     3.905e+03  1.600e+03   2.440 0.014744 *  
avgAge                 -4.050e+02  3.579e+03  -0.113 0.909914    
maxAge                 -3.166e+03  1.842e+03  -1.718 0.085804 .  
minAge                 -2.154e+03  2.423e+03  -0.889 0.374255    
educBasicEducation     -3.320e+04  3.289e+04  -1.009 0.312885    
educSecondaryEducation  7.350e+04  8.948e+04   0.821 0.411469    
educTertiaryEducation  -8.837e+04  7.737e+04  -1.142 0.253494    
educOther              -1.876e+05  1.750e+05  -1.072 0.283873    
marketTRUE             -9.510e+04  4.119e+04  -2.309 0.021017 *  
transportTRUE          -2.059e+04  3.026e+04  -0.681 0.496191    
livstcdTypeCount       -6.877e+03  1.269e+04  -0.542 0.587965    
livstcd1               -1.155e+04  3.835e+04  -0.301 0.763335    
livstcd2               -2.666e+03  5.866e+03  -0.454 0.649505    
livstcd3               -2.455e-01  5.336e-01  -0.460 0.645512    
livstcd4               -1.318e-01  2.959e-01  -0.445 0.656044    
livstcd5                4.406e+04  6.197e+03   7.110 1.40e-12 ***
livstcd6                1.450e+05  4.777e+04   3.037 0.002411 ** 
livstcd7               -1.248e+01  6.172e+00  -2.022 0.043273 *  
livstcd8               -3.450e+03  2.906e+03  -1.187 0.235234    
livstcd9                1.871e+04  1.298e+04   1.442 0.149521    
livstcd10              -8.227e+03  3.768e+03  -2.184 0.029059 *  
livstcd11              -1.127e+05  2.187e+05  -0.515 0.606245    
livstcd12              -3.972e+02  1.658e+04  -0.024 0.980891    
equipTypeCount         -6.038e+04  2.862e+04  -2.109 0.034989 *  
eqcdown21               1.021e+05  3.091e+05   0.330 0.741238    
eqcdown22              -6.542e+04  1.325e+05  -0.494 0.621535    
```

```
eqcdown31           -5.400e+04  2.466e+05  -0.219 0.826694
eqcdown51           -7.935e+02  4.891e+03  -0.162 0.871115
eqcdown61           -2.109e+06  2.767e+05  -7.623 3.18e-14 ***
eqcdown62            3.385e+03  6.850e+03   0.494 0.621172
eqcdown63           -7.457e+02  5.806e+03  -0.128 0.897819
eqcdown64            7.207e+04  9.454e+04   0.762 0.445890
eqcdown65           -3.892e+02  9.763e+02  -0.399 0.690164
cropTypeCount       -1.186e+04  1.683e+04  -0.705 0.480902
cropcd0                     NA         NA      NA       NA
cropcd1             -1.932e+02  4.620e+02  -0.418 0.675833
cropcd2             -2.091e+03  3.470e+03  -0.603 0.546816
cropcd3             -1.497e+04  2.621e+05  -0.057 0.954461
cropcd4              7.705e+01  1.996e+02   0.386 0.699475
cropcd5              4.081e+05  2.472e+05   1.651 0.098897 .
cropcd6             -1.580e+05  2.705e+05  -0.584 0.559116
cropcd8             -1.482e+03  4.839e+02  -3.062 0.002216 **
cropcd9             -8.346e+03  5.246e+04  -0.159 0.873590
cropcd10             1.783e+01  6.821e+02   0.026 0.979143
cropcd11            -8.952e+04  1.756e+04  -5.099 3.60e-07 ***
cropcd12             3.008e+03  1.290e+04   0.233 0.815664
cropcd13            -1.600e+04  4.106e+04  -0.390 0.696752
cropcd14            -3.387e+02  1.705e+03  -0.199 0.842497
cropcd15             1.995e+03  4.547e+03   0.439 0.660795
cropcd16             1.486e+04  1.571e+04   0.946 0.344249
cropcd17             2.534e+02  4.916e+02   0.516 0.606229
cropcd18             5.804e+02  1.810e+03   0.321 0.748524
cropcd19            -2.968e+02  2.927e+03  -0.101 0.919234
cropcd20             1.824e+04  3.999e+04   0.456 0.648278
cropcd21             3.078e+04  7.429e+04   0.414 0.678686
cropcd22            -4.025e+01  1.668e+02  -0.241 0.809310
cropcd23             2.136e+03  2.485e+03   0.859 0.390242
cropcd24            -4.521e+03  6.434e+03  -0.703 0.482287
cropcd25             9.963e+04  8.311e+03  11.988  < 2e-16 ***
cropcd26             1.108e+04  5.549e+04   0.200 0.841796
cropcd27            -5.238e+03  2.079e+04  -0.252 0.801116
cropcd28             1.381e+03  2.157e+03   0.640 0.521981
cropcd29            -6.901e+03  3.997e+03  -1.727 0.084298 .
cropcd31             1.468e+05  1.959e+05   0.749 0.453815
cropcd32            -1.836e+02  7.815e+02  -0.235 0.814233
cropcd33            -1.174e+04  2.763e+04  -0.425 0.670842
cropcd34             1.720e+03  3.145e+03   0.547 0.584485
cropcd35            -3.955e+02  1.244e+04  -0.032 0.974643
rootTypeCount        2.498e+03  4.209e+03   0.593 0.552936
rootcd0                     NA         NA      NA       NA
rootcd5              9.402e+02  1.310e+03   0.718 0.473066
rootcd6              2.582e+03  1.718e+03   1.503 0.133013
rootcd7              2.045e+04  7.834e+03   2.610 0.009097 **
```

```
rootcd8                  5.665e+02  6.516e+02   0.869 0.384716
rootcd9                 -1.315e+03  3.763e+03  -0.350 0.726726
rootcd11                -9.676e+01  1.366e+03  -0.071 0.943541
rootcd14                        NA         NA      NA       NA
rootcd16                 1.392e+03  2.704e+03   0.515 0.606768
rootcd18                 2.340e+04  3.453e+03   6.779 1.41e-11 ***
rootcd19                -3.822e+01  4.430e+02  -0.086 0.931262
rootcd20                 4.535e+04  1.056e+04   4.294 1.81e-05 ***
rootcd21                -1.873e+02  4.863e+03  -0.039 0.969273
rootcd22                -1.023e+03  7.899e+03  -0.129 0.896983
rootcd25                -1.450e+03  4.088e+03  -0.355 0.722916
rootcd26                -1.152e+03  1.450e+03  -0.794 0.427080
rootcd27                -1.669e+04  3.276e+03  -5.095 3.68e-07 ***
rootcd29                 6.320e+02  3.090e+03   0.205 0.837959
rootcd30                 1.534e+02  2.828e+03   0.054 0.956734
rootcd31                 1.078e+03  1.041e+04   0.103 0.917595
rootcd33                 1.408e+04  8.392e+03   1.677 0.093536 .
rootcd34                 1.647e+02  8.711e+02   0.189 0.850032
rootcd35                -3.864e+03  2.366e+03  -1.633 0.102498
rootcd36                 1.993e+04  5.531e+03   3.603 0.000319 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 782200 on 3510 degrees of freedom
Multiple R-squared:  0.137, Adjusted R-squared:  0.1119
F-statistic: 5.461 on 102 and 3510 DF,  p-value: < 2.2e-16

====================== model alias ========================
Model :
profit ~ reslan + ez + loc2 + loc5 + loc3 + female + age + avgAge +
    maxAge + minAge + educ + market + transport + livstcdTypeCount +
    livstcd1 + livstcd2 + livstcd3 + livstcd4 + livstcd5 + livstcd6 +
    livstcd7 + livstcd8 + livstcd9 + livstcd10 + livstcd11 +
    livstcd12 + equipTypeCount + eqcdown21 + eqcdown22 + eqcdown31 +
    eqcdown51 + eqcdown61 + eqcdown62 + eqcdown63 + eqcdown64 +
    eqcdown65 + cropTypeCount + cropcd0 + cropcd1 + cropcd2 +
    cropcd3 + cropcd4 + cropcd5 + cropcd6 + cropcd8 + cropcd9 +
    cropcd10 + cropcd11 + cropcd12 + cropcd13 + cropcd14 + cropcd15 +
    cropcd16 + cropcd17 + cropcd18 + cropcd19 + cropcd20 + cropcd21 +
    cropcd22 + cropcd23 + cropcd24 + cropcd25 + cropcd26 + cropcd27 +
    cropcd28 + cropcd29 + cropcd31 + cropcd32 + cropcd33 + cropcd34 +
    cropcd35 + rootTypeCount + rootcd0 + rootcd5 + rootcd6 +
    rootcd7 + rootcd8 + rootcd9 + rootcd11 + rootcd14 + rootcd16 +
    rootcd18 + rootcd19 + rootcd20 + rootcd21 + rootcd22 + rootcd25 +
    rootcd26 + rootcd27 + rootcd29 + rootcd30 + rootcd31 + rootcd33 +
    rootcd34 + rootcd35 + rootcd36
```

```
Complete :
                 (Intercept) reslanAkan reslanEwe reslanGaAdangbe
loc5RuralCoastal   1           0          0         0
loc5RuralSavannah -1           0          0         0
loc3Rural          0           0          0         0
cropcd0            0           0          0         0
rootcd0            0           0          0         0
rootcd14           0           0          0         0
                 reslanDagbani reslanHausa reslanOther reslanUnknown
loc5RuralCoastal   0             0           0           0
loc5RuralSavannah  0             0           0           0
loc3Rural          0             0           0           0
cropcd0            0             0           0           0
rootcd0            0             0           0           0
rootcd14           0             0           0           0
                 ezForest ezSavannah loc2Rural loc5RuralForest femaleTRUE
loc5RuralCoastal  -1       -1          0         0               0
loc5RuralSavannah  1        1          1        -1               0
loc3Rural          0        0          1         0               0
cropcd0            0        0          0         0               0
rootcd0            0        0          0         0               0
rootcd14           0        0          0         0               0
                 age avgAge maxAge minAge educBasicEducation
loc5RuralCoastal   0   0      0      0      0
loc5RuralSavannah  0   0      0      0      0
loc3Rural          0   0      0      0      0
cropcd0            0   0      0      0      0
rootcd0            0   0      0      0      0
rootcd14           0   0      0      0      0
                 educSecondaryEducation educTertiaryEducation educOther
loc5RuralCoastal   0                      0                     0
loc5RuralSavannah  0                      0                     0
loc3Rural          0                      0                     0
cropcd0            0                      0                     0
rootcd0            0                      0                     0
rootcd14           0                      0                     0
                 marketTRUE transportTRUE livstcdTypeCount livstcd1
loc5RuralCoastal   0          0             0                 0
loc5RuralSavannah  0          0             0                 0
loc3Rural          0          0             0                 0
cropcd0            0          0             0                 0
rootcd0            0          0             0                 0
rootcd14           0          0             0                 0
                 livstcd2 livstcd3 livstcd4 livstcd5 livstcd6 livstcd7
loc5RuralCoastal   0        0        0        0        0        0
loc5RuralSavannah  0        0        0        0        0        0
loc3Rural          0        0        0        0        0        0
```

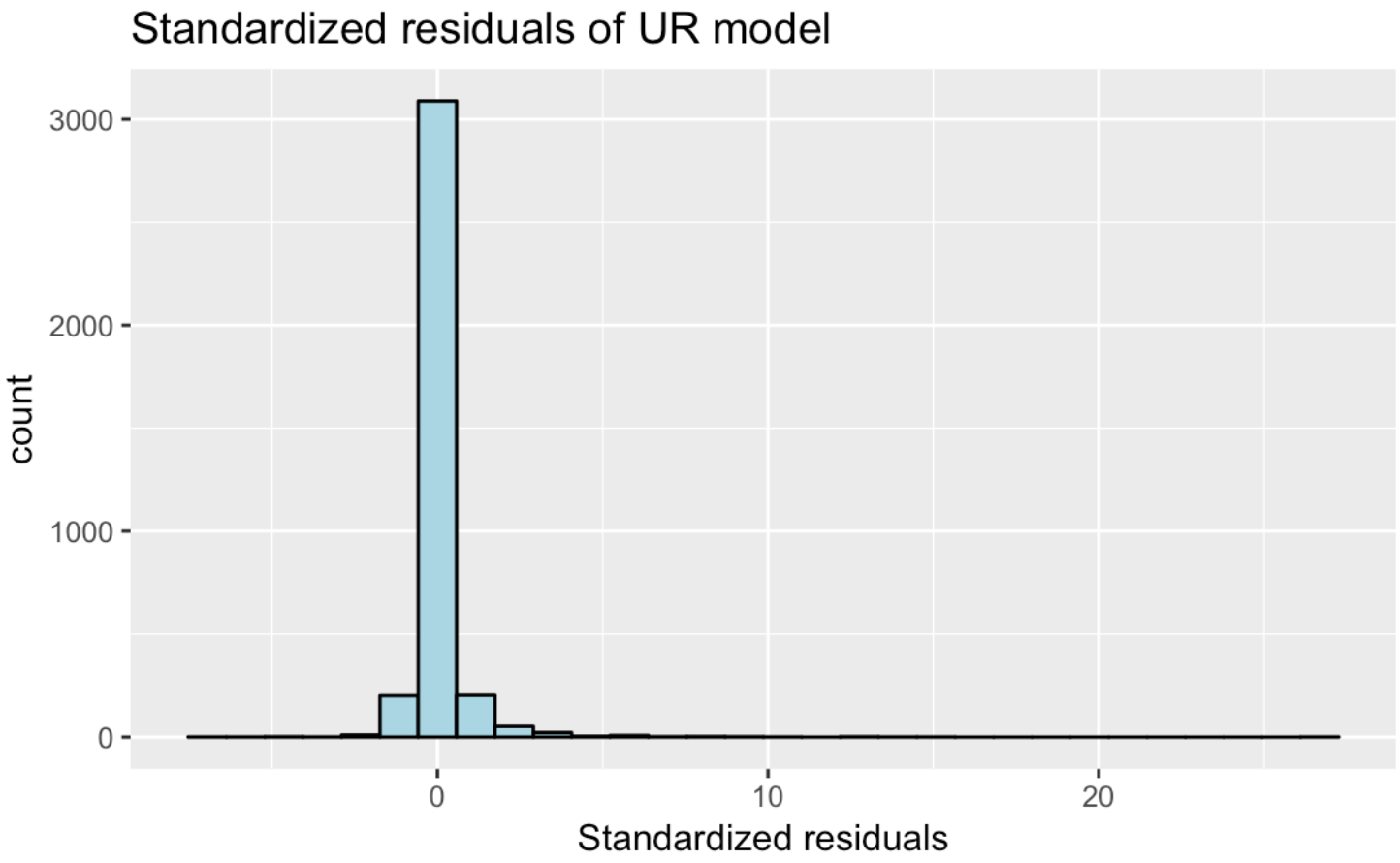
```
cropcd0             0         0         0         0         0         0
rootcd0             0         0         0         0         0         0
rootcd14            0         0         0         0         0         0
                  livstcd8 livstcd9 livstcd10 livstcd11 livstcd12
loc5RuralCoastal    0         0         0          0          0
loc5RuralSavannah   0         0         0          0          0
loc3Rural           0         0         0          0          0
cropcd0             0         0         0          0          0
rootcd0             0         0         0          0          0
rootcd14            0         0         0          0          0
                  equipTypeCount eqcdown21 eqcdown22 eqcdown31 eqcdown51
loc5RuralCoastal    0               0          0          0          0
loc5RuralSavannah   0               0          0          0          0
loc3Rural           0               0          0          0          0
cropcd0             0               0          0          0          0
rootcd0             0               0          0          0          0
rootcd14            0               0          0          0          0
                  eqcdown61 eqcdown62 eqcdown63 eqcdown64 eqcdown65
loc5RuralCoastal    0          0          0          0          0
loc5RuralSavannah   0          0          0          0          0
loc3Rural           0          0          0          0          0
cropcd0             0          0          0          0          0
rootcd0             0          0          0          0          0
rootcd14            0          0          0          0          0
                  cropTypeCount cropcd1 cropcd2 cropcd3 cropcd4 cropcd5
loc5RuralCoastal    0              0        0        0        0        0
loc5RuralSavannah   0              0        0        0        0        0
loc3Rural           0              0        0        0        0        0
cropcd0             0              0        0        0        0        0
rootcd0             0              0        0        0        0        0
rootcd14            0              0        0        0        0        0
                  cropcd6 cropcd8 cropcd9 cropcd10 cropcd11 cropcd12
loc5RuralCoastal    0        0        0        0         0         0
loc5RuralSavannah   0        0        0        0         0         0
loc3Rural           0        0        0        0         0         0
cropcd0             0        0        0        0         0         0
rootcd0             0        0        0        0         0         0
rootcd14            0        0        0        0         0         0
                  cropcd13 cropcd14 cropcd15 cropcd16 cropcd17 cropcd18
loc5RuralCoastal    0         0         0         0         0         0
loc5RuralSavannah   0         0         0         0         0         0
loc3Rural           0         0         0         0         0         0
cropcd0             0         0         0         0         0         0
rootcd0             0         0         0         0         0         0
rootcd14            0         0         0         0         0         0
                  cropcd19 cropcd20 cropcd21 cropcd22 cropcd23 cropcd24
loc5RuralCoastal    0         0         0         0         0         0
```

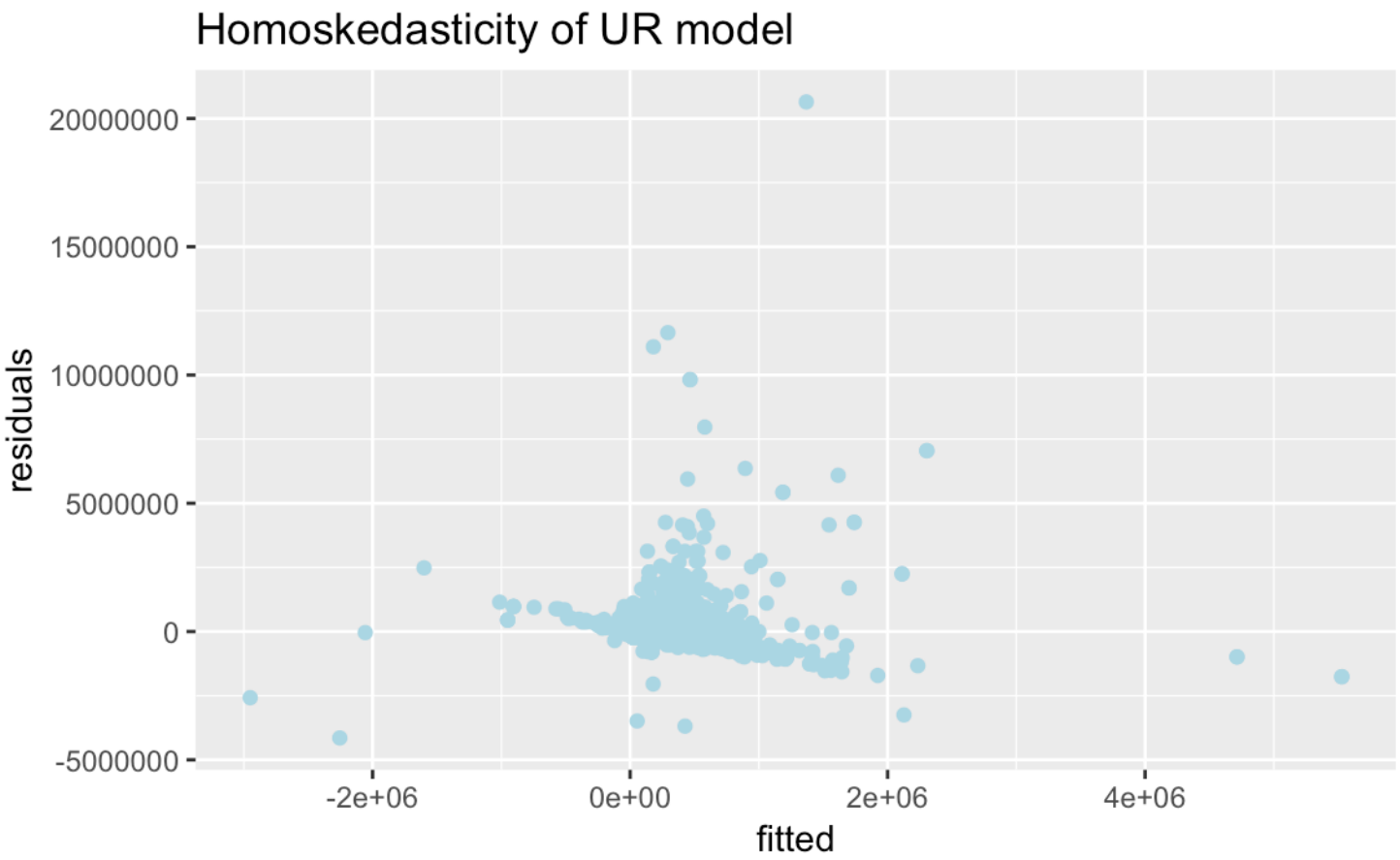
```
loc5RuralSavannah  0         0         0         0         0         0
loc3Rural          0         0         0         0         0         0
cropcd0            0         0         0         0         0         0
rootcd0            0         0         0         0         0         0
rootcd14           0         0         0         0         0         0
                  cropcd25 cropcd26 cropcd27 cropcd28 cropcd29 cropcd31
loc5RuralCoastal   0         0         0         0         0         0
loc5RuralSavannah  0         0         0         0         0         0
loc3Rural          0         0         0         0         0         0
cropcd0            0         0         0         0         0         0
rootcd0            0         0         0         0         0         0
rootcd14           0         0         0         0         0         0
                  cropcd32 cropcd33 cropcd34 cropcd35 rootTypeCount rootcd5
loc5RuralCoastal   0         0         0         0         0              0
loc5RuralSavannah  0         0         0         0         0              0
loc3Rural          0         0         0         0         0              0
cropcd0            0         0         0         0         0              0
rootcd0            0         0         0         0         0              0
rootcd14           0         0         0         0         0              0
                  rootcd6 rootcd7 rootcd8 rootcd9 rootcd11 rootcd16
loc5RuralCoastal   0        0        0        0        0         0
loc5RuralSavannah  0        0        0        0        0         0
loc3Rural          0        0        0        0        0         0
cropcd0            0        0        0        0        0         0
rootcd0            0        0        0        0        0         0
rootcd14           0        0        0        0        0         0
                  rootcd18 rootcd19 rootcd20 rootcd21 rootcd22 rootcd25
loc5RuralCoastal   0         0         0         0         0         0
loc5RuralSavannah  0         0         0         0         0         0
loc3Rural          0         0         0         0         0         0
cropcd0            0         0         0         0         0         0
rootcd0            0         0         0         0         0         0
rootcd14           0         0         0         0         0         0
                  rootcd26 rootcd27 rootcd29 rootcd30 rootcd31 rootcd33
loc5RuralCoastal   0         0         0         0         0         0
loc5RuralSavannah  0         0         0         0         0         0
loc3Rural          0         0         0         0         0         0
cropcd0            0         0         0         0         0         0
rootcd0            0         0         0         0         0         0
rootcd14           0         0         0         0         0         0
                  rootcd34 rootcd35 rootcd36
loc5RuralCoastal   0         0         0
loc5RuralSavannah  0         0         0
loc3Rural          0         0         0
cropcd0            0         0         0
rootcd0            0         0         0
rootcd14           0         0         0
```

```
There are correlated variables. See above model alias.
```

Hide

```
suppressWarnings(plotResiduals(hh_profit, hh_profit_model_ur, "lightblue", "UR"))
```





Restricted model with significant variables (R1)

We pick the significant variables from the above unrestricted model and fit a restricted model.

Hide

```
hh_profit_model_r1 <- lm(profit ~ reslan + ez + age + market + livstcd5 + livstcd6 +
livstcd7 + livstcd10 +
                           equipTypeCount + eqcdown61 + cropcd5 + cropcd8 + cropcd11
+ cropcd25 + cropcd29 +
                           rootcd7 + rootcd18 + rootcd20 + rootcd27 + rootcd33 + root
cd36,
                         data = hh_profit)
checkCorrVarAndTestHnull(hh_profit_model_r1)
```

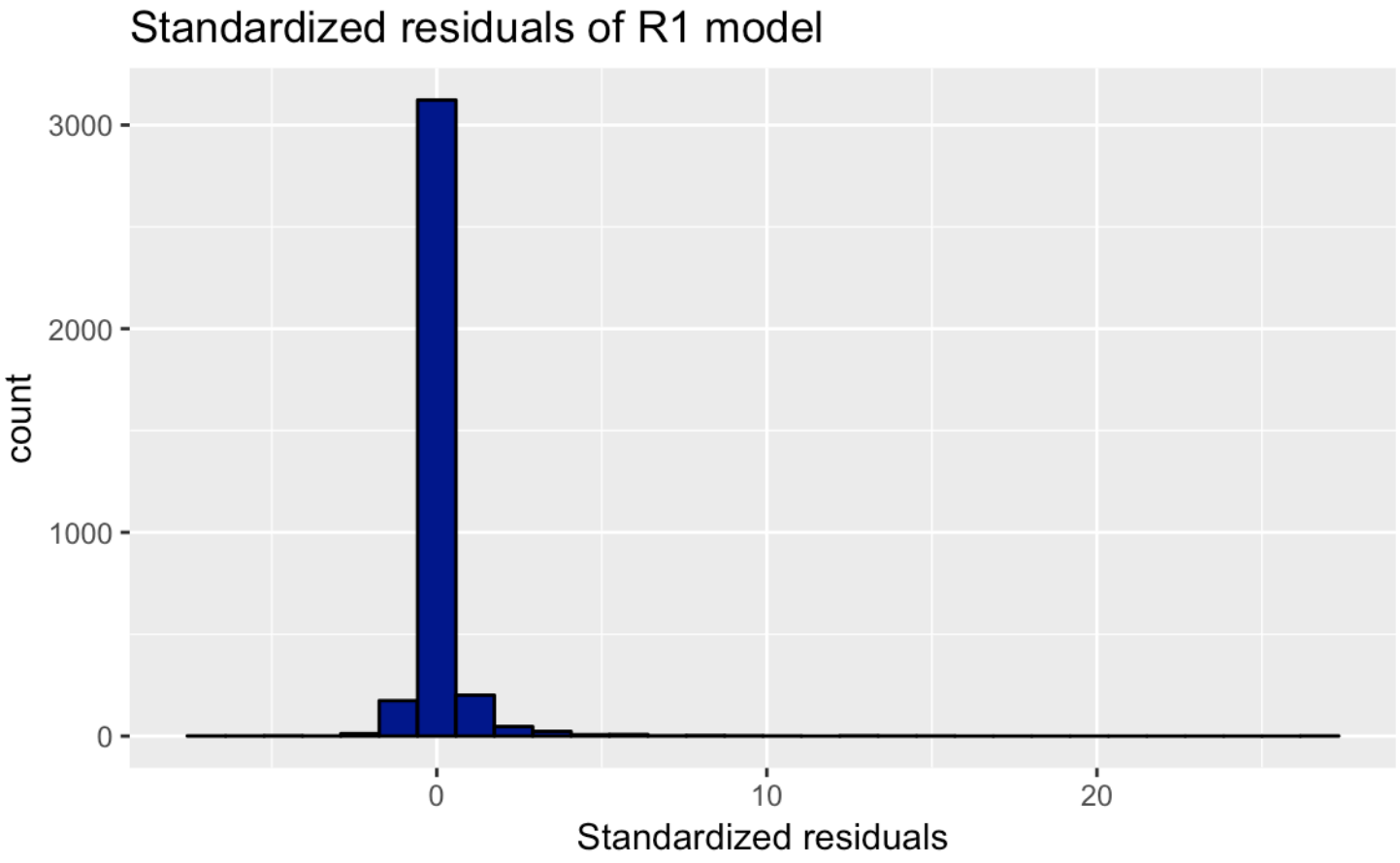
```
====================== model summary ======================

Call:
lm(formula = profit ~ reslan + ez + age + market + livstcd5 +
    livstcd6 + livstcd7 + livstcd10 + equipTypeCount + eqcdown61 +
    cropcd5 + cropcd8 + cropcd11 + cropcd25 + cropcd29 + rootcd7 +
    rootcd18 + rootcd20 + rootcd27 + rootcd33 + rootcd36, data = hh_profit)
```

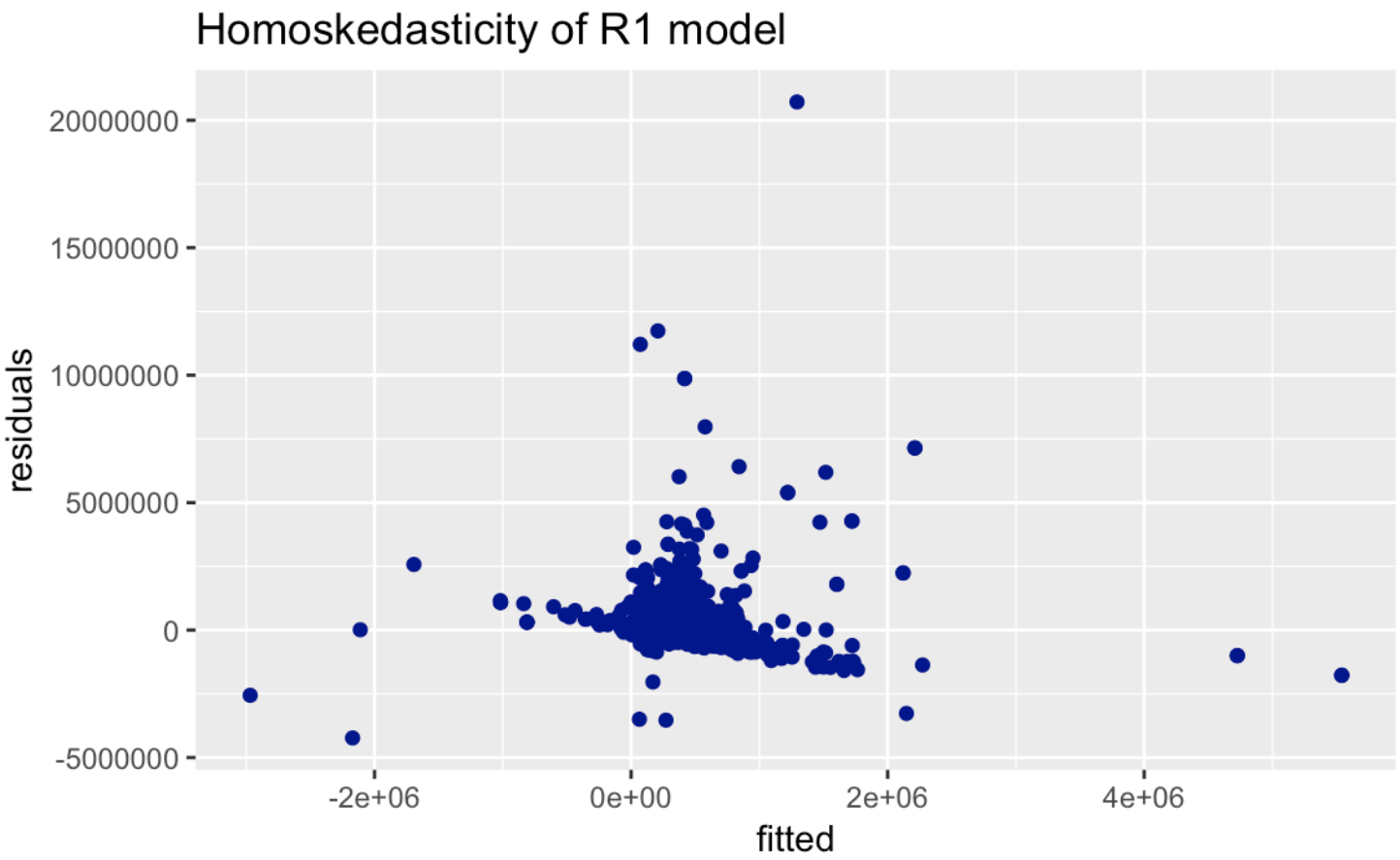
```
Residuals:
     Min       1Q   Median       3Q      Max
-4230445  -234684  -129489    33517 20721029

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.166e+05  7.787e+04   2.781  0.00545 **
reslanAkan       4.694e+03  5.728e+04   0.082  0.93469
reslanEwe       -4.304e+04  7.487e+04  -0.575  0.56544
reslanGaAdangbe  2.707e+05  1.092e+05   2.480  0.01319 *
reslanDagbani   -1.701e+05  1.048e+05  -1.624  0.10448
reslanHausa     -1.750e+05  1.245e+05  -1.405  0.16015
reslanOther     -2.324e+05  8.018e+04  -2.899  0.00377 **
reslanUnknown   -6.183e+04  7.529e+04  -0.821  0.41159
ezForest         4.855e+04  3.988e+04   1.218  0.22347
ezSavannah       6.429e+04  4.706e+04   1.366  0.17194
age              1.676e+02  8.861e+02   0.189  0.84997
marketTRUE      -9.375e+04  3.902e+04  -2.403  0.01632 *
livstcd5         4.293e+04  5.847e+03   7.341 2.60e-13 ***
livstcd6         1.508e+05  4.643e+04   3.247  0.00118 **
livstcd7        -9.917e+00  5.336e+00  -1.859  0.06316 .
livstcd10       -8.339e+03  3.712e+03  -2.246  0.02476 *
equipTypeCount  -5.785e+04  2.463e+04  -2.349  0.01890 *
eqcdown61       -2.123e+06  2.644e+05  -8.031 1.30e-15 ***
cropcd5          3.344e+05  2.348e+05   1.424  0.15441
cropcd8         -1.430e+03  4.700e+02  -3.043  0.00236 **
cropcd11        -9.308e+04  1.711e+04  -5.441 5.65e-08 ***
cropcd25         1.002e+05  8.164e+03  12.278  < 2e-16 ***
cropcd29        -5.640e+03  3.735e+03  -1.510  0.13115
rootcd7          2.022e+04  7.682e+03   2.632  0.00852 **
rootcd18         2.408e+04  3.248e+03   7.413 1.53e-13 ***
rootcd20         4.457e+04  1.023e+04   4.356 1.36e-05 ***
rootcd27        -1.617e+04  3.195e+03  -5.062 4.36e-07 ***
rootcd33         1.436e+04  7.675e+03   1.872  0.06133 .
rootcd36         2.221e+04  5.141e+03   4.320 1.60e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 779000 on 3584 degrees of freedom
Multiple R-squared:  0.1259,    Adjusted R-squared:  0.1191
F-statistic: 18.44 on 28 and 3584 DF,  p-value: < 2.2e-16

==================== hypothesis test ====================
Linear hypothesis test

Hypothesis:
reslanAkan = 0
```

```
reslanEwe = 0
reslanGaAdangbe = 0
reslanDagbani = 0
reslanHausa = 0
reslanOther = 0
reslanUnknown = 0
ezForest = 0
ezSavannah = 0
age = 0
marketTRUE = 0
livstcd5 = 0
livstcd6 = 0
livstcd7 = 0
livstcd10 = 0
equipTypeCount = 0
eqcdown61 = 0
cropcd5 = 0
cropcd8 = 0
cropcd11 = 0
cropcd25 = 0
cropcd29 = 0
rootcd7 = 0
rootcd18 = 0
rootcd20 = 0
rootcd27 = 0
rootcd33 = 0
rootcd36 = 0

Model 1: restricted model
Model 2: profit ~ reslan + ez + age + market + livstcd5 + livstcd6 + livstcd7 +
    livstcd10 + equipTypeCount + eqcdown61 + cropcd5 + cropcd8 +
    cropcd11 + cropcd25 + cropcd29 + rootcd7 + rootcd18 + rootcd20 +
    rootcd27 + rootcd33 + rootcd36

  Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
1   3612 2.4883e+15
2   3584 2.1750e+15 28 3.1327e+14 18.436 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
suppressWarnings(plotResiduals(hh_profit, hh_profit_model_r1, "darkblue", "R1"))
```


Standardized residuals of R1 model
3000
2000
1000
0
count
0
10
20
Standardized residuals



Restricted model with top features from agricultural characteristics information (TOP)

We are also interested to see how will the model perform if fitted only with top 15 features from agricultural characteristics information. These top features are the variables that have higher correlation with profit, no matter it's positive or negative correlation.

Hide

```
hh_profit_agri <- hh_profit[, -c(2:14)] # get agricultural characteristics variables
all_correlations <- findAbsoluteCorrelation(hh_profit_agri)
```

```
the standard deviation is zerothe standard deviation is zerothe standard deviation is
zero
```

Hide

```
hh_profit_agri_topFeatures <- hh_profit[c("profit", all_correlations$colName[1:15])]
hh_profit_model_topfeatures <- lm(profit ~ . , data = hh_profit_agri_topFeatures)
checkCorrVarAndTestHnull(hh_profit_model_topfeatures)
```

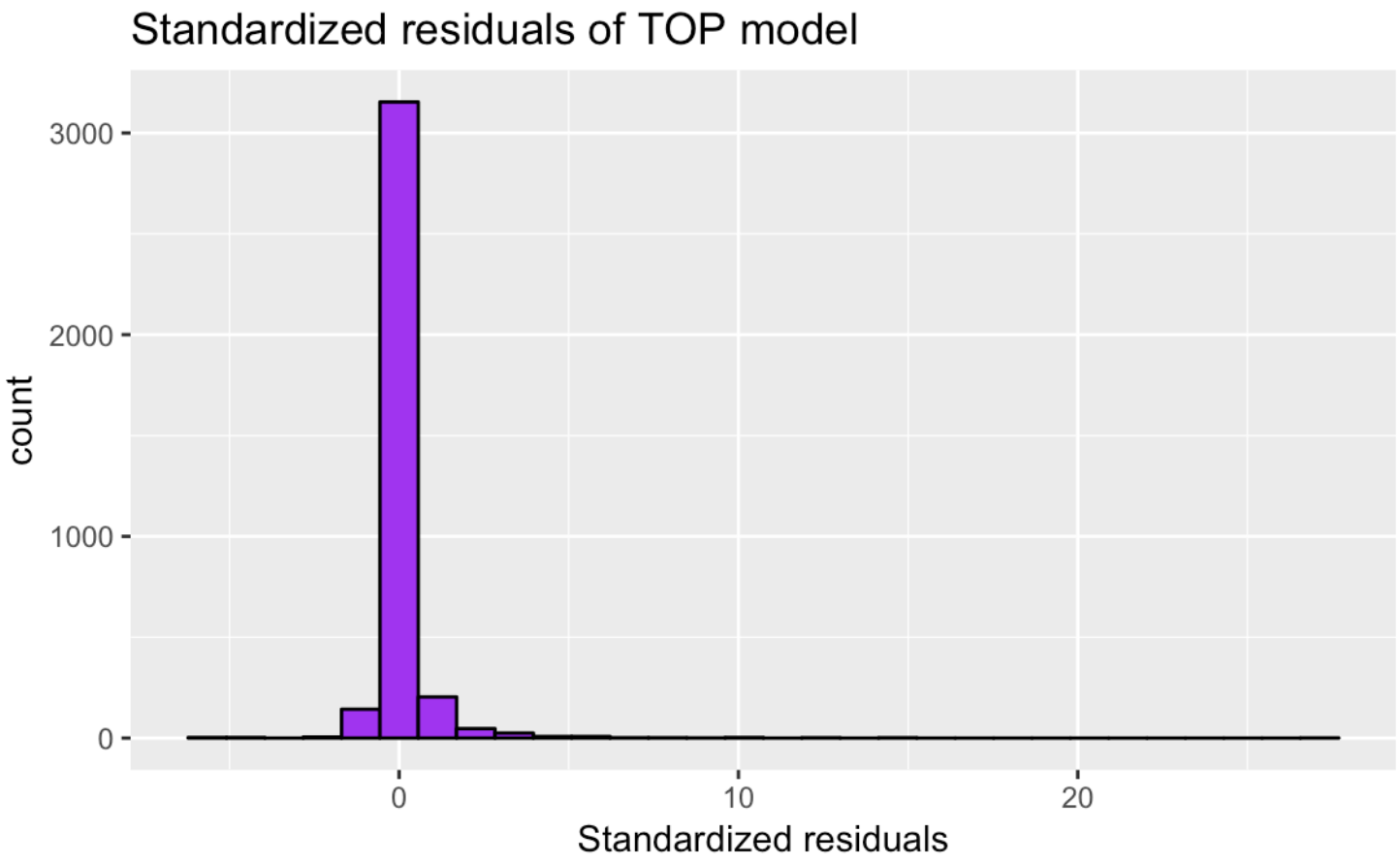
```
==================== model summary ======================
```

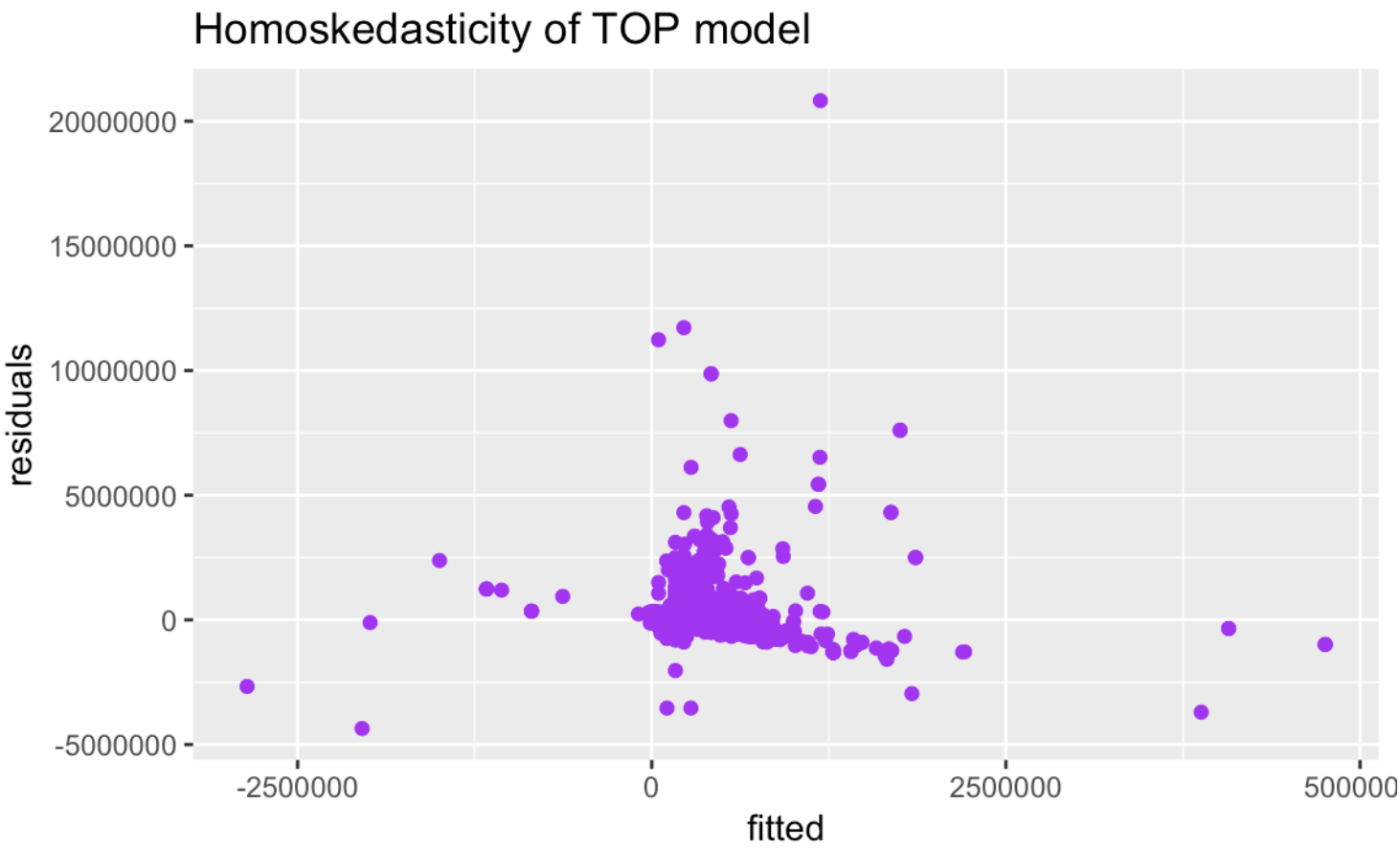
```
Call:
lm(formula = profit ~ ., data = hh_profit_agri_topFeatures)

Residuals:
     Min       1Q   Median       3Q      Max 
-4356876  -223572  -143826     9718 20824127 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      226865.6    21550.4  10.527  < 2e-16 ***
cropcd25          84128.0     7257.6  11.592  < 2e-16 ***
rootcd18          23465.0     3380.7   6.941 4.60e-12 ***
rootcd20          44255.6    10409.3   4.252 2.18e-05 ***
rootTypeCount      2448.2     3722.7   0.658  0.51082    
eqcdown61      -1969761.3   270480.3  -7.282 4.01e-13 ***
rootcd36          17450.2     5337.4   3.269  0.00109 ** 
livstcd5          32012.9     5377.6   5.953 2.89e-09 ***
equipTypeCount   -59648.8    23358.8  -2.554  0.01070 *  
rootcd6            2489.4     1662.2   1.498  0.13432    
rootcd7           16346.7     7718.2   2.118  0.03425 *  
livstcd10         -8578.5     3735.6  -2.296  0.02171 *  
rootcd30            276.4     2479.5   0.111  0.91124    
rootcd5            1477.8     1296.6   1.140  0.25448    
rootcd8             430.4      600.4   0.717  0.47355    
livstcd2          -6932.1     5149.1  -1.346  0.17830    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 788200 on 3597 degrees of freedom
Multiple R-squared:  0.102, Adjusted R-squared:  0.09828 
F-statistic: 27.25 on 15 and 3597 DF,  p-value: < 2.2e-16

====================== hypothesis test ====================
Linear hypothesis test

Hypothesis:
cropcd25 = 0
rootcd18 = 0
rootcd20 = 0
rootTypeCount = 0
eqcdown61 = 0
rootcd36 = 0
livstcd5 = 0
equipTypeCount = 0
rootcd6 = 0
rootcd7 = 0
```

```
livstcd10 = 0
rootcd30 = 0
rootcd5 = 0
rootcd8 = 0
livstcd2 = 0

Model 1: restricted model
Model 2: profit ~ cropcd25 + rootcd18 + rootcd20 + rootTypeCount + eqcdown61 +
    rootcd36 + livstcd5 + equipTypeCount + rootcd6 + rootcd7 +
    livstcd10 + rootcd30 + rootcd5 + rootcd8 + livstcd2

  Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
1   3612 2.4883e+15
2   3597 2.2344e+15 15 2.5387e+14 27.246 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
suppressWarnings(plotResiduals(hh_profit, hh_profit_model_topfeatures, "purple", "TOP
"))
```





Unrestricted model for rural area (RURAL)

Fit unrestricted model for rural area profit.

Hide

```
hh_profit_model_rural <- lm(profit ~ .,
                            data = hh_profit_rural)
checkCorrVarAndTestHnull(hh_profit_model_rural)
```

```
====================== model summary ======================

Call:
lm(formula = profit ~ ., data = hh_profit_rural)

Residuals:
     Min       1Q   Median       3Q      Max
-4146964  -238546  -118235    46546 20643791

Coefficients: (5 not defined because of singularities)
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             3.165e+05  9.745e+04   3.248 0.001174 **
reslanAkan             -3.980e+04  6.191e+04  -0.643 0.520400
```

```
reslanEwe               -9.813e+04  8.080e+04  -1.214 0.224663
reslanGaAdangbe          2.437e+05  1.157e+05   2.107 0.035192 *
reslanDagbani           -1.920e+05  1.091e+05  -1.761 0.078401 .
reslanHausa             -1.617e+05  1.288e+05  -1.256 0.209224
reslanOther             -2.378e+05  8.432e+04  -2.820 0.004826 **
reslanUnknown           -1.122e+05  7.966e+04  -1.408 0.159184
ezForest                 4.293e+04  4.459e+04   0.963 0.335781
ezSavannah               9.988e+04  5.515e+04   1.811 0.070223 .
loc5RuralForest                 NA         NA      NA       NA
loc5RuralSavannah               NA         NA      NA       NA
femaleTRUE               1.203e+04  3.416e+04   0.352 0.724752
age                      3.770e+03  1.622e+03   2.324 0.020190 *
avgAge                  -2.234e+02  3.628e+03  -0.062 0.950902
maxAge                  -3.088e+03  1.867e+03  -1.654 0.098229 .
minAge                  -2.242e+03  2.456e+03  -0.913 0.361347
educBasicEducation      -3.337e+04  3.331e+04  -1.002 0.316476
educSecondaryEducation   7.066e+04  9.065e+04   0.780 0.435738
educTertiaryEducation   -8.169e+04  7.886e+04  -1.036 0.300339
educOther               -1.880e+05  1.760e+05  -1.068 0.285462
marketTRUE              -1.099e+05  4.196e+04  -2.620 0.008844 **
transportTRUE           -1.863e+04  3.045e+04  -0.612 0.540606
livstcdTypeCount        -7.444e+03  1.281e+04  -0.581 0.561133
livstcd1                -1.234e+04  3.862e+04  -0.320 0.749338
livstcd2                -2.516e+03  5.901e+03  -0.426 0.669861
livstcd3                -2.526e-01  5.366e-01  -0.471 0.637801
livstcd4                -1.312e-01  2.975e-01  -0.441 0.659209
livstcd5                 4.394e+04  6.286e+03   6.989 3.30e-12 ***
livstcd6                 1.448e+05  4.803e+04   3.014 0.002599 **
livstcd7                -1.244e+01  6.207e+00  -2.004 0.045190 *
livstcd8                -3.226e+03  2.936e+03  -1.099 0.271995
livstcd9                 1.853e+04  1.305e+04   1.420 0.155793
livstcd10               -8.260e+03  3.788e+03  -2.180 0.029292 *
livstcd11               -1.117e+05  2.199e+05  -0.508 0.611625
livstcd12               -3.790e+02  1.667e+04  -0.023 0.981864
equipTypeCount          -6.493e+04  2.889e+04  -2.247 0.024675 *
eqcdown21                1.085e+05  3.108e+05   0.349 0.726991
eqcdown22               -6.254e+04  1.333e+05  -0.469 0.638887
eqcdown31               -5.073e+04  2.480e+05  -0.205 0.837927
eqcdown51               -7.668e+02  4.917e+03  -0.156 0.876087
eqcdown61               -2.103e+06  2.783e+05  -7.557 5.27e-14 ***
eqcdown62                3.453e+03  6.887e+03   0.501 0.616111
eqcdown63               -6.506e+02  5.839e+03  -0.111 0.911284
eqcdown64                7.366e+04  9.505e+04   0.775 0.438424
eqcdown65               -3.703e+02  9.817e+02  -0.377 0.706041
cropTypeCount           -1.197e+04  1.696e+04  -0.706 0.480214
cropcd0                         NA         NA      NA       NA
cropcd1                 -1.807e+02  4.647e+02  -0.389 0.697401
```

```
cropcd2               -2.185e+03  3.490e+03  -0.626 0.531245
cropcd3               -1.369e+04  2.636e+05  -0.052 0.958589
cropcd4                8.770e+01  2.011e+02   0.436 0.662844
cropcd5                4.061e+05  2.486e+05   1.634 0.102416
cropcd6               -1.595e+05  2.720e+05  -0.586 0.557587
cropcd8               -1.475e+03  4.866e+02  -3.032 0.002450 **
cropcd9               -8.029e+03  5.274e+04  -0.152 0.879010
cropcd10               1.812e+02  7.170e+02   0.253 0.800458
cropcd11              -8.957e+04  1.765e+04  -5.074 4.10e-07 ***
cropcd12               3.217e+03  1.298e+04   0.248 0.804330
cropcd13              -1.618e+04  4.128e+04  -0.392 0.695136
cropcd14              -3.311e+02  1.714e+03  -0.193 0.846893
cropcd15               1.923e+03  4.573e+03   0.420 0.674180
cropcd16               1.499e+04  1.579e+04   0.950 0.342383
cropcd17               2.698e+02  4.950e+02   0.545 0.585786
cropcd18               6.021e+02  1.820e+03   0.331 0.740859
cropcd19              -3.004e+02  2.943e+03  -0.102 0.918688
cropcd20               1.923e+04  4.020e+04   0.478 0.632395
cropcd21               3.164e+04  7.476e+04   0.423 0.672133
cropcd22              -4.868e+01  1.678e+02  -0.290 0.771793
cropcd23               2.084e+03  2.499e+03   0.834 0.404478
cropcd24              -4.612e+03  6.471e+03  -0.713 0.476056
cropcd25               9.962e+04  8.357e+03  11.921  < 2e-16 ***
cropcd26               1.044e+04  5.579e+04   0.187 0.851625
cropcd27              -4.896e+03  2.091e+04  -0.234 0.814870
cropcd28               1.346e+03  2.169e+03   0.621 0.534910
cropcd29              -6.881e+03  4.019e+03  -1.712 0.086926 .
cropcd31               1.466e+05  1.970e+05   0.744 0.456695
cropcd32              -1.818e+02  7.857e+02  -0.231 0.816998
cropcd33              -1.161e+04  2.779e+04  -0.418 0.675999
cropcd34               1.733e+03  3.162e+03   0.548 0.583810
cropcd35              -3.535e+02  1.251e+04  -0.028 0.977460
rootTypeCount          3.166e+03  4.278e+03   0.740 0.459246
rootcd0                       NA         NA      NA       NA
rootcd5                9.504e+02  1.318e+03   0.721 0.470781
rootcd6                2.612e+03  1.729e+03   1.511 0.130914
rootcd7                2.034e+04  7.901e+03   2.574 0.010096 *
rootcd8                5.491e+02  6.555e+02   0.838 0.402265
rootcd9               -9.208e+03  1.021e+04  -0.902 0.367376
rootcd11              -1.110e+02  1.374e+03  -0.081 0.935580
rootcd14                      NA         NA      NA       NA
rootcd16               1.442e+03  2.719e+03   0.531 0.595771
rootcd18               2.329e+04  3.478e+03   6.695 2.51e-11 ***
rootcd19              -4.070e+01  4.454e+02  -0.091 0.927205
rootcd20               4.520e+04  1.063e+04   4.254 2.16e-05 ***
rootcd21              -2.646e+02  4.892e+03  -0.054 0.956864
rootcd22              -1.128e+03  7.942e+03  -0.142 0.887095
```

```
rootcd25               -1.697e+03  4.113e+03  -0.412 0.680027
rootcd26               -1.180e+03  1.459e+03  -0.809 0.418485
rootcd27               -1.682e+04  3.342e+03  -5.033 5.08e-07 ***
rootcd29                6.127e+02  3.111e+03   0.197 0.843884
rootcd30               -2.119e+02  2.880e+03  -0.074 0.941367
rootcd31                1.178e+03  1.048e+04   0.112 0.910460
rootcd33                1.425e+04  8.667e+03   1.644 0.100187
rootcd34                1.766e+02  8.761e+02   0.202 0.840252
rootcd35               -3.915e+03  2.379e+03  -1.646 0.099954 .
rootcd36                1.968e+04  5.578e+03   3.528 0.000425 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 786400 on 3456 degrees of freedom
Multiple R-squared:  0.1377,    Adjusted R-squared:  0.1127
F-statistic: 5.518 on 100 and 3456 DF,  p-value: < 2.2e-16

====================== model alias ========================
Model :
profit ~ reslan + ez + loc5 + female + age + avgAge + maxAge +
    minAge + educ + market + transport + livstcdTypeCount + livstcd1 +
    livstcd2 + livstcd3 + livstcd4 + livstcd5 + livstcd6 + livstcd7 +
    livstcd8 + livstcd9 + livstcd10 + livstcd11 + livstcd12 +
    equipTypeCount + eqcdown21 + eqcdown22 + eqcdown31 + eqcdown51 +
    eqcdown61 + eqcdown62 + eqcdown63 + eqcdown64 + eqcdown65 +
    cropTypeCount + cropcd0 + cropcd1 + cropcd2 + cropcd3 + cropcd4 +
    cropcd5 + cropcd6 + cropcd8 + cropcd9 + cropcd10 + cropcd11 +
    cropcd12 + cropcd13 + cropcd14 + cropcd15 + cropcd16 + cropcd17 +
    cropcd18 + cropcd19 + cropcd20 + cropcd21 + cropcd22 + cropcd23 +
    cropcd24 + cropcd25 + cropcd26 + cropcd27 + cropcd28 + cropcd29 +
    cropcd31 + cropcd32 + cropcd33 + cropcd34 + cropcd35 + rootTypeCount +
    rootcd0 + rootcd5 + rootcd6 + rootcd7 + rootcd8 + rootcd9 +
    rootcd11 + rootcd14 + rootcd16 + rootcd18 + rootcd19 + rootcd20 +
    rootcd21 + rootcd22 + rootcd25 + rootcd26 + rootcd27 + rootcd29 +
    rootcd30 + rootcd31 + rootcd33 + rootcd34 + rootcd35 + rootcd36

Complete :
                 (Intercept) reslanAkan reslanEwe reslanGaAdangbe
loc5RuralForest  0           0          0         0
loc5RuralSavannah 0          0          0         0
cropcd0          0           0          0         0
rootcd0          0           0          0         0
rootcd14         0           0          0         0
                 reslanDagbani reslanHausa reslanOther reslanUnknown
loc5RuralForest  0             0           0           0
loc5RuralSavannah 0            0           0           0
cropcd0          0             0           0           0
```

```
rootcd0           0             0           0           0
rootcd14          0             0           0           0
                  ezForest ezSavannah femaleTRUE age avgAge maxAge minAge
loc5RuralForest   1         0          0          0   0      0      0
loc5RuralSavannah 0         1          0          0   0      0      0
cropcd0           0         0          0          0   0      0      0
rootcd0           0         0          0          0   0      0      0
rootcd14          0         0          0          0   0      0      0
                  educBasicEducation educSecondaryEducation
loc5RuralForest   0                  0
loc5RuralSavannah 0                  0
cropcd0           0                  0
rootcd0           0                  0
rootcd14          0                  0
                  educTertiaryEducation educOther marketTRUE transportTRUE
loc5RuralForest   0                     0         0          0
loc5RuralSavannah 0                     0         0          0
cropcd0           0                     0         0          0
rootcd0           0                     0         0          0
rootcd14          0                     0         0          0
                  livstcdTypeCount livstcd1 livstcd2 livstcd3 livstcd4
loc5RuralForest   0                0        0        0        0
loc5RuralSavannah 0                0        0        0        0
cropcd0           0                0        0        0        0
rootcd0           0                0        0        0        0
rootcd14          0                0        0        0        0
                  livstcd5 livstcd6 livstcd7 livstcd8 livstcd9 livstcd10
loc5RuralForest   0        0        0        0        0        0
loc5RuralSavannah 0        0        0        0        0        0
cropcd0           0        0        0        0        0        0
rootcd0           0        0        0        0        0        0
rootcd14          0        0        0        0        0        0
                  livstcd11 livstcd12 equipTypeCount eqcdown21 eqcdown22
loc5RuralForest   0         0         0              0         0
loc5RuralSavannah 0         0         0              0         0
cropcd0           0         0         0              0         0
rootcd0           0         0         0              0         0
rootcd14          0         0         0              0         0
                  eqcdown31 eqcdown51 eqcdown61 eqcdown62 eqcdown63
loc5RuralForest   0         0         0         0         0
loc5RuralSavannah 0         0         0         0         0
cropcd0           0         0         0         0         0
rootcd0           0         0         0         0         0
rootcd14          0         0         0         0         0
                  eqcdown64 eqcdown65 cropTypeCount cropcd1 cropcd2 cropcd3
loc5RuralForest   0         0         0             0       0       0
loc5RuralSavannah 0         0         0             0       0       0
```

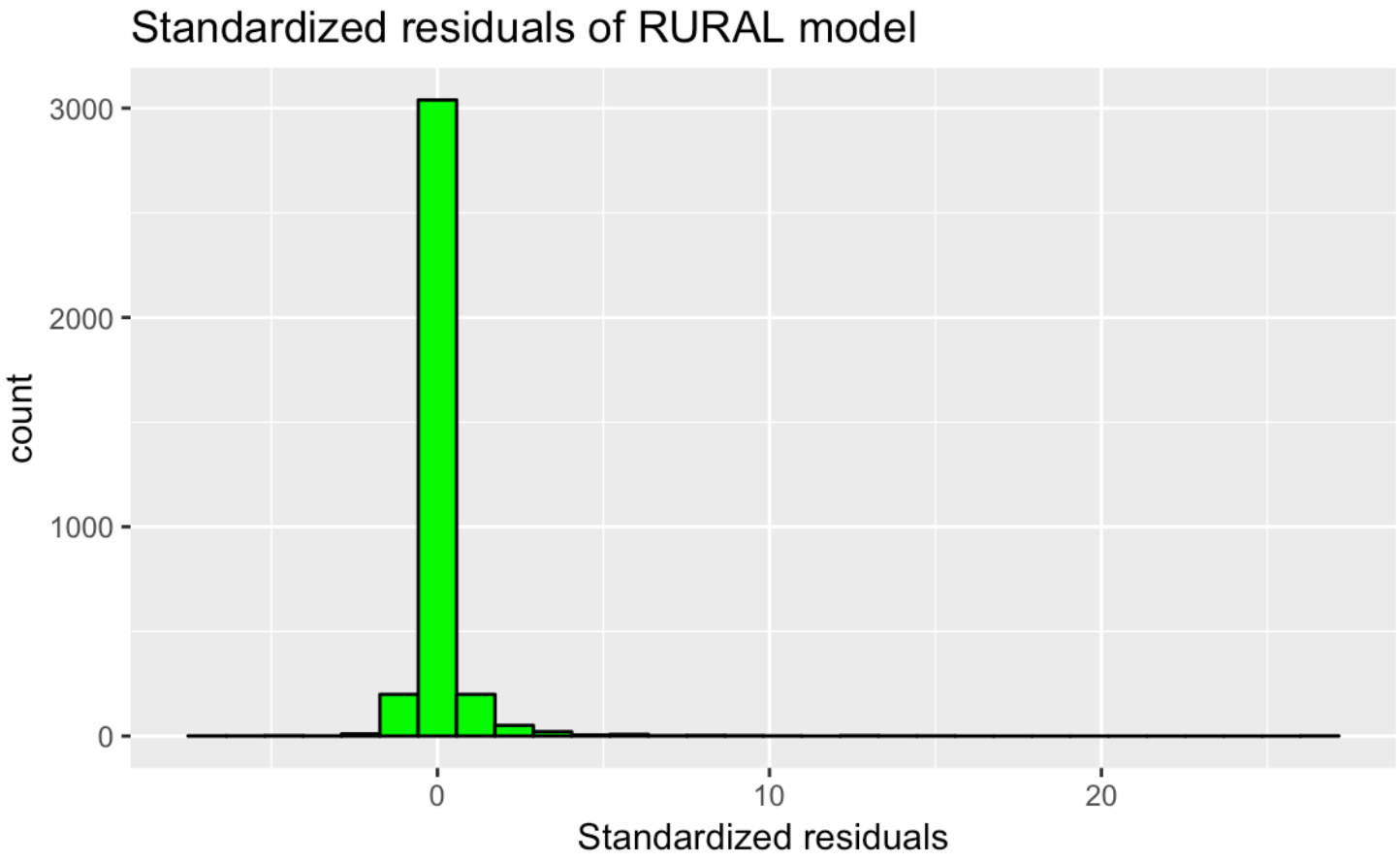
```
cropcd0            0          0          0              0       0       0
rootcd0            0          0          0              0       0       0
rootcd14           0          0          0              0       0       0
                   cropcd4 cropcd5 cropcd6 cropcd8 cropcd9 cropcd10 cropcd11
loc5RuralForest    0       0       0       0       0       0        0
loc5RuralSavannah  0       0       0       0       0       0        0
cropcd0            0       0       0       0       0       0        0
rootcd0            0       0       0       0       0       0        0
rootcd14           0       0       0       0       0       0        0
                   cropcd12 cropcd13 cropcd14 cropcd15 cropcd16 cropcd17
loc5RuralForest    0        0        0        0        0        0
loc5RuralSavannah  0        0        0        0        0        0
cropcd0            0        0        0        0        0        0
rootcd0            0        0        0        0        0        0
rootcd14           0        0        0        0        0        0
                   cropcd18 cropcd19 cropcd20 cropcd21 cropcd22 cropcd23
loc5RuralForest    0        0        0        0        0        0
loc5RuralSavannah  0        0        0        0        0        0
cropcd0            0        0        0        0        0        0
rootcd0            0        0        0        0        0        0
rootcd14           0        0        0        0        0        0
                   cropcd24 cropcd25 cropcd26 cropcd27 cropcd28 cropcd29
loc5RuralForest    0        0        0        0        0        0
loc5RuralSavannah  0        0        0        0        0        0
cropcd0            0        0        0        0        0        0
rootcd0            0        0        0        0        0        0
rootcd14           0        0        0        0        0        0
                   cropcd31 cropcd32 cropcd33 cropcd34 cropcd35
loc5RuralForest    0        0        0        0        0
loc5RuralSavannah  0        0        0        0        0
cropcd0            0        0        0        0        0
rootcd0            0        0        0        0        0
rootcd14           0        0        0        0        0
                   rootTypeCount rootcd5 rootcd6 rootcd7 rootcd8 rootcd9
loc5RuralForest    0             0       0       0       0       0
loc5RuralSavannah  0             0       0       0       0       0
cropcd0            0             0       0       0       0       0
rootcd0            0             0       0       0       0       0
rootcd14           0             0       0       0       0       0
                   rootcd11 rootcd16 rootcd18 rootcd19 rootcd20 rootcd21
loc5RuralForest    0        0        0        0        0        0
loc5RuralSavannah  0        0        0        0        0        0
cropcd0            0        0        0        0        0        0
rootcd0            0        0        0        0        0        0
rootcd14           0        0        0        0        0        0
                   rootcd22 rootcd25 rootcd26 rootcd27 rootcd29 rootcd30
loc5RuralForest    0        0        0        0        0        0
```

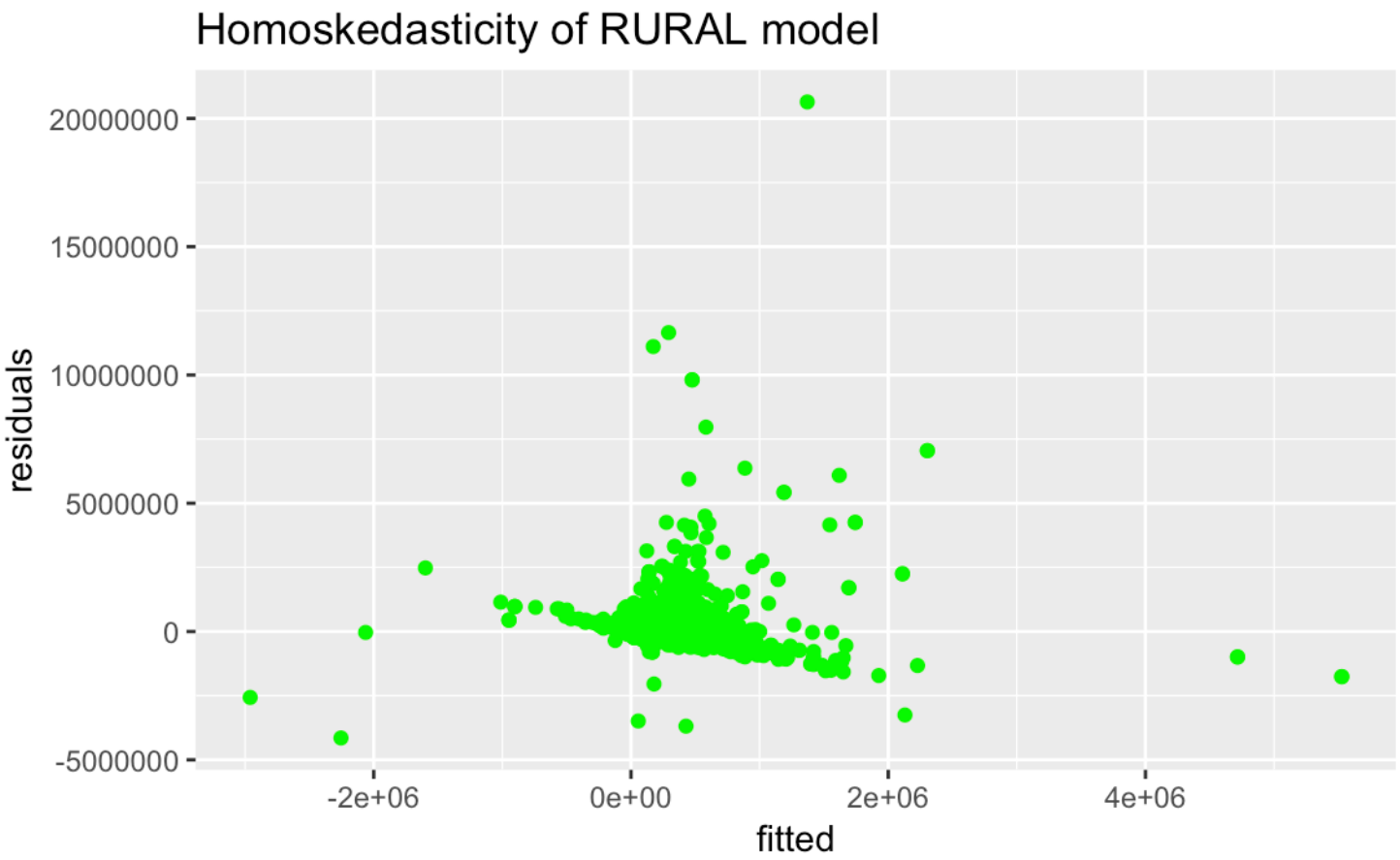
```
loc5RuralSavannah 0         0         0         0         0         0
cropcd0           0         0         0         0         0         0
rootcd0           0         0         0         0         0         0
rootcd14          0         0         0         0         0         0
                  rootcd31 rootcd33 rootcd34 rootcd35 rootcd36
loc5RuralForest   0         0         0         0         0
loc5RuralSavannah 0         0         0         0         0
cropcd0           0         0         0         0         0
rootcd0           0         0         0         0         0
rootcd14          0         0         0         0         0
```

```
There are correlated variables. See above model alias.
```

Hide

```
suppressWarnings(plotResiduals(hh_profit_rural, hh_profit_model_rural, "green", "RURAL"))
```





Unrestricted model for urban area

Fit unrestricted model for urban area profit. We notice that the adjusted R squared is really low, and there are a lot of correlated variable. By looking at data size, we see that there are less observations than variables for urban are profit data. So we don't have enough data to fit a model.

Hide

```
hh_profit_model_urban <- lm(profit ~ .,
                            data = hh_profit_urban)
checkCorrVarAndTestHnull(hh_profit_model_urban)
```

```
====================== model summary ======================

Call:
lm(formula = profit ~ ., data = hh_profit_urban)

Residuals:
    Min      1Q  Median      3Q     Max
-609766  -62750       0   43147  709516

Coefficients: (53 not defined because of singularities)
                        Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)             -3.969e+05  1.048e+06  -0.379    0.712
reslanAkan               1.276e+05  7.214e+05   0.177    0.863
reslanUnknown            2.854e+05  8.071e+05   0.354    0.730
ezSavannah              -9.612e+05  6.879e+05  -1.397    0.188
femaleTRUE              -1.599e+04  2.718e+05  -0.059    0.954
age                      1.074e+04  1.302e+04   0.825    0.425
avgAge                  -1.654e+04  3.730e+04  -0.443    0.665
maxAge                   6.680e+03  1.790e+04   0.373    0.716
minAge                   4.451e+02  2.880e+04   0.015    0.988
educBasicEducation       3.686e+05  3.712e+05   0.993    0.340
educSecondaryEducation   2.229e+05  8.876e+05   0.251    0.806
educTertiaryEducation    9.755e+02  5.419e+05   0.002    0.999
marketTRUE               1.075e+06  5.401e+05   1.989    0.070 .
transportTRUE                   NA         NA      NA       NA
livstcdTypeCount        -8.311e+05  8.671e+05  -0.959    0.357
livstcd1                        NA         NA      NA       NA
livstcd2                        NA         NA      NA       NA
livstcd3                 9.080e+04  1.196e+05   0.759    0.462
livstcd4                 5.326e+04  1.320e+05   0.403    0.694
livstcd5                 1.049e+05  6.914e+04   1.517    0.155
livstcd6                        NA         NA      NA       NA
livstcd7                 7.646e+04  7.178e+04   1.065    0.308
livstcd8                 8.764e+04  9.539e+04   0.919    0.376
livstcd9                        NA         NA      NA       NA
livstcd10                       NA         NA      NA       NA
livstcd11                       NA         NA      NA       NA
livstcd12                       NA         NA      NA       NA
equipTypeCount           7.001e+05  1.150e+06   0.609    0.554
eqcdown21                       NA         NA      NA       NA
eqcdown22                       NA         NA      NA       NA
eqcdown31                       NA         NA      NA       NA
eqcdown51                       NA         NA      NA       NA
eqcdown61                       NA         NA      NA       NA
eqcdown62                       NA         NA      NA       NA
eqcdown63                       NA         NA      NA       NA
eqcdown64                       NA         NA      NA       NA
eqcdown65               -2.093e+05  1.025e+06  -0.204    0.842
cropTypeCount           -9.111e+04  4.412e+05  -0.206    0.840
cropcd0                         NA         NA      NA       NA
cropcd1                 -6.547e+04  1.327e+05  -0.493    0.631
cropcd2                         NA         NA      NA       NA
cropcd3                         NA         NA      NA       NA
cropcd4                         NA         NA      NA       NA
cropcd5                         NA         NA      NA       NA
cropcd6                         NA         NA      NA       NA
cropcd8                         NA         NA      NA       NA
cropcd9                         NA         NA      NA       NA
```

```
cropcd10              3.242e+03  6.012e+03   0.539    0.600
cropcd11                     NA         NA      NA       NA
cropcd12                     NA         NA      NA       NA
cropcd13                     NA         NA      NA       NA
cropcd14                     NA         NA      NA       NA
cropcd15                     NA         NA      NA       NA
cropcd16                     NA         NA      NA       NA
cropcd17             -1.669e+02  9.282e+04  -0.002    0.999
cropcd18                     NA         NA      NA       NA
cropcd19                     NA         NA      NA       NA
cropcd20                     NA         NA      NA       NA
cropcd21                     NA         NA      NA       NA
cropcd22             -1.723e+04  1.136e+04  -1.517    0.155
cropcd23                     NA         NA      NA       NA
cropcd24                     NA         NA      NA       NA
cropcd25                     NA         NA      NA       NA
cropcd26                     NA         NA      NA       NA
cropcd27                     NA         NA      NA       NA
cropcd28              9.335e+05  2.474e+06   0.377    0.713
cropcd29                     NA         NA      NA       NA
cropcd31                     NA         NA      NA       NA
cropcd32                     NA         NA      NA       NA
cropcd33                     NA         NA      NA       NA
cropcd34                     NA         NA      NA       NA
cropcd35                     NA         NA      NA       NA
rootTypeCount         1.883e+04  1.323e+05   0.142    0.889
rootcd0                      NA         NA      NA       NA
rootcd5               3.974e+04  9.861e+04   0.403    0.694
rootcd6               1.564e+05  2.891e+05   0.541    0.598
rootcd7              -4.601e+04  2.293e+05  -0.201    0.844
rootcd8               2.544e+04  9.786e+04   0.260    0.799
rootcd9                      NA         NA      NA       NA
rootcd11              2.299e+07  6.715e+07   0.342    0.738
rootcd14                     NA         NA      NA       NA
rootcd16             -1.143e+07  3.387e+07  -0.337    0.742
rootcd18             -1.380e+04  1.851e+05  -0.075    0.942
rootcd19             -1.535e+05  7.176e+05  -0.214    0.834
rootcd20              2.374e+05  3.858e+05   0.615    0.550
rootcd21                     NA         NA      NA       NA
rootcd22                     NA         NA      NA       NA
rootcd25              9.814e+04  2.739e+05   0.358    0.726
rootcd26             -1.006e+05  1.687e+05  -0.597    0.562
rootcd27              2.846e+05  8.460e+05   0.336    0.742
rootcd29              1.115e+04  1.757e+05   0.063    0.950
rootcd30              5.394e+03  1.116e+05   0.048    0.962
rootcd31                     NA         NA      NA       NA
rootcd33             -3.586e+04  3.993e+04  -0.898    0.387
```

```
rootcd34                        NA         NA      NA       NA
rootcd35                        NA         NA      NA       NA
rootcd36              -2.105e+04  1.195e+05  -0.176    0.863
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 475100 on 12 degrees of freedom
Multiple R-squared:  0.7237,    Adjusted R-squared:  -0.2664
F-statistic: 0.7309 on 43 and 12 DF,  p-value: 0.7818

==================== model alias ========================
Model :
profit ~ reslan + ez + female + age + avgAge + maxAge + minAge +
    educ + market + transport + livstcdTypeCount + livstcd1 +
    livstcd2 + livstcd3 + livstcd4 + livstcd5 + livstcd6 + livstcd7 +
    livstcd8 + livstcd9 + livstcd10 + livstcd11 + livstcd12 +
    equipTypeCount + eqcdown21 + eqcdown22 + eqcdown31 + eqcdown51 +
    eqcdown61 + eqcdown62 + eqcdown63 + eqcdown64 + eqcdown65 +
    cropTypeCount + cropcd0 + cropcd1 + cropcd2 + cropcd3 + cropcd4 +
    cropcd5 + cropcd6 + cropcd8 + cropcd9 + cropcd10 + cropcd11 +
    cropcd12 + cropcd13 + cropcd14 + cropcd15 + cropcd16 + cropcd17 +
    cropcd18 + cropcd19 + cropcd20 + cropcd21 + cropcd22 + cropcd23 +
    cropcd24 + cropcd25 + cropcd26 + cropcd27 + cropcd28 + cropcd29 +
    cropcd31 + cropcd32 + cropcd33 + cropcd34 + cropcd35 + rootTypeCount +
    rootcd0 + rootcd5 + rootcd6 + rootcd7 + rootcd8 + rootcd9 +
    rootcd11 + rootcd14 + rootcd16 + rootcd18 + rootcd19 + rootcd20 +
    rootcd21 + rootcd22 + rootcd25 + rootcd26 + rootcd27 + rootcd29 +
    rootcd30 + rootcd31 + rootcd33 + rootcd34 + rootcd35 + rootcd36

Complete :
             (Intercept) reslanAkan reslanUnknown ezSavannah femaleTRUE
transportTRUE           1          0            0          0          0
livstcd1                0          0            0          0          0
livstcd2                0          0            0          0          0
livstcd6                0          0            0          0          0
livstcd9                0          0            0          0          0
livstcd10               0          0            0          0          0
livstcd11               0          0            0          0          0
livstcd12               0          0            0          0          0
eqcdown21               0          0            0          0          0
eqcdown22               0          0            0          0          0
eqcdown31               0          0            0          0          0
eqcdown51               0          0            0          0          0
eqcdown61               0          0            0          0          0
eqcdown62               0          0            0          0          0
eqcdown63               0          0            0          0          0
eqcdown64               0          0            0          0          0
```

```
cropcd0          0            0           0               0           0
cropcd2          0            0           0               0           0
cropcd3          0            0           0               0           0
cropcd4          0            0           0               0           0
cropcd5          0            0           0               0           0
cropcd6          0            0           0               0           0
              age avgAge maxAge minAge educBasicEducation
transportTRUE   0    0      0      0      0
livstcd1        0    0      0      0      0
livstcd2        0    0      0      0      0
livstcd6        0    0      0      0      0
livstcd9        0    0      0      0      0
livstcd10       0    0      0      0      0
livstcd11       0    0      0      0      0
livstcd12       0    0      0      0      0
eqcdown21       0    0      0      0      0
eqcdown22       0    0      0      0      0
eqcdown31       0    0      0      0      0
eqcdown51       0    0      0      0      0
eqcdown61       0    0      0      0      0
eqcdown62       0    0      0      0      0
eqcdown63       0    0      0      0      0
eqcdown64       0    0      0      0      0
cropcd0         0    0      0      0      0
cropcd2         0    0      0      0      0
cropcd3         0    0      0      0      0
cropcd4         0    0      0      0      0
cropcd5         0    0      0      0      0
cropcd6         0    0      0      0      0
              educSecondaryEducation educTertiaryEducation marketTRUE
transportTRUE   0                      0                     0
livstcd1        0                      0                     0
livstcd2        0                      0                     0
livstcd6        0                      0                     0
livstcd9        0                      0                     0
livstcd10       0                      0                     0
livstcd11       0                      0                     0
livstcd12       0                      0                     0
eqcdown21       0                      0                     0
eqcdown22       0                      0                     0
eqcdown31       0                      0                     0
eqcdown51       0                      0                     0
eqcdown61       0                      0                     0
eqcdown62       0                      0                     0
eqcdown63       0                      0                     0
eqcdown64       0                      0                     0
cropcd0         0                      0                     0
```

cropcd2	0	0	0		
cropcd3	0	0	0		
cropcd4	0	0	0		
cropcd5	0	0	0		
cropcd6	0	0	0		

	livstcdTypeCount	livstcd3	livstcd4	livstcd5	livstcd7	livstcd8
transportTRUE	0	0	0	0	0	0
livstcd1	0	0	0	0	0	0
livstcd2	0	0	0	0	0	0
livstcd6	0	0	0	0	0	0
livstcd9	0	0	0	0	0	0
livstcd10	0	0	0	0	0	0
livstcd11	0	0	0	0	0	0
livstcd12	0	0	0	0	0	0
eqcdown21	0	0	0	0	0	0
eqcdown22	0	0	0	0	0	0
eqcdown31	0	0	0	0	0	0
eqcdown51	0	0	0	0	0	0
eqcdown61	0	0	0	0	0	0
eqcdown62	0	0	0	0	0	0
eqcdown63	0	0	0	0	0	0
eqcdown64	0	0	0	0	0	0
cropcd0	0	0	0	0	0	0
cropcd2	0	0	0	0	0	0
cropcd3	0	0	0	0	0	0
cropcd4	0	0	0	0	0	0
cropcd5	0	0	0	0	0	0
cropcd6	0	0	0	0	0	0

	equipTypeCount	eqcdown65	cropTypeCount	cropcd1	cropcd10
transportTRUE	0	0	0	0	0
livstcd1	0	0	0	0	0
livstcd2	0	0	0	0	0
livstcd6	0	0	0	0	0
livstcd9	0	0	0	0	0
livstcd10	0	0	0	0	0
livstcd11	0	0	0	0	0
livstcd12	0	0	0	0	0
eqcdown21	0	0	0	0	0
eqcdown22	0	0	0	0	0
eqcdown31	0	0	0	0	0
eqcdown51	0	0	0	0	0
eqcdown61	0	0	0	0	0
eqcdown62	0	0	0	0	0
eqcdown63	0	0	0	0	0
eqcdown64	0	0	0	0	0
cropcd0	0	0	0	0	0
cropcd2	0	0	0	0	0

cropcd3	0	0	0	0	0
cropcd4	0	0	0	0	0
cropcd5	0	0	0	0	0
cropcd6	0	0	0	0	0

	cropcd17	cropcd22	cropcd28	rootTypeCount	rootcd5	rootcd6
transportTRUE	0	0	0	0	0	0
livstcd1	0	0	0	0	0	0
livstcd2	0	0	0	0	0	0
livstcd6	0	0	0	0	0	0
livstcd9	0	0	0	0	0	0
livstcd10	0	0	0	0	0	0
livstcd11	0	0	0	0	0	0
livstcd12	0	0	0	0	0	0
eqcdown21	0	0	0	0	0	0
eqcdown22	0	0	0	0	0	0
eqcdown31	0	0	0	0	0	0
eqcdown51	0	0	0	0	0	0
eqcdown61	0	0	0	0	0	0
eqcdown62	0	0	0	0	0	0
eqcdown63	0	0	0	0	0	0
eqcdown64	0	0	0	0	0	0
cropcd0	0	0	0	0	0	0
cropcd2	0	0	0	0	0	0
cropcd3	0	0	0	0	0	0
cropcd4	0	0	0	0	0	0
cropcd5	0	0	0	0	0	0
cropcd6	0	0	0	0	0	0

	rootcd7	rootcd8	rootcd11	rootcd16	rootcd18	rootcd19	rootcd20
transportTRUE	0	0	0	0	0	0	0
livstcd1	0	0	0	0	0	0	0
livstcd2	0	0	0	0	0	0	0
livstcd6	0	0	0	0	0	0	0
livstcd9	0	0	0	0	0	0	0
livstcd10	0	0	0	0	0	0	0
livstcd11	0	0	0	0	0	0	0
livstcd12	0	0	0	0	0	0	0
eqcdown21	0	0	0	0	0	0	0
eqcdown22	0	0	0	0	0	0	0
eqcdown31	0	0	0	0	0	0	0
eqcdown51	0	0	0	0	0	0	0
eqcdown61	0	0	0	0	0	0	0
eqcdown62	0	0	0	0	0	0	0
eqcdown63	0	0	0	0	0	0	0
eqcdown64	0	0	0	0	0	0	0
cropcd0	0	0	0	0	0	0	0
cropcd2	0	0	0	0	0	0	0
cropcd3	0	0	0	0	0	0	0

```
cropcd4          0        0        0         0         0         0         0
cropcd5          0        0        0         0         0         0         0
cropcd6          0        0        0         0         0         0         0
              rootcd25 rootcd26 rootcd27 rootcd29 rootcd30 rootcd33
transportTRUE    0         0         0         0         0         0
livstcd1         0         0         0         0         0         0
livstcd2         0         0         0         0         0         0
livstcd6         0         0         0         0         0         0
livstcd9         0         0         0         0         0         0
livstcd10        0         0         0         0         0         0
livstcd11        0         0         0         0         0         0
livstcd12        0         0         0         0         0         0
eqcdown21        0         0         0         0         0         0
eqcdown22        0         0         0         0         0         0
eqcdown31        0         0         0         0         0         0
eqcdown51        0         0         0         0         0         0
eqcdown61        0         0         0         0         0         0
eqcdown62        0         0         0         0         0         0
eqcdown63        0         0         0         0         0         0
eqcdown64        0         0         0         0         0         0
cropcd0          0         0         0         0         0         0
cropcd2          0         0         0         0         0         0
cropcd3          0         0         0         0         0         0
cropcd4          0         0         0         0         0         0
cropcd5          0         0         0         0         0         0
cropcd6          0         0         0         0         0         0
              rootcd36
transportTRUE    0
livstcd1         0
livstcd2         0
livstcd6         0
livstcd9         0
livstcd10        0
livstcd11        0
livstcd12        0
eqcdown21        0
eqcdown22        0
eqcdown31        0
eqcdown51        0
eqcdown61        0
eqcdown62        0
eqcdown63        0
eqcdown64        0
cropcd0          0
cropcd2          0
cropcd3          0
cropcd4          0
```

```
cropcd5         0
cropcd6         0
 [ reached getOption("max.print") -- omitted 31 rows ]
```

```
There are correlated variables. See above model alias.
```

Hide

```
# check data size
dim(hh_profit_urban)
```

```
[1] 56 94
```

Restricted model with only education and local characteristic information (R2, R3)

We also want to see how education and local characteristic info alone influence profit. Looking at hypothesis test result, loc5 and loc3 are correlated with other variables. So remove them and fit model again. Looking at the summary, there are not many significant variables, also the R squared is very low.

Hide

```
hh_profit_model_r2 <- lm(profit ~ educ + ez + loc2 + loc5 + loc3 + market + transport
,
                         data = hh_profit)
checkCorrVarAndTestHnull(hh_profit_model_r2)
```

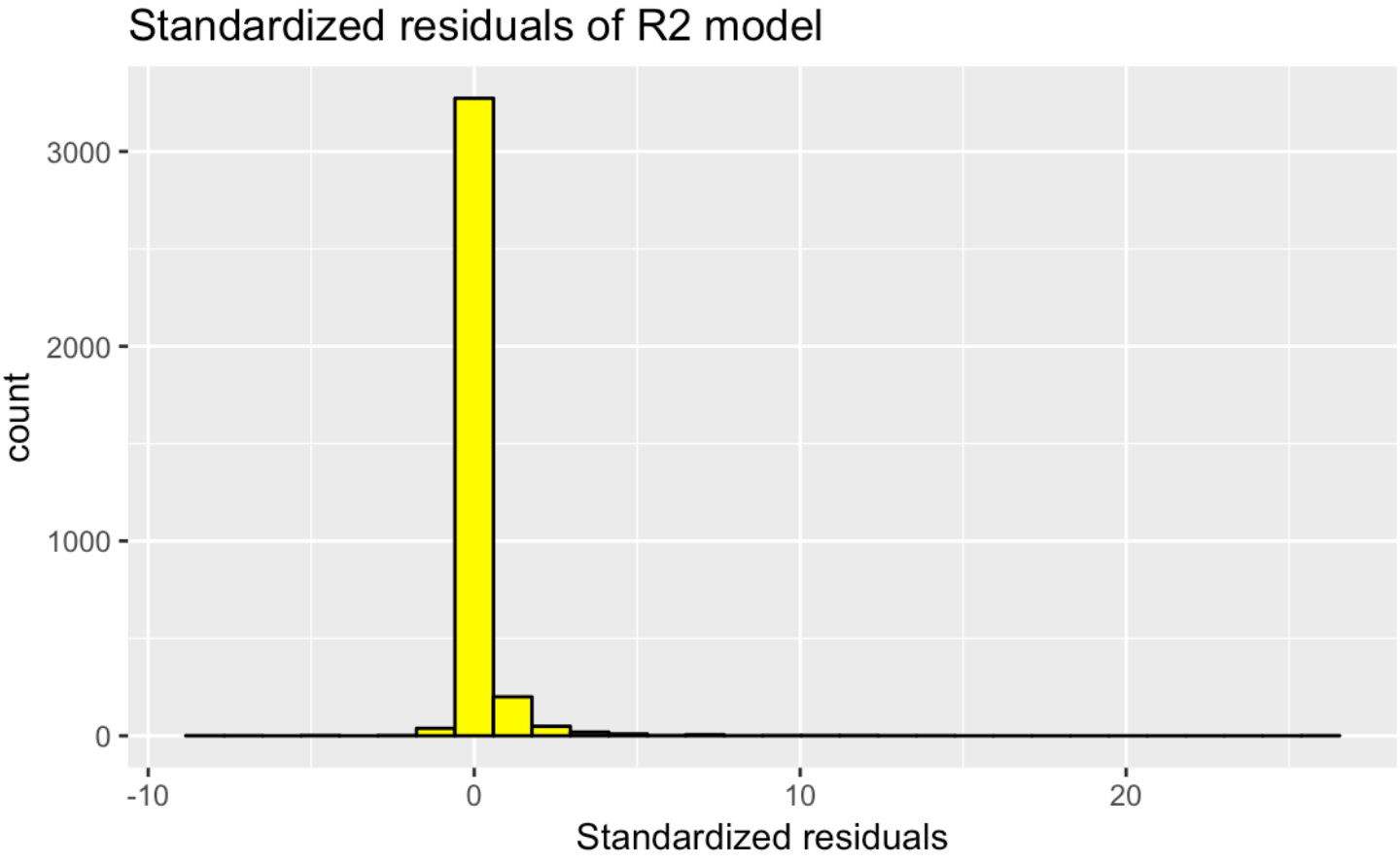
```
====================== model summary ======================

Call:
lm(formula = profit ~ educ + ez + loc2 + loc5 + loc3 + market +
    transport, data = hh_profit)

Residuals:
     Min       1Q   Median       3Q      Max
-6686068  -259175  -173421     3700 21648271

Coefficients: (3 not defined because of singularities)
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)              341148.54  271694.77   1.256  0.20933
educBasicEducation          604.85   31312.81   0.019  0.98459
educSecondaryEducation    20078.89   88658.86   0.226  0.82085
educTertiaryEducation   -114169.37   78495.67  -1.454  0.14590
educOther               -245371.94  181893.76  -1.349  0.17743
ezForest                  79473.13  294500.56   0.270  0.78729
ezSavannah               -25449.92   44857.89  -0.567  0.57051
```

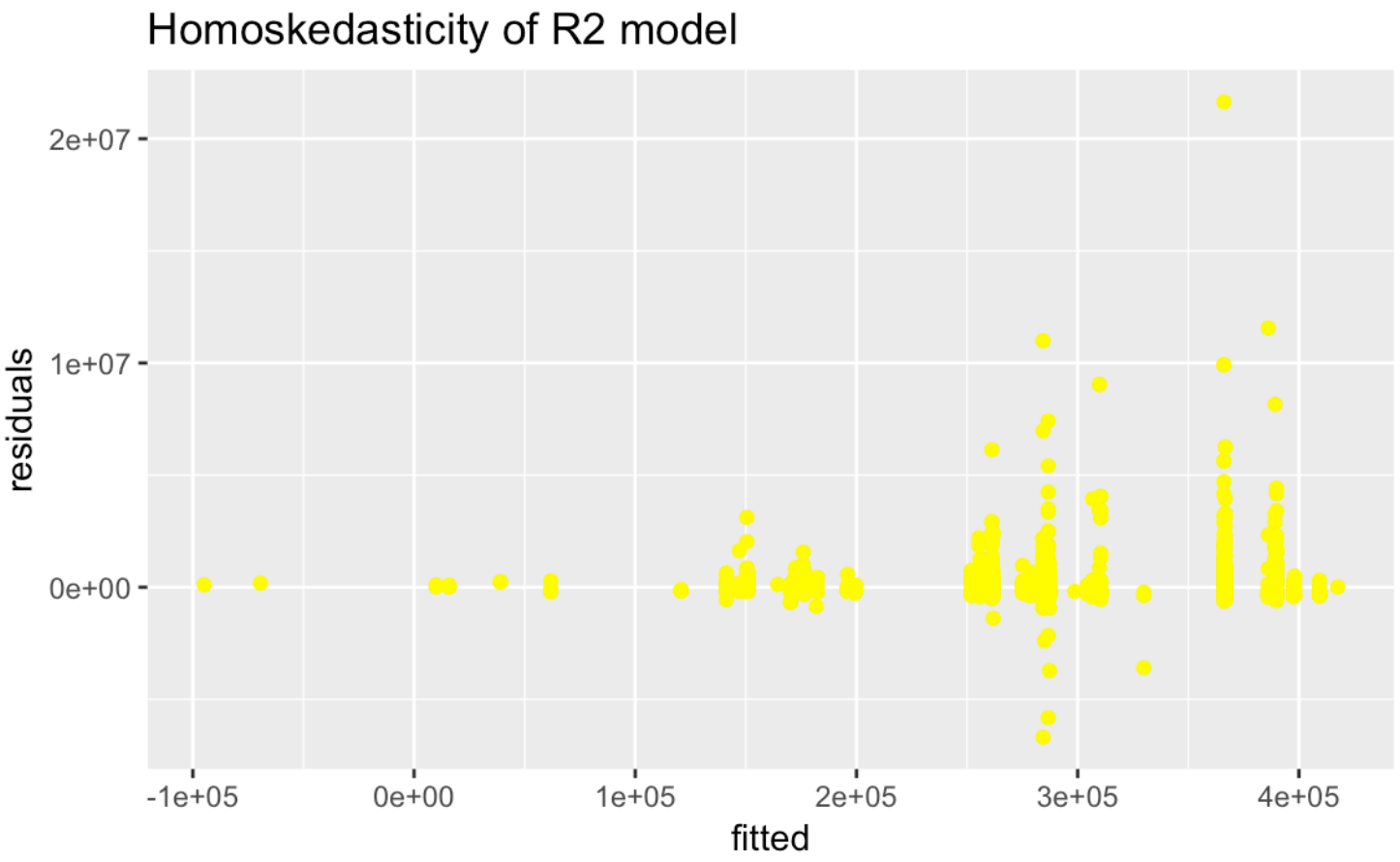
```
loc2Rural               -31347.21  266430.71  -0.118  0.90635
loc5RuralCoastal               NA         NA      NA       NA
loc5RuralForest            -36.17  292751.08   0.000  0.99990
loc5RuralSavannah              NA         NA      NA       NA
loc3Rural                      NA         NA      NA       NA
marketTRUE             -110715.56   41813.82  -2.648  0.00814 **
transportTRUE           -23157.10   30171.81  -0.768  0.44283
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 828500 on 3602 degrees of freedom
Multiple R-squared:  0.006371,  Adjusted R-squared:  0.003612
F-statistic: 2.309 on 10 and 3602 DF,  p-value: 0.01058

====================== model alias ========================
Model :
profit ~ educ + ez + loc2 + loc5 + loc3 + market + transport

Complete :
                (Intercept) educBasicEducation educSecondaryEducation
loc5RuralCoastal   1           0                  0
loc5RuralSavannah -1           0                  0
loc3Rural          0           0                  0
                educTertiaryEducation educOther ezForest ezSavannah
loc5RuralCoastal   0                     0        -1       -1
loc5RuralSavannah  0                     0         1        1
loc3Rural          0                     0         0        0
                loc2Rural loc5RuralForest marketTRUE transportTRUE
loc5RuralCoastal   0         0               0          0
loc5RuralSavannah  1        -1               0          0
loc3Rural          1         0               0          0
```

```
There are correlated variables. See above model alias.
```

Hide

```
suppressWarnings(plotResiduals(hh_profit, hh_profit_model_r2, "yellow", "R2"))
```


Standardized residuals of R2 model
count
3000
2000
1000
0
-10
0
10
20
Standardized residuals



Hide

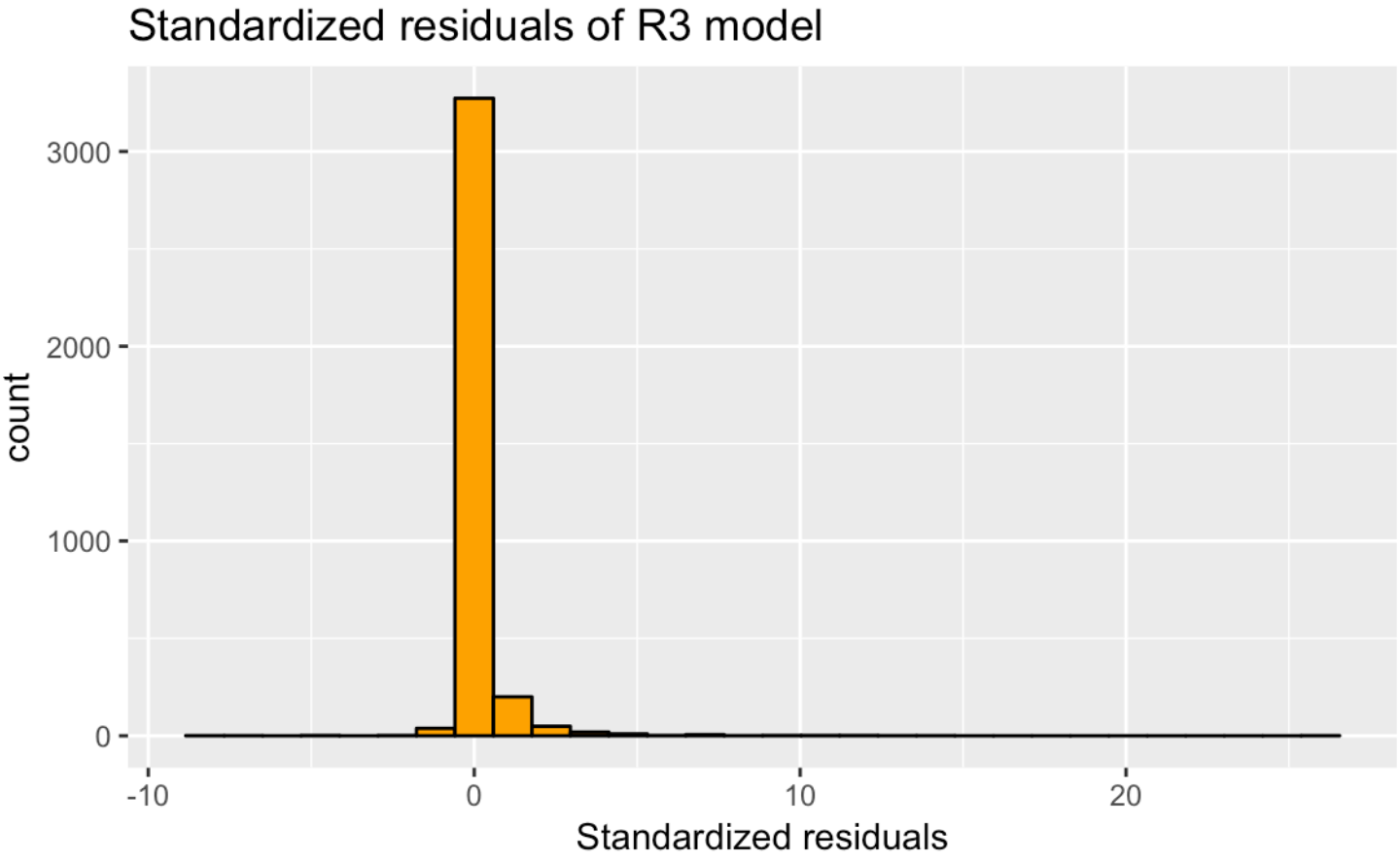
```
# remove loc5 and loc3 and fit model again
hh_profit_model_r3 <- lm(profit ~ educ + ez + loc2 + market + transport,
                         data = hh_profit)
checkCorrVarAndTestHnull(hh_profit_model_r3)
```

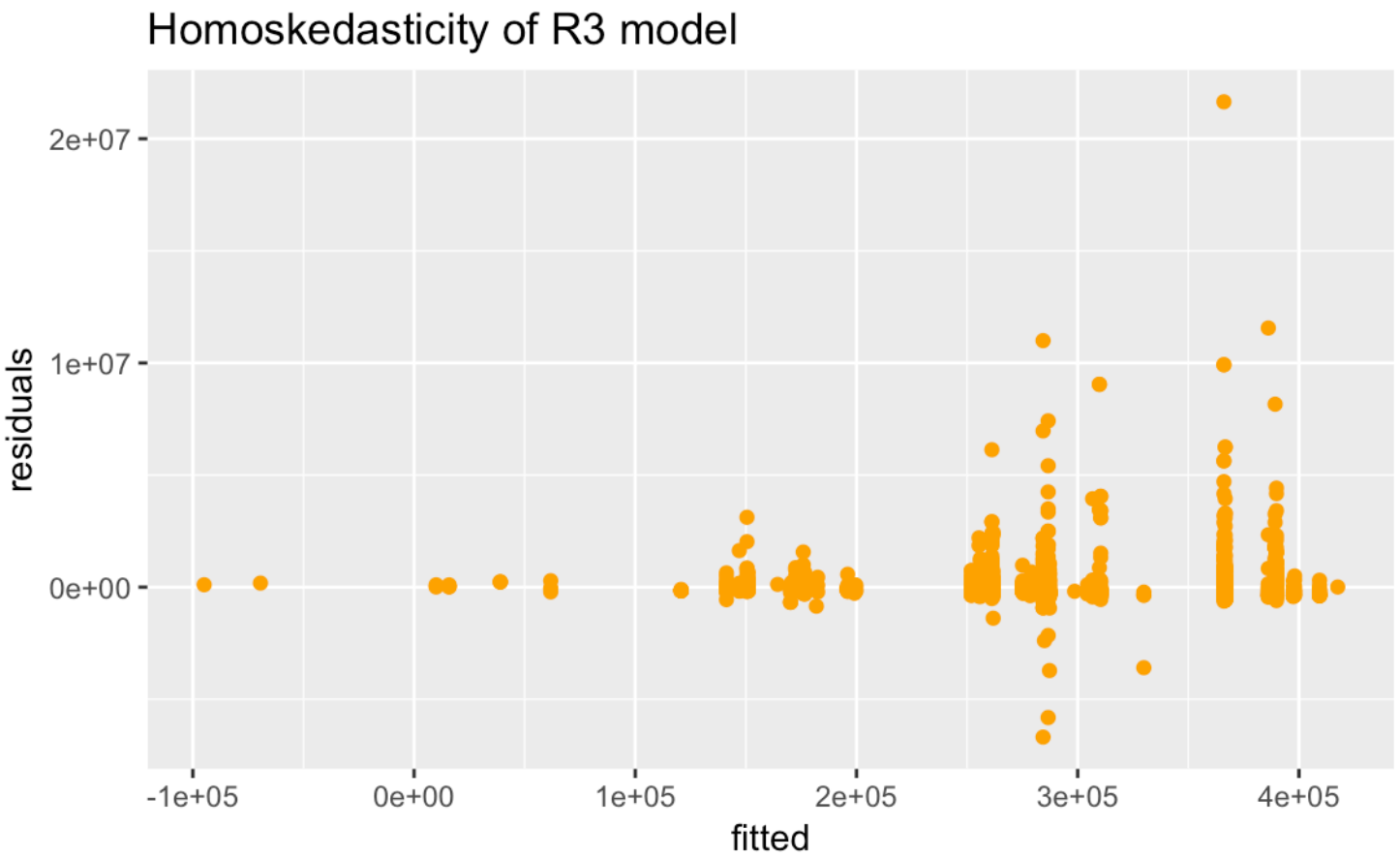
```
==================== model summary ======================

Call:
lm(formula = profit ~ educ + ez + loc2 + market + transport,
    data = hh_profit)

Residuals:
     Min       1Q   Median       3Q      Max
-6686068  -259175  -173421     3700 21648271

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             341178.5   122474.1   2.786  0.00537 **
educBasicEducation         604.7    31295.3   0.019  0.98458
educSecondaryEducation   20079.0    88642.2   0.227  0.82081
```

```
educTertiaryEducation  -114169.3    78481.0  -1.455  0.14583
educOther              -245371.9   181867.3  -1.349  0.17736
ezForest                 79437.1    39472.3   2.012  0.04424 *
ezSavannah              -25450.3    44729.0  -0.569  0.56940
loc2Rural               -31377.0   112961.1  -0.278  0.78121
marketTRUE             -110716.1    41581.9  -2.663  0.00779 **
transportTRUE           -23157.1    30167.4  -0.768  0.44276
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 828400 on 3603 degrees of freedom
Multiple R-squared:  0.006371,  Adjusted R-squared:  0.003889
F-statistic: 2.567 on 9 and 3603 DF,  p-value: 0.006086

==================== hypothesis test ====================
Linear hypothesis test

Hypothesis:
educBasicEducation = 0
educSecondaryEducation = 0
educTertiaryEducation = 0
educOther = 0
ezForest = 0
ezSavannah = 0
loc2Rural = 0
marketTRUE = 0
transportTRUE = 0

Model 1: restricted model
Model 2: profit ~ educ + ez + loc2 + market + transport

  Res.Df        RSS Df  Sum of Sq      F   Pr(>F)
1   3612 2.4883e+15
2   3603 2.4724e+15  9 1.5852e+13 2.5668 0.006086 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
suppressWarnings(plotResiduals(hh_profit, hh_profit_model_r3, "orange", "R3"))
```


Standardized residuals of R3 model
count
3000
2000
1000
0
-10
0
10
20
Standardized residuals



Restricted model with educ * age (R4)

Although the education in the above model is not significant, we are curious to see if educ and age together would become significant. So we fit a model with `educ * age` . However, looking at the summary it's still not significant.

Hide

```
hh_profit_model_r4 <- lm(profit ~ educ * age + female + ez + loc2 + market + transpor
t,
                         data = hh_profit)
checkCorrVarAndTestHnull(hh_profit_model_r4)
```

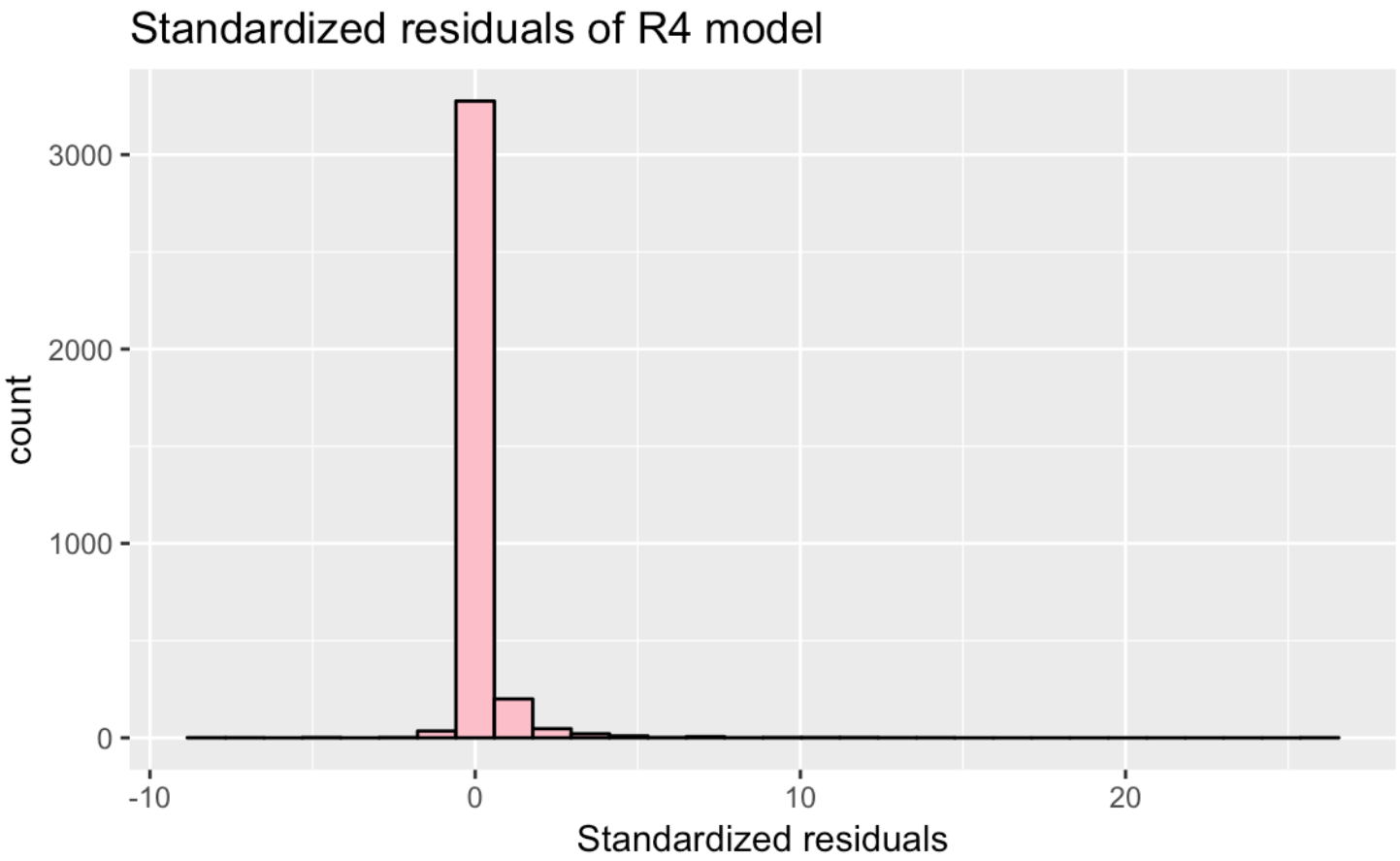
```
==================== model summary ======================

Call:
lm(formula = profit ~ educ * age + female + ez + loc2 + market +
    transport, data = hh_profit)

Residuals:
     Min       1Q   Median       3Q      Max
-6693082  -260609  -173405     4360 21635925
```

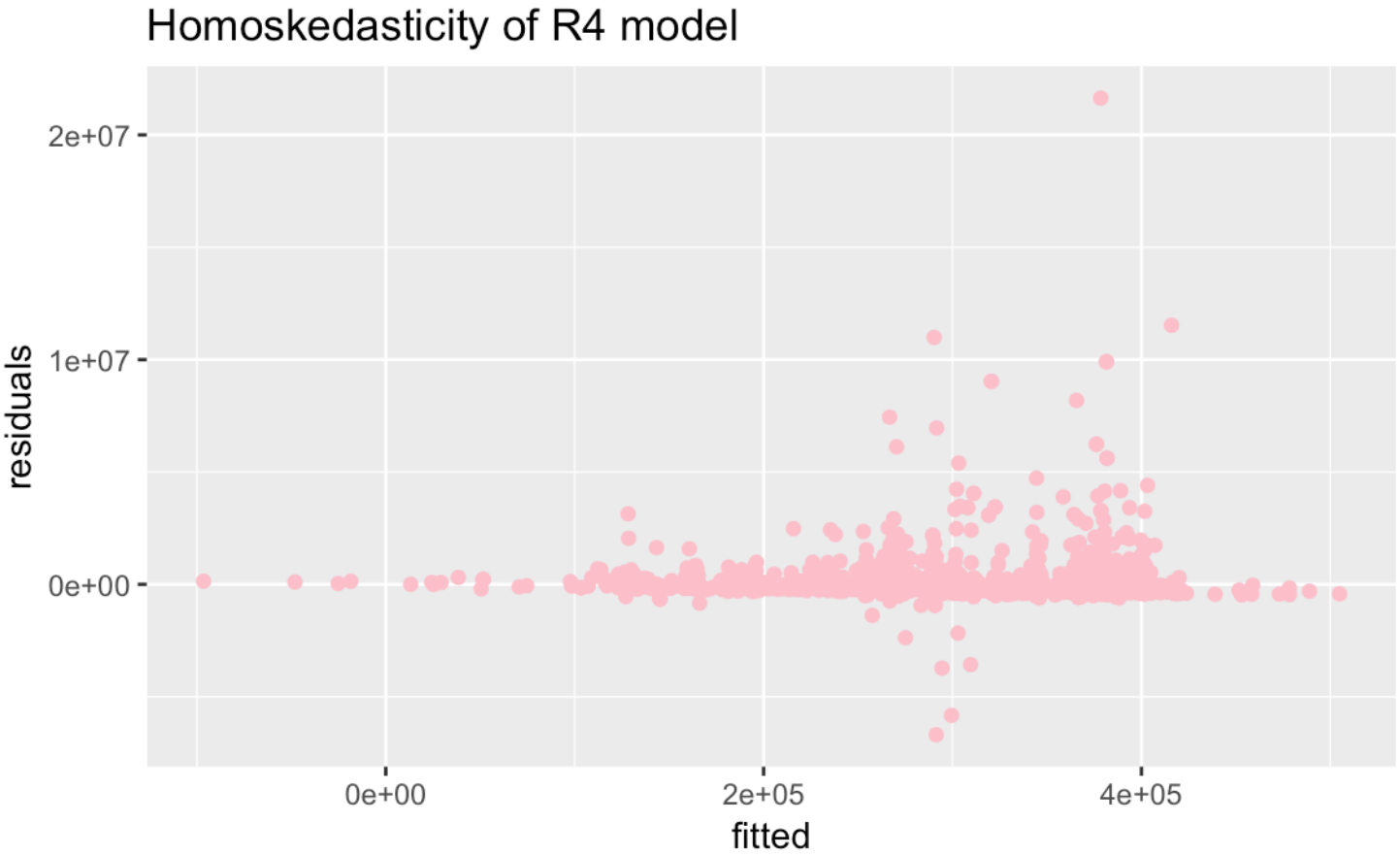
```
Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                363514.37  136714.01   2.659  0.00787 **
educBasicEducation         -43926.23  108243.80  -0.406  0.68491
educSecondaryEducation     206525.08  346672.42   0.596  0.55139
educTertiaryEducation     -221002.20  358817.33  -0.616  0.53799
educOther                  19470.37  903976.83   0.022  0.98282
age                          -75.88    1130.08  -0.067  0.94647
femaleTRUE                 -35363.11   33340.55  -1.061  0.28891
ezForest                    78569.77   39620.05   1.983  0.04743 *
ezSavannah                 -32145.59   45435.95  -0.707  0.47931
loc2Rural                  -37651.00  113301.97  -0.332  0.73968
marketTRUE                -106856.37   41927.77  -2.549  0.01086 *
transportTRUE              -19935.82   30302.98  -0.658  0.51066
educBasicEducation:age        866.44    2429.91   0.357  0.72143
educSecondaryEducation:age  -5229.05    8809.74  -0.594  0.55285
educTertiaryEducation:age    2087.10    7643.66   0.273  0.78483
educOther:age               -5491.26   17510.71  -0.314  0.75385
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 828900 on 3597 degrees of freedom
Multiple R-squared:  0.006882,  Adjusted R-squared:  0.002741
F-statistic: 1.662 on 15 and 3597 DF,  p-value: 0.0515

====================== hypothesis test ======================
Linear hypothesis test

Hypothesis:
educBasicEducation = 0
educSecondaryEducation = 0
educTertiaryEducation = 0
educOther = 0
age = 0
femaleTRUE = 0
ezForest = 0
ezSavannah = 0
loc2Rural = 0
marketTRUE = 0
transportTRUE = 0
educBasicEducation:age = 0
educSecondaryEducation:age = 0
educTertiaryEducation:age = 0
educOther:age = 0

Model 1: restricted model
Model 2: profit ~ educ * age + female + ez + loc2 + market + transport
```

```
  Res.Df        RSS Df  Sum of Sq      F Pr(>F)
1   3612 2.4883e+15
2   3597 2.4711e+15 15 1.7125e+13 1.6618 0.0515 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
suppressWarnings(plotResiduals(hh_profit, hh_profit_model_r4, "pink", "R4"))
```




Homoskedasticity of R4 model
residuals
2e+07
1e+07
0e+00
0e+00
2e+05
4e+05
fitted