MA 615 Refrigerator EDA Project

Shuyi Jiang, Liz Morris, Shan Shan, Luna Tang

10/8/2017

Goal of the project:

Exploring the EIA data, we figured out that household income should be the most important factor for planning and marketing refrigerators. We are going to look at the type, size, age and number of refrigerators owned when the population is broken down by income.

```
library(ggplot2)
```

```
use <- read.csv("use.csv", header = TRUE)
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : incomplete final line found by readTableHeader on 'use.csv'
```

```
use1 <- read.csv("use1.csv", header = TRUE)
incomelevel <- c("under 20", "20 to 40", "40 to 60", "60 to 80", "80 to 100", "100 to 120", "120 to 140
use1$income <- factor(use1$income, levels = incomelevel)
```

Five number summary

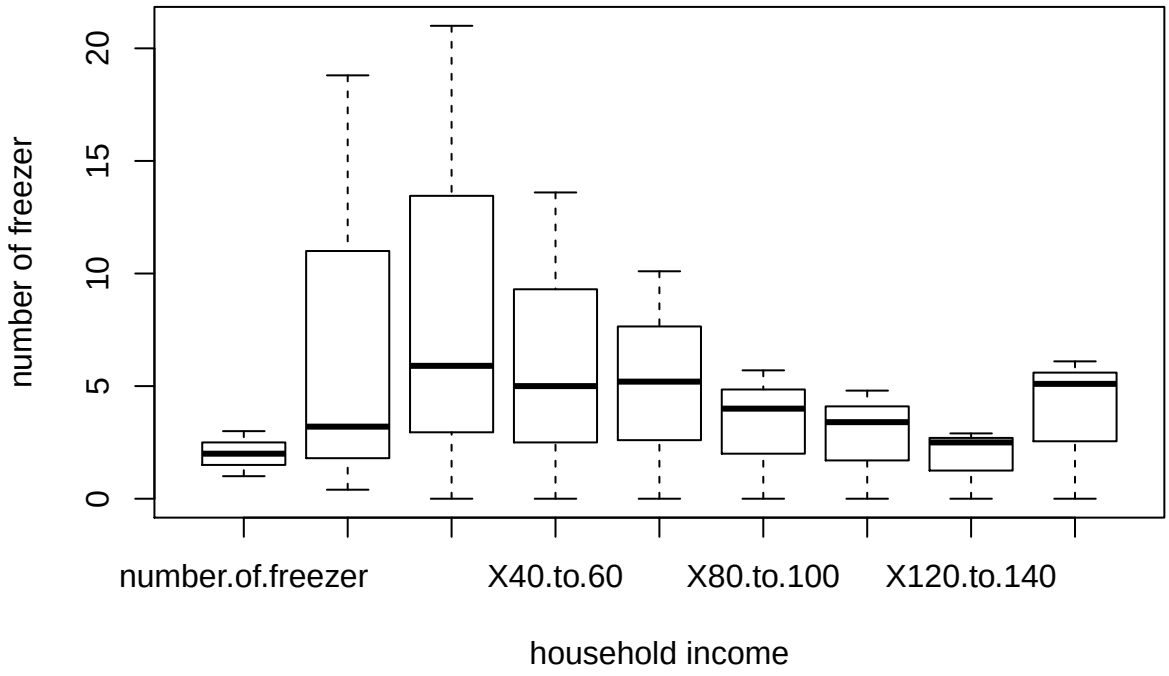
```
summary(use)
```

```
##  number.of.freezer    under.20        X20.to.40        X40.to.60
##  0        :1       Min.   : 0.400   Min.   : 0.000   Min.   : 0.0
##  1        :1       1st Qu.: 1.800   1st Qu.: 2.950   1st Qu.: 2.5
##  2 or more:1       Median : 3.200   Median : 5.900   Median : 5.0
##                    Mean   : 7.467   Mean   : 8.967   Mean   : 6.2
##                    3rd Qu.:11.000   3rd Qu.:13.450   3rd Qu.: 9.3
##                    Max.   :18.800   Max.   :21.000   Max.   :13.6
##    X60.to.80       X80.to.100     X100.to.120     X120.to.140
##  Min.   : 0.00   Min.   :0.000   Min.   :0.000   Min.   :0.00
##  1st Qu.: 2.60   1st Qu.:2.000   1st Qu.:1.700   1st Qu.:1.25
##  Median : 5.20   Median :4.000   Median :3.400   Median :2.50
##  Mean   : 5.10   Mean   :3.233   Mean   :2.733   Mean   :1.80
##  3rd Qu.: 7.65   3rd Qu.:4.850   3rd Qu.:4.100   3rd Qu.:2.70
##  Max.   :10.10   Max.   :5.700   Max.   :4.800   Max.   :2.90
##   X140.or.more
##  Min.   :0.000
##  1st Qu.:2.550
##  Median :5.100
##  Mean   :3.733
##  3rd Qu.:5.600
##  Max.   :6.100
```

boxplot: number of freezer vs household income

The boxplot indicates that around half of the families with household income of 80000 or more have more than 1 freezer. The majority of the riches families with household income of 140000 or more have more than 1 freezer.

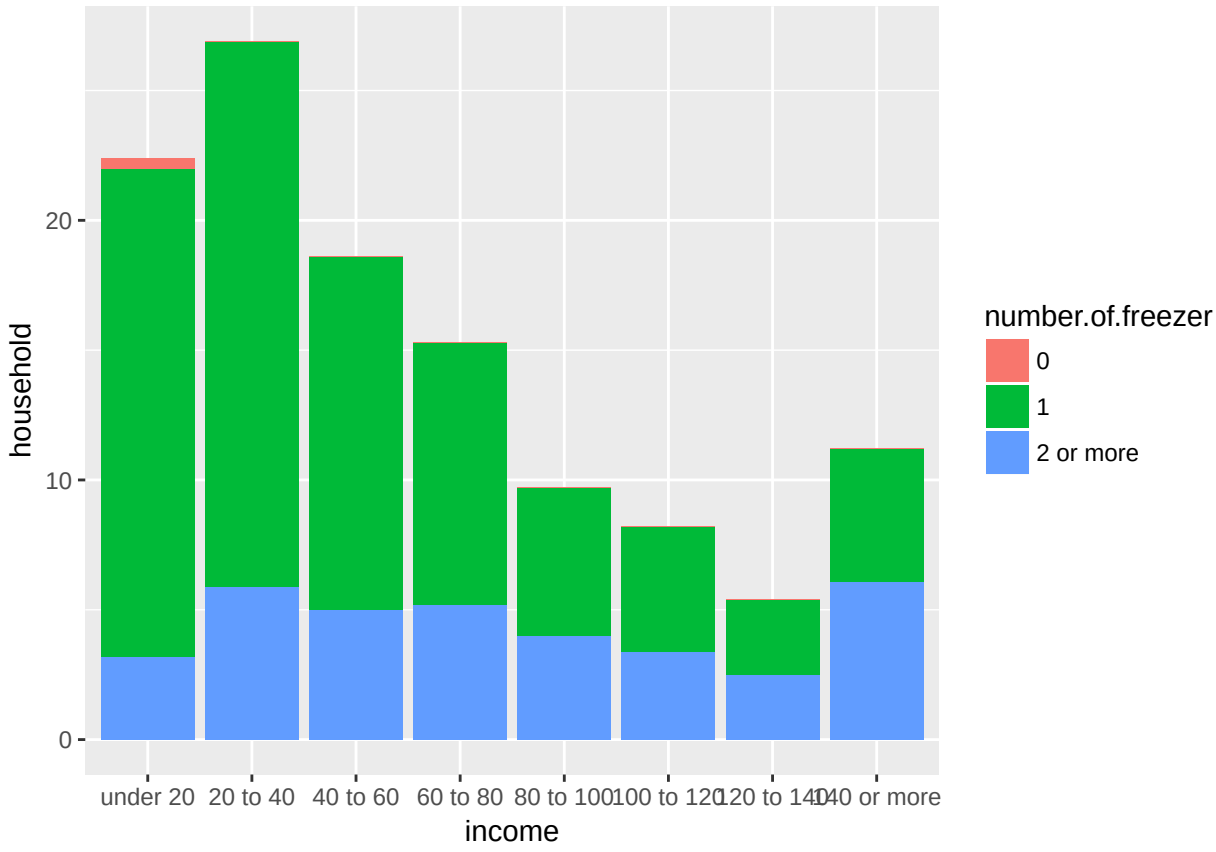
```
boxplot(use, ylab="number of freezer", xlab="household income")
```



histogram: number of freezer vs household income

The histogram indicates the number of freezers own by family with different household income. All the families that do not use freezer have household income under 20000.

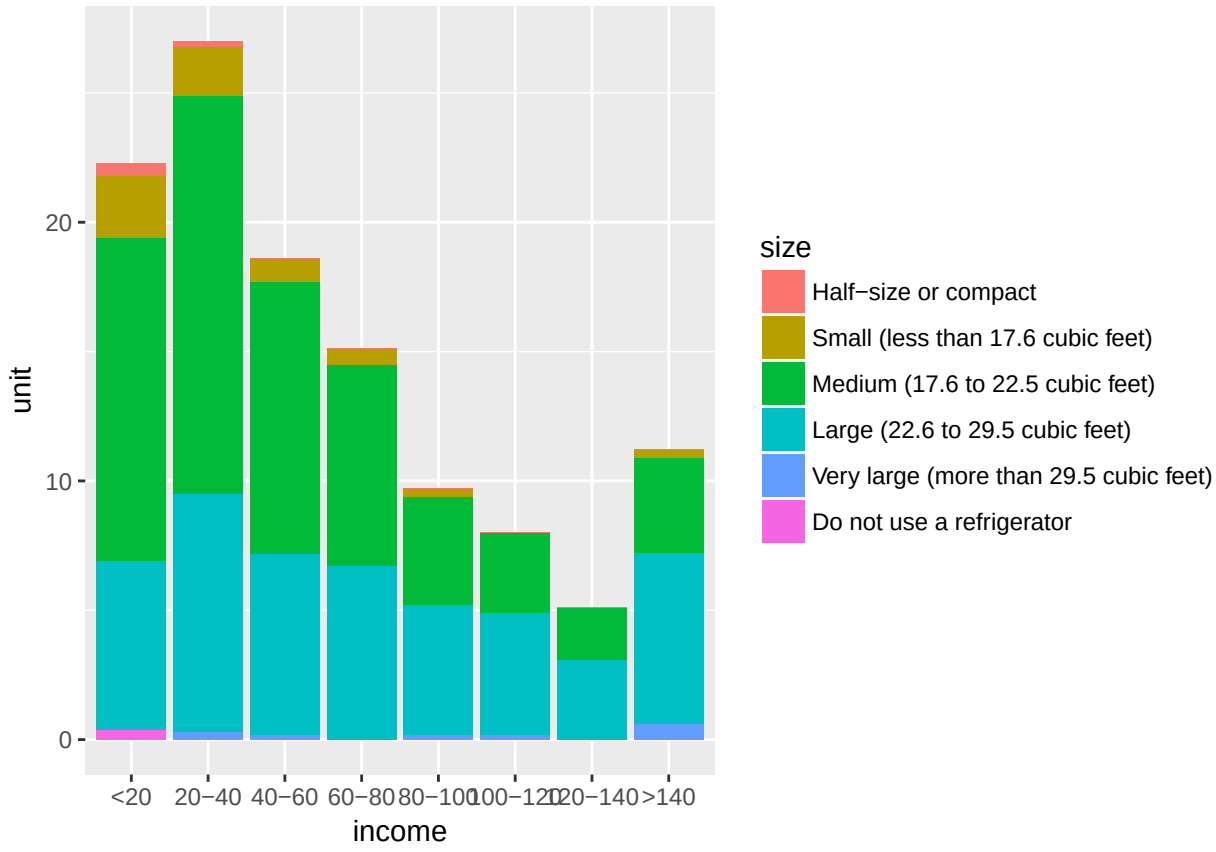
```
ggplot(data = use1, aes(x= income, y = household, fill= number.of.freezer)) + geom_bar(stat="identity")
```



Boxplot for size

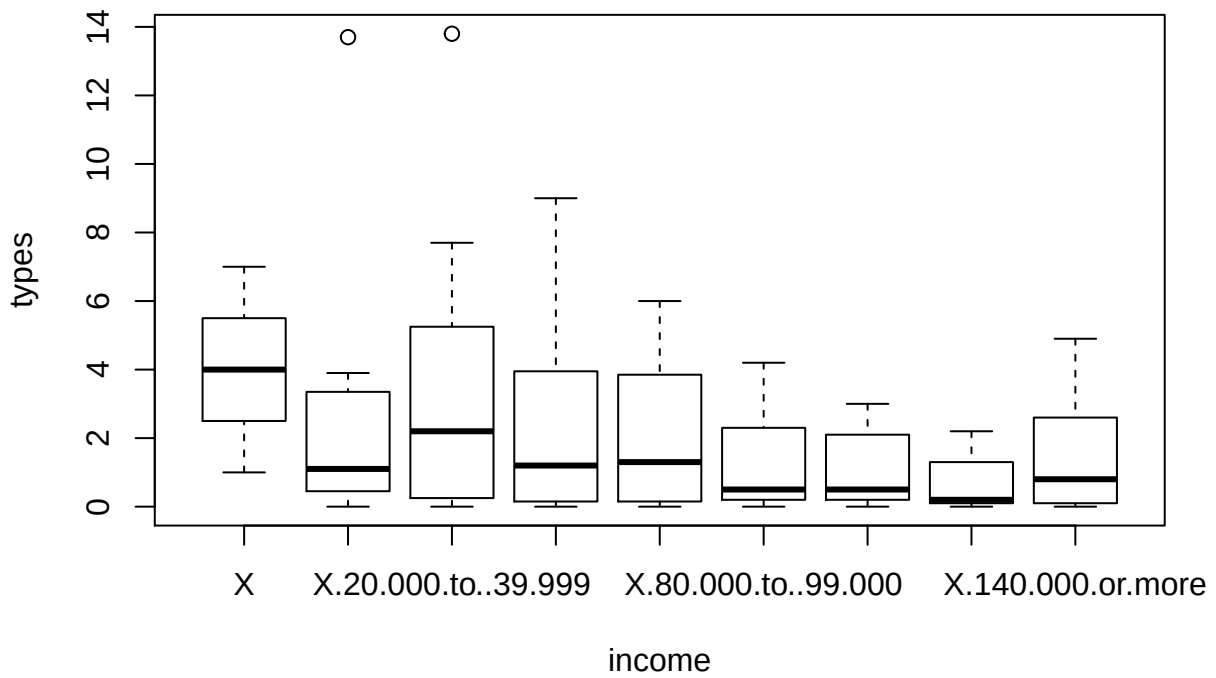
```
size1<-read.csv("size.csv", header=TRUE)
size <-read.csv("size.csv", header=TRUE)

incomeorder<- c('<20', '20-40', '40-60', '60-80','80-100','100-120', '120-140','>140')
sizeorder<-c('Half-size or compact','Small (less than 17.6 cubic feet)','Medium (17.6 to 22.5 cubic fee
size1$size<- factor(size1$size, levels=sizeorder)
size1$income <- factor(size1$income, levels=incomeorder)
```

```
ggplot (data=size1, aes(x=income,y=unit,fill=size))+geom_bar(stat="identity")
```

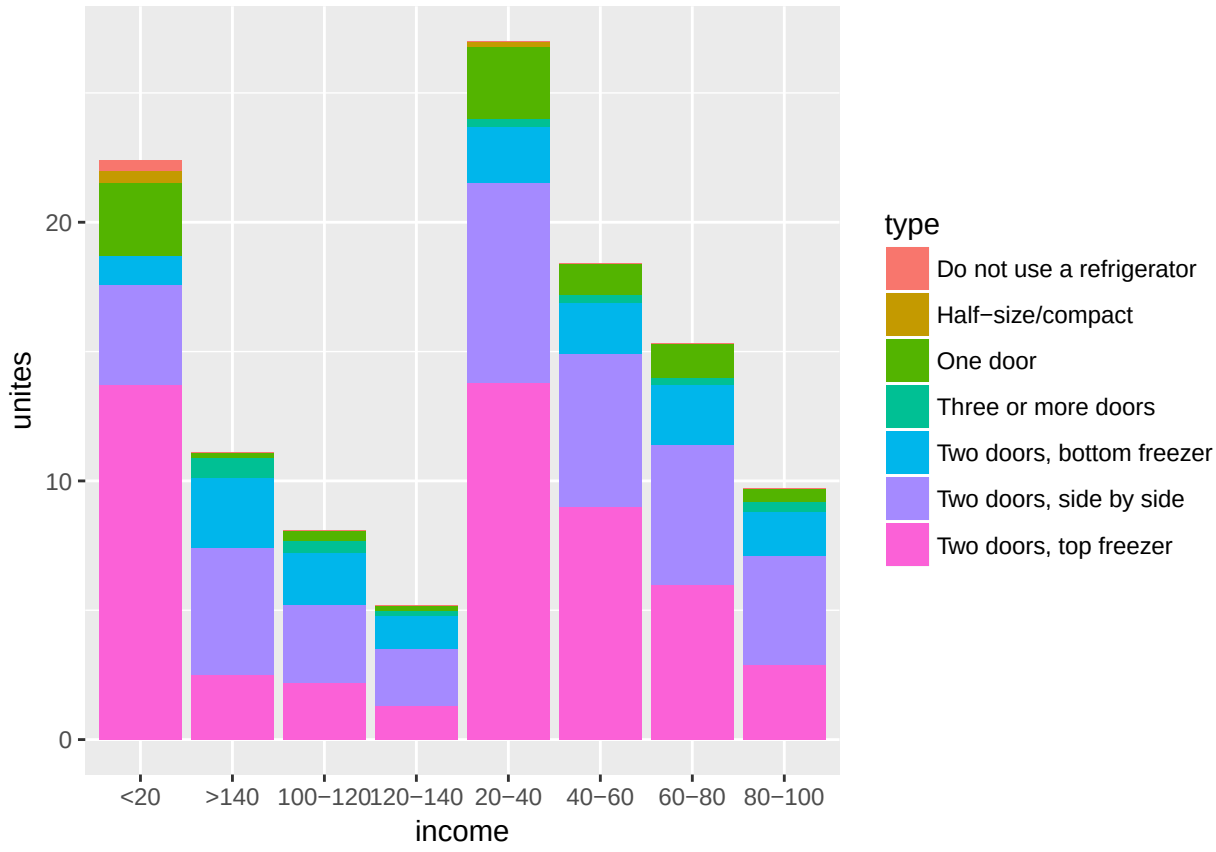



```
type <- read.csv("type.csv")
```

```
boxplot(type,xlab="income",ylab="types")
```

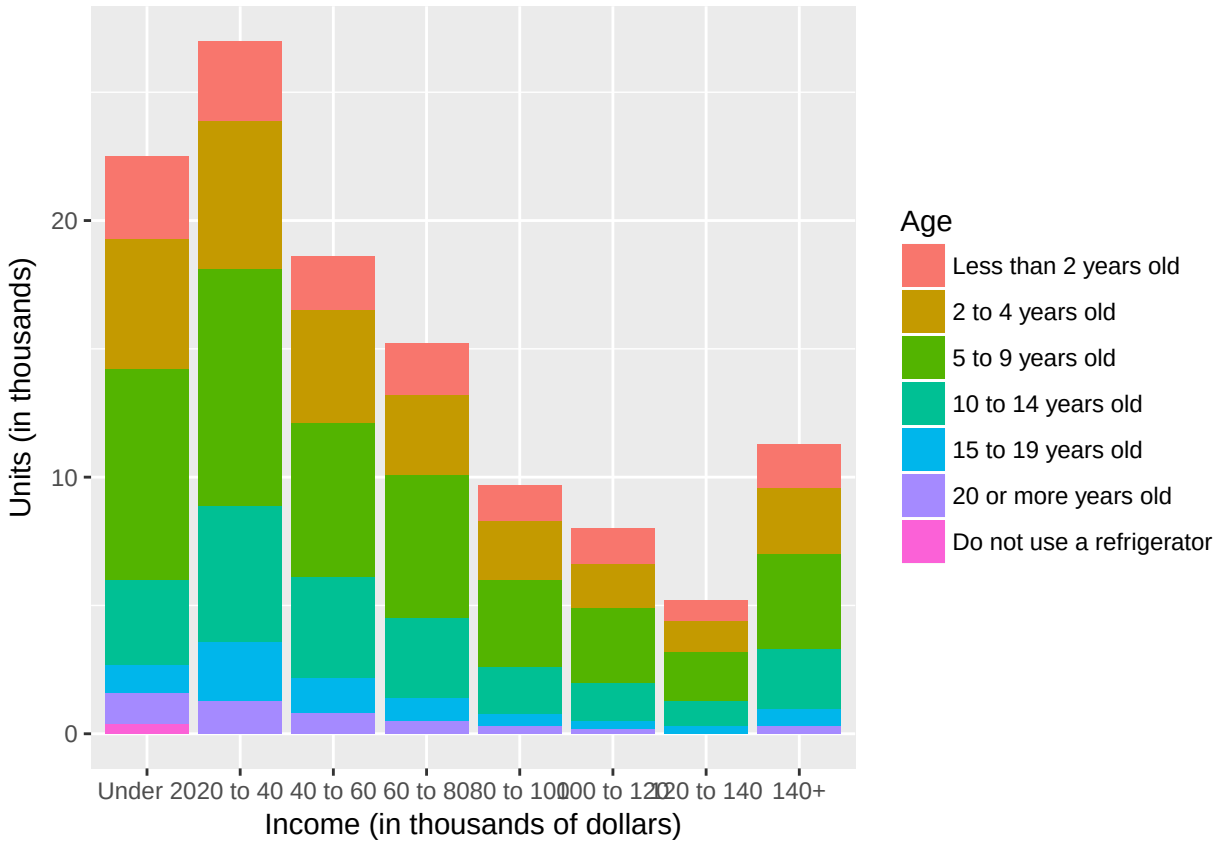



```
type1 <- read.csv("type1.csv")
ggplot(data=type1, aes(x=income, y=unites, fill=type)) + geom_bar(stat="identity")
```




```
age1 <- read.csv("rg age1.csv", header = TRUE)
age <- read.csv("rg age.csv", header = TRUE)

incomeorder <- c('Under 20', '20 to 40', '40 to 60', '60 to 80', '80 to 100', '100 to 120', '120 to 140
ageorder <- c('Less than 2 years old', '2 to 4 years old', '5 to 9 years old', '10 to 14 years old', '15

age1$Age <- factor(age1$Age, levels = ageorder)
age1$Income <- factor(age1$Income, levels=incomeorder)
```

```
ggplot(data=age1, aes(x=Income, y=Units, fill=Age)) +
  geom_bar(stat="identity") +
  labs(x= "Income (in thousands of dollars)", y = "Units (in thousands)")
```



Looking at this plot, you can see that the largest number of people with new refrigerators is in the lower income brackets. In terms of marketing, this means that people who are lower income are buying refrigerators more often, possibly because they have less money to spend and so buy lower quality refrigerators when they do buy them.