Refrigerator Report

Xuetong Ma,Yitong Zhang,Shuai Dong, Ang Li, Yuting Ma

2017/10/9

We would like to explore the data to find useful information that related to planning and marketing refrigerators.
Therefore, we want to find some variables that may influence the number of fridges. To visualize the information, we first summarize and determine the variables to analyze after cleaning and recoding.

1.Potential variables

1.1Region

First, we summarize a table by region to see which part have more refrigerators on average.

```
## # A tibble: 4 x 4
##   `d$REGIONC` AvgNumFrige MostAtAgeLevel              MostChosenFrigeType
##         <chr>       <dbl>          <chr>                            <chr>
## 1     Midwest    1.449887   5 to 9 years   Two doors, freezer above frige
## 2   Northeast    1.416877   5 to 9 years   Two doors, freezer above frige
## 3       South    1.364179   5 to 9 years Two doors, freezer next to frige
## 4        West    1.378135   5 to 9 years Two doors, freezer next to frige
```

From the table we can see, Midwest and Northeast have higher average number of fridges than South and West, though the difference is not too large.

1.2Income level

Another variable might be people's income level, here's a table that summarize the average number of fridges of different income levels. Here '1 =Less than 20,000 dollars''2=20,000 - 39,999 dollars' etc."8= $140,000 or more"

```
## # A tibble: 8 x 4
##   `d$MONEYPY` AvgNumFrige MostAtAgeLevel              MostChosenFrigeType
##         <int>       <dbl>          <chr>                            <chr>
## 1           1    1.158055   5 to 9 years   Two doors, freezer above frige
## 2           2    1.274525   5 to 9 years   Two doors, freezer above frige
## 3           3    1.316964   5 to 9 years   Two doors, freezer above frige
## 4           4    1.447368   5 to 9 years Two doors, freezer next to frige
## 5           5    1.519192   5 to 9 years Two doors, freezer next to frige
## 6           6    1.529551   5 to 9 years Two doors, freezer next to frige
## 7           7    1.613718   5 to 9 years Two doors, freezer next to frige
## 8           8    1.804795   5 to 9 years Two doors, freezer next to frige
```

Then from the table, we can see that income Level is significant in determining the number of frige purchased. The average numeber of fridges increases with higher income level. However, through data summary, we found that some income levels share a similar purchase patern. Therefore, we decided to combine some of the income levels to better demostrate the effect of income.

Create a new variable for combined Income Level

So we recode the data so that 1-3 = "$59,999 and Less", 4-6 = 60,000 to 119,999 dollars, and 7-8 = 120,000 dollars and more.

The new table shown as follows:

```
## # A tibble: 3 x 4
##       `d$IncomeLevel` AvgNumFrige MostAtAgeLevel
##                 <chr>       <dbl>          <chr>
## 1   $120,000 and More    1.743322   5 to 9 years
## 2    $59,999 and Less    1.250079   5 to 9 years
## 3 $60,000 to $119,999    1.489273   5 to 9 years
## # ... with 1 more variables: MostChosenFrigeType <chr>
```

This time the differences among average number of fridges become more obvious. Higher income level has higher average number of fridge per household.

1.3 Number of child & Fridges

Is the number of child in a household a potential variables that may influence the number of fridges? Since with more children, there seems to need more space to storage snacks and dessert.

```
## # A tibble: 10 x 4
##    `d$NUMCHILD` AvgNumFrige    MostAtAgeLevel
##           <int>       <dbl>             <chr>
##  1            0    1.375390      5 to 9 years
##  2            1    1.382093      5 to 9 years
##  3            2    1.463746      5 to 9 years
##  4            3    1.555556      5 to 9 years
##  5            4    1.363636      5 to 9 years
##  6            5    1.526316      5 to 9 years
##  7            6    1.285714    10 to 14 years
##  8            7    1.000000 Less than 2 years
##  9            9    1.000000      5 to 9 years
## 10           10    2.000000 20 years or older
## # ... with 1 more variables: MostChosenType <chr>
```

However, the average number of fridges doesn't show a tendency with more children in a household, which means that the number of children in a household is not significant enough to determine the number of fridges that the household choose to use.

1.4 Conclusion about the mode

From all these tables above, we can see that most of the fridges are at age level 5 to 9 regardless of region, income, and number of child; Most of the fridges owned are two doors, with different design of the locations of freezer.Further analysis can focus on the distribution of age level of friges, and people's choice of frige type. Therefore, we conclude that the average usage life of a frige is around 5 to 9 years. For design and production, the mainly focus should be two doors fridges.

2.Plots and Findings

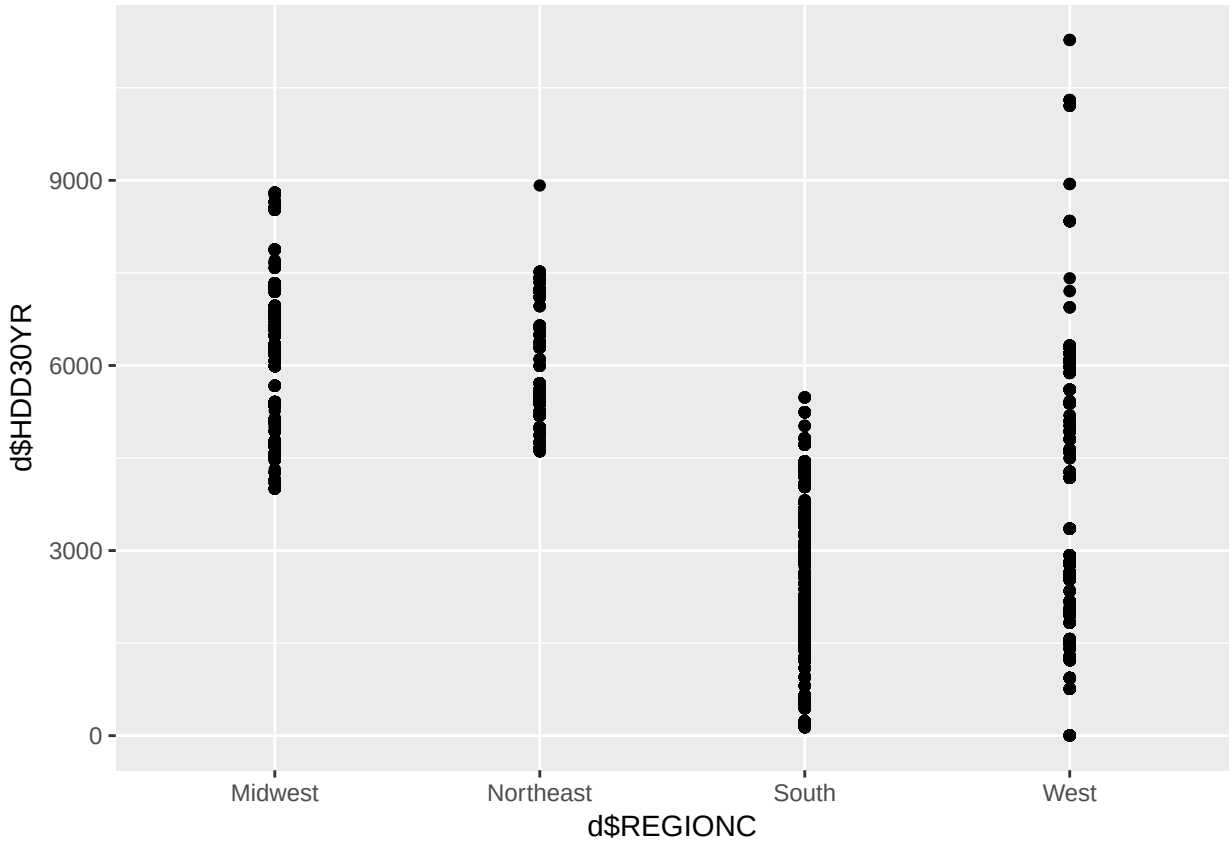
To visualize the data, we draw different plots for comparision.

2.1Region

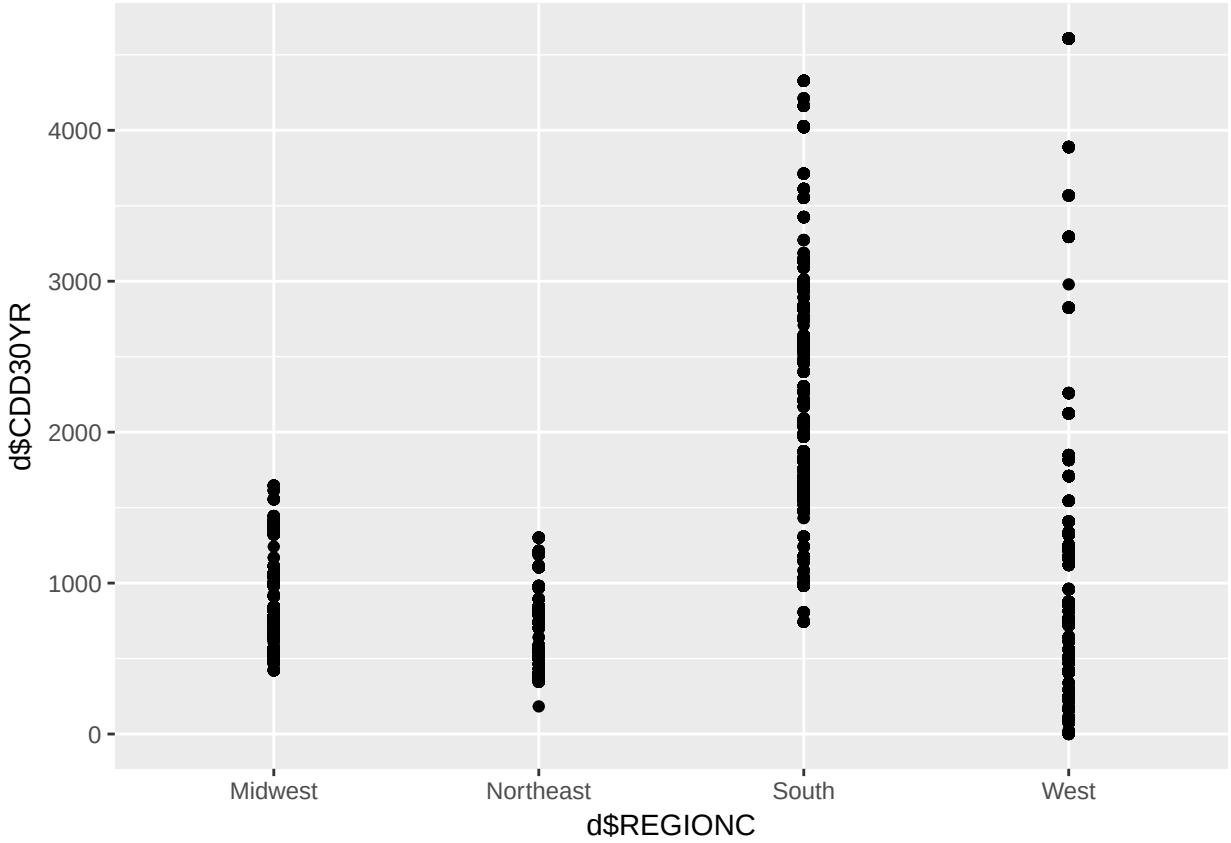
By summarise Number of Frige by Region, we found that Midwest and Northeast regions seem to have a larger average of number of frige per household. To understand the reason behind this pattern and then to confirm our finding, we make hypothesis that the it is the temperature difference in different regions that causes the pattern.

First, we want to analyze the relationship between Heating Days / Cooling Days with Region.

2.1.2 Plot of Heating Days & Region

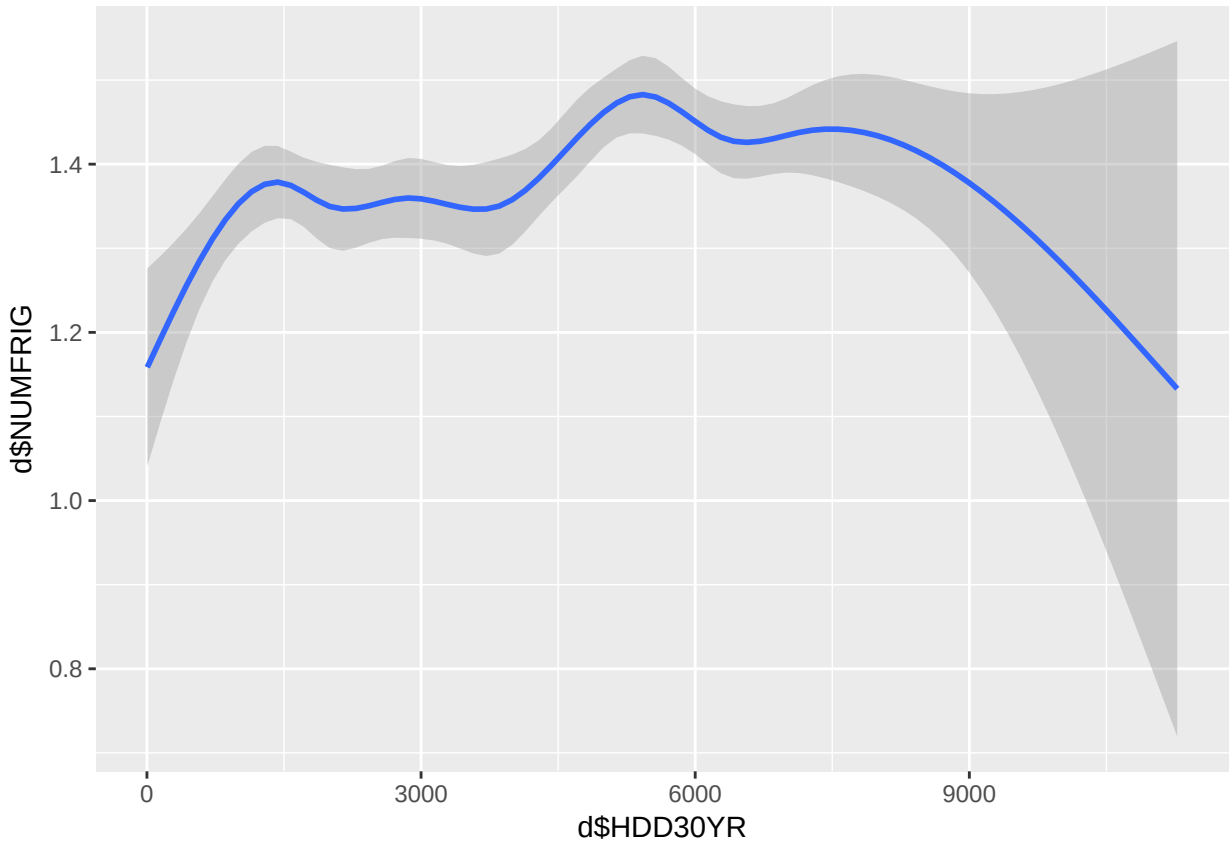


2.1.3 Plot of Cooling Days & Region

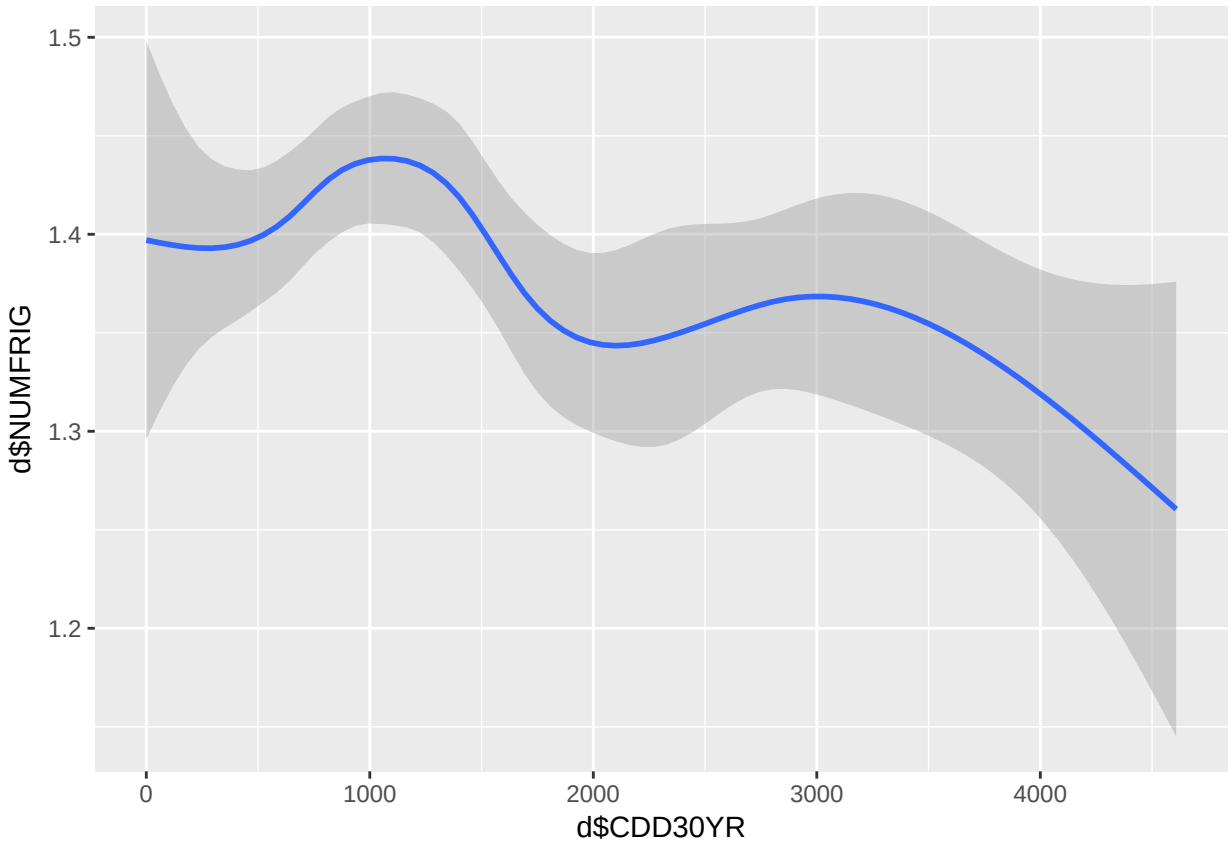


2.1.4 Number of Fridge by Heating Days / Cooling Days: Fitted Line Graph

```
## `geom_smooth()` using method = 'gam'
```




```
## `geom_smooth()` using method = 'gam'
```



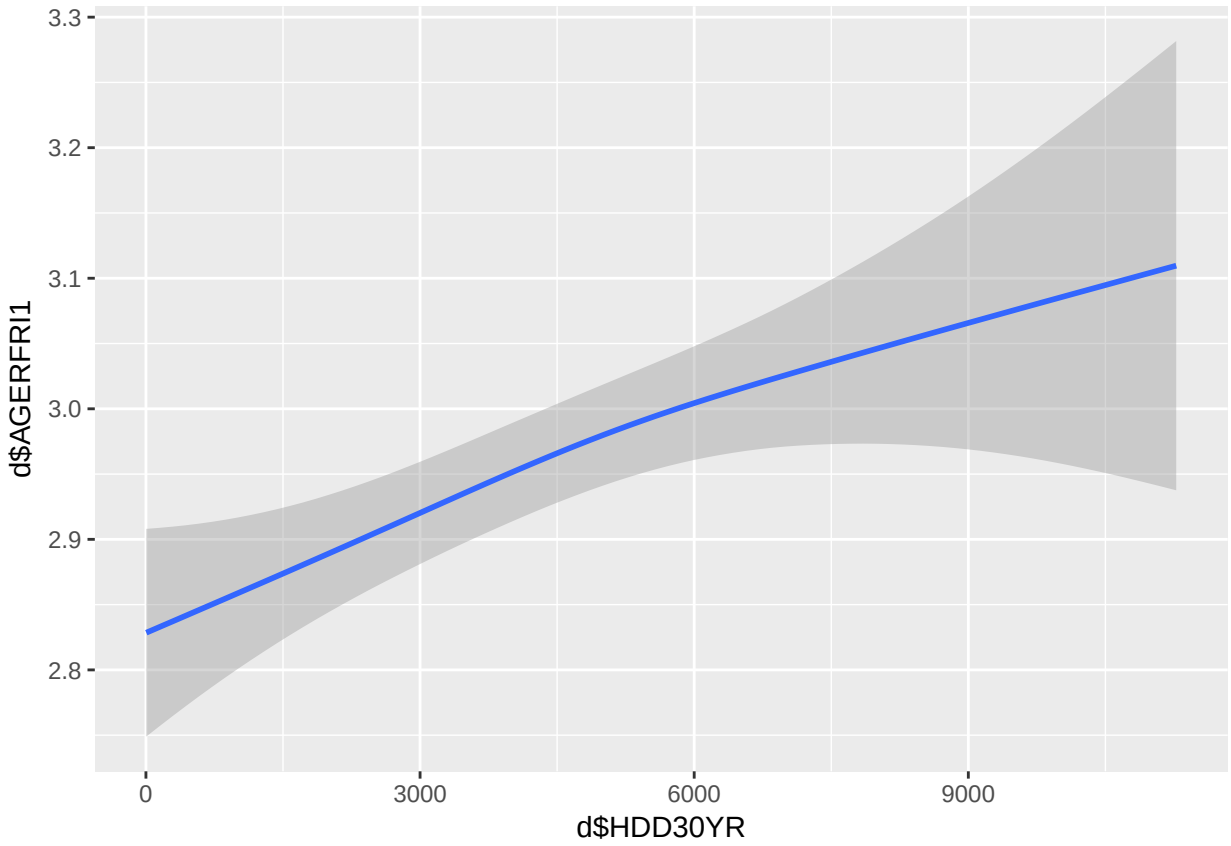
From the last two graphs, it is shown that the average of around 4500 to 6000 Heating Days in 30 years result in the highest number of fridge purchased, whereas the average of around 1000 Cooling Days in the 30 years result in the highest number of frige purchased Checking the two previous charts, Midwest and Northeast have the Heating Days closest to 4500 to 6000, and the Cooling Days closest to 1000, and the distribution is more concentrated. This matches our previous finding that Midwest and Northeast have more fridges on average. Therefore, temperature difference might be one of the reasons for this circumstance.

2.1.5 Age of Fridge by Heating Days / Cooling Days:

Then we want to find the tendency of age of fridge influenced by heating/cooling days. Graphs are as follows:

```
## `geom_smooth()` using method = 'gam'
```

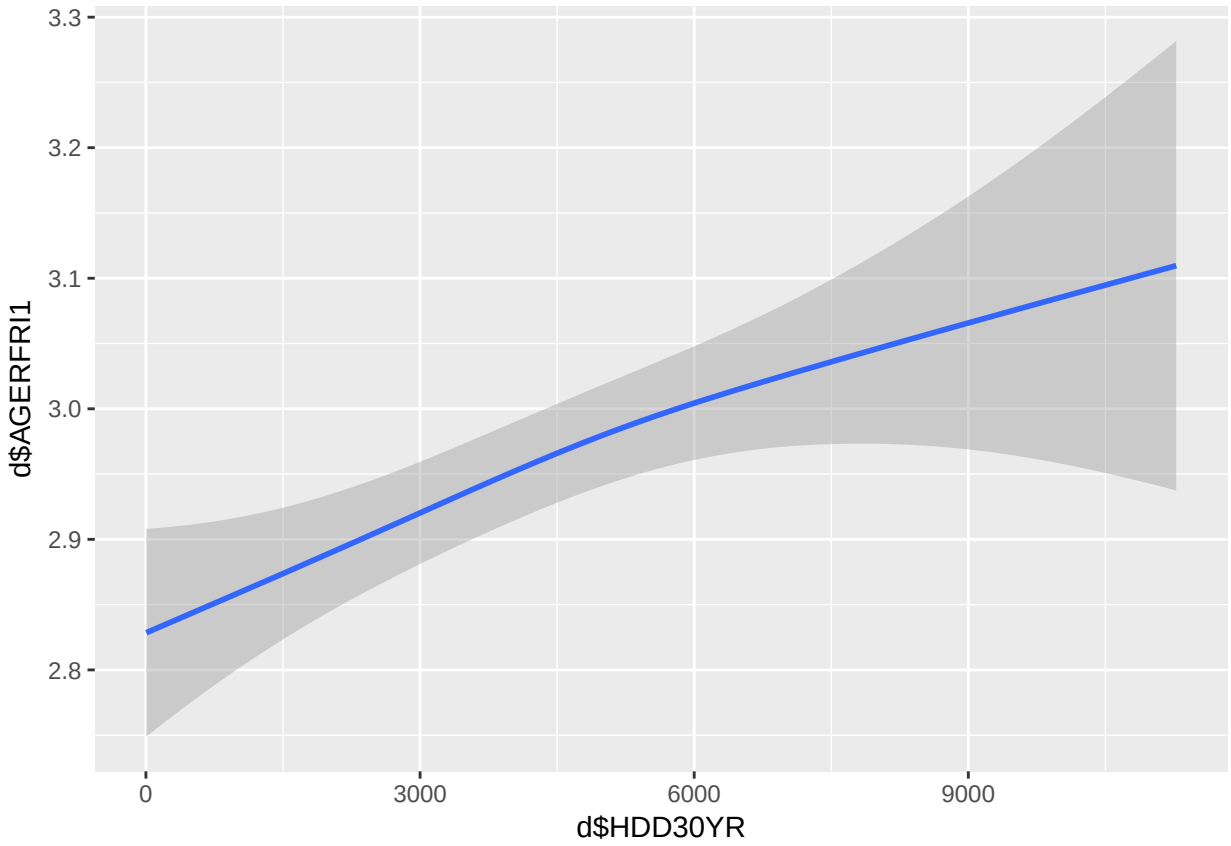
```
## Warning: Removed 31 rows containing non-finite values (stat_smooth).
```



Age level of the frige increases with the increase of Heating Days.

```
## `geom_smooth()` using method = 'gam'
```

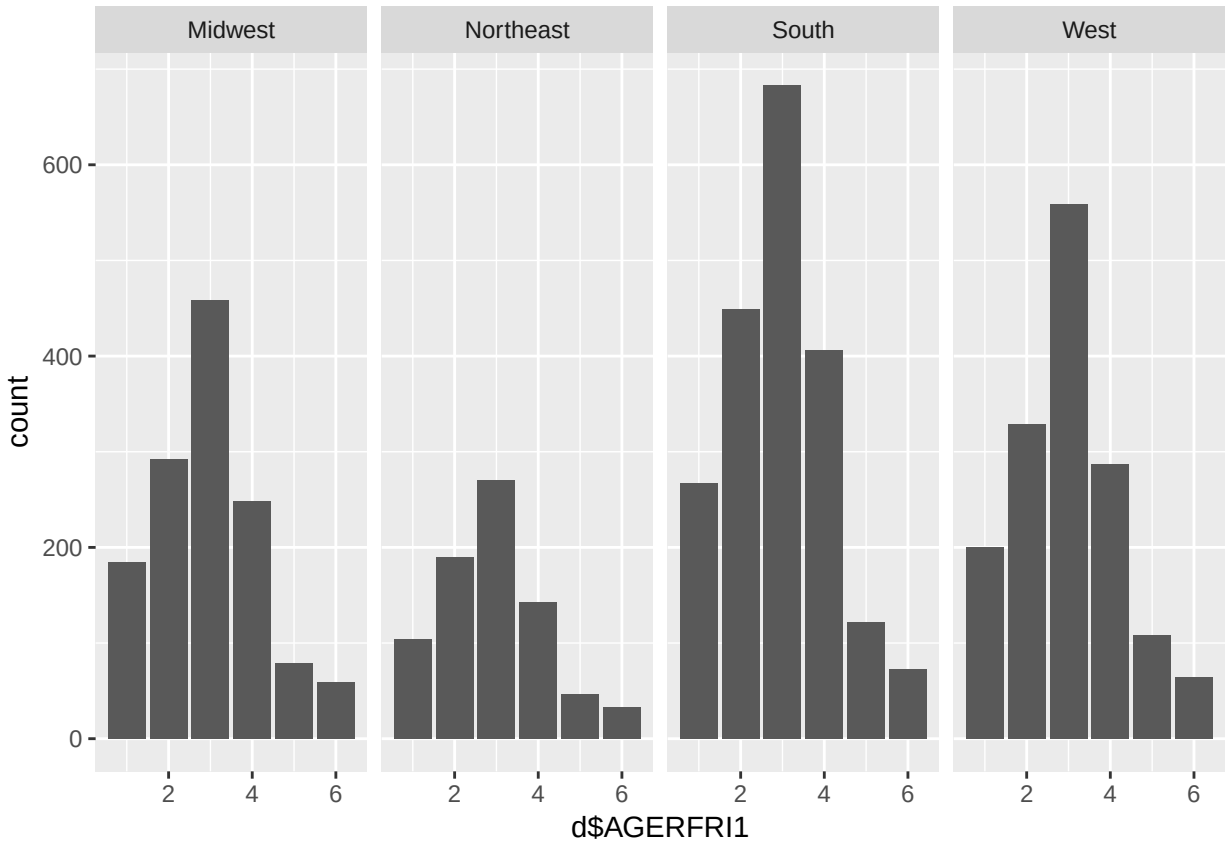
```
## Warning: Removed 31 rows containing non-finite values (stat_smooth).
```



Age level of the frige decreases with the increase of Cooling Days. The two graphs have inverse tendency towards cooling/heating days, thus justifies our finding.

2.1.6 Compare age level for different region.

```
## Warning: Removed 31 rows containing non-finite values (stat_count).
```



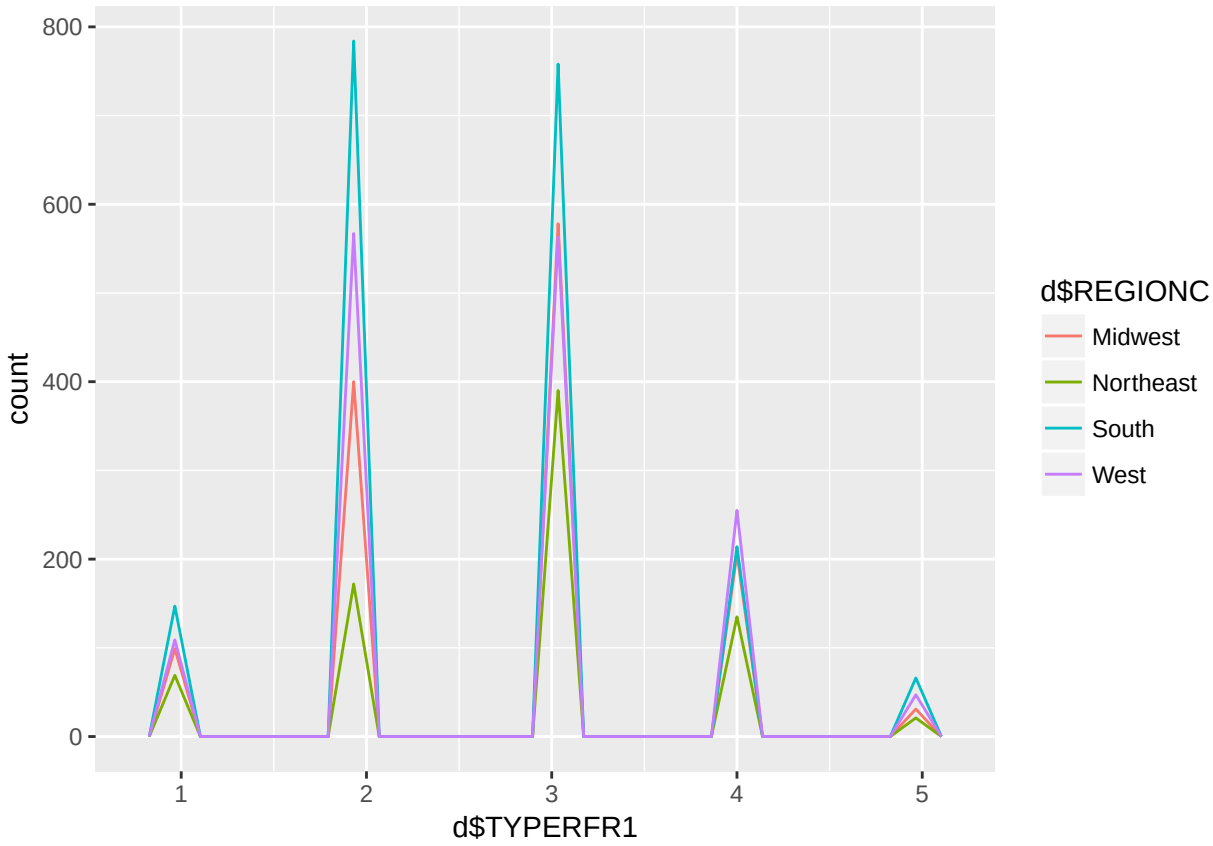
Regardless the number of observations, the patterns of these bar charts are similar: most of the ages are concentrated on "3", which means 5 to 9 years. However, if we compare the scale, the proportions of 4-6 in Midwest and Northeast are larger than that in South and West, which means higer age level in these regions. The conclusion in the same with our previous finding.

2.1.7 Type of Fridges by Region:

Here 1= "One door", 2 = "Two doors, freezer next to fridge", 3 = "Two doors, freezer above fridge", 4 = "Two doors, freezer below the fridge", 5 = "Three or more doors"

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 75 rows containing non-finite values (stat_bin).
```



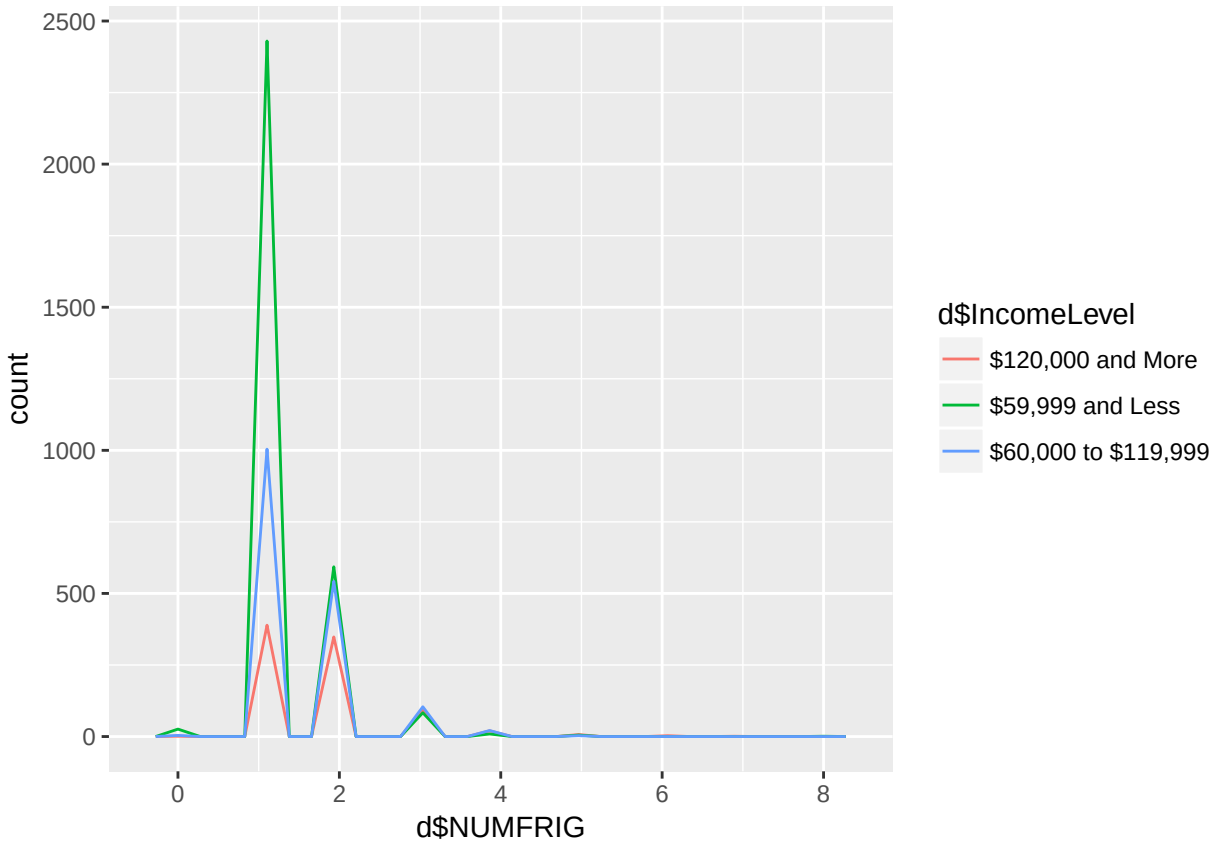
We can see that the data concentrates on 2, 3 and 4, means that people would like to choose two doors fridges. Among them, the most popular style is freezer next to and above fridge, which means that the company should design and produce more these two styles.

2.2 Income level

As we already found income level a significnat variable that may influence the number of fridges, here are some graphs to visaulize our findings.

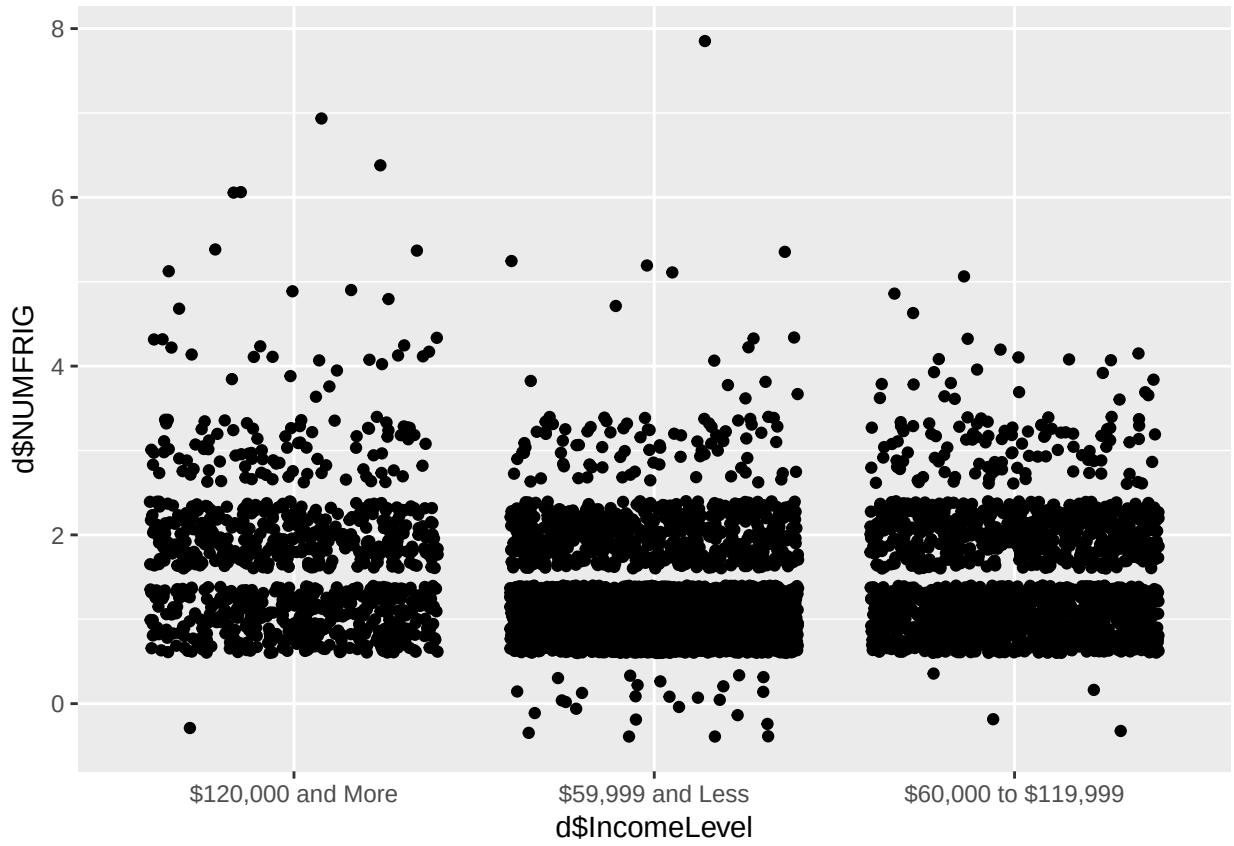
2.2.1 Number of Frige by Income Level

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



People who earn $59,999 or less are concentrated to have 1 fridges, with a small proportion have 2 fridges. The proportion of having 3 or more is quite small compared to the large observations. However, the proportion of having 2 fridges for people who earn 60,000 to 119,999 is larger, nearly half of the proportion of having 1 fridge. And for those who earn 120,000 dollars and more the two proportions are approximately the same.

Here's another scatter plot that shows the same circumstance.

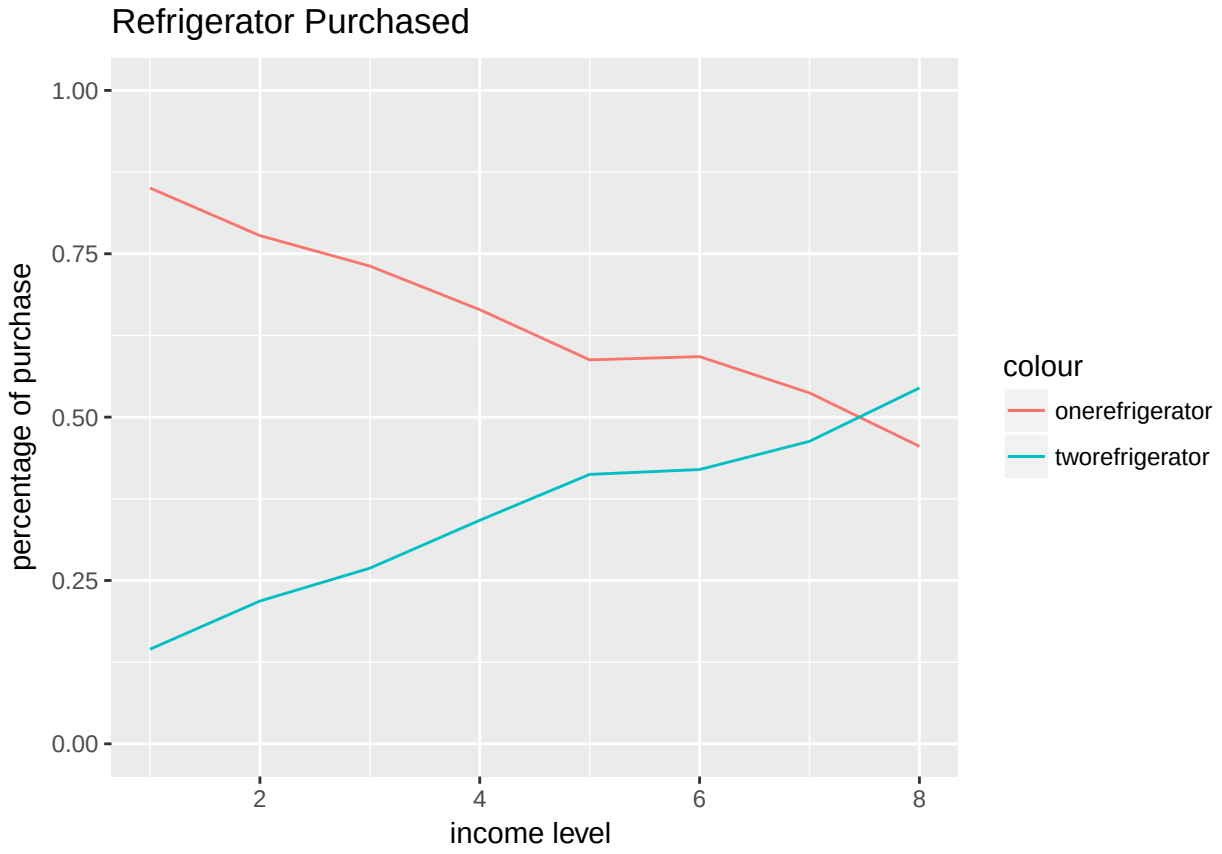


From the scatter plot, we can get the same conclusion. People who earn less are concentrated to have 0-2 fridges. With higher income level, the proportion of people having 2 or more fridges increases, means that higher income level tend to own more fridges than lower income level.

We divided people into eight income levels. Since we found that the population of he low income level is large while that of the high income level is small, and this would result in bias of this analysis. We changed the input into percent of population.

The line graph below shows the relation of income levels and number of refrigerators purchased.

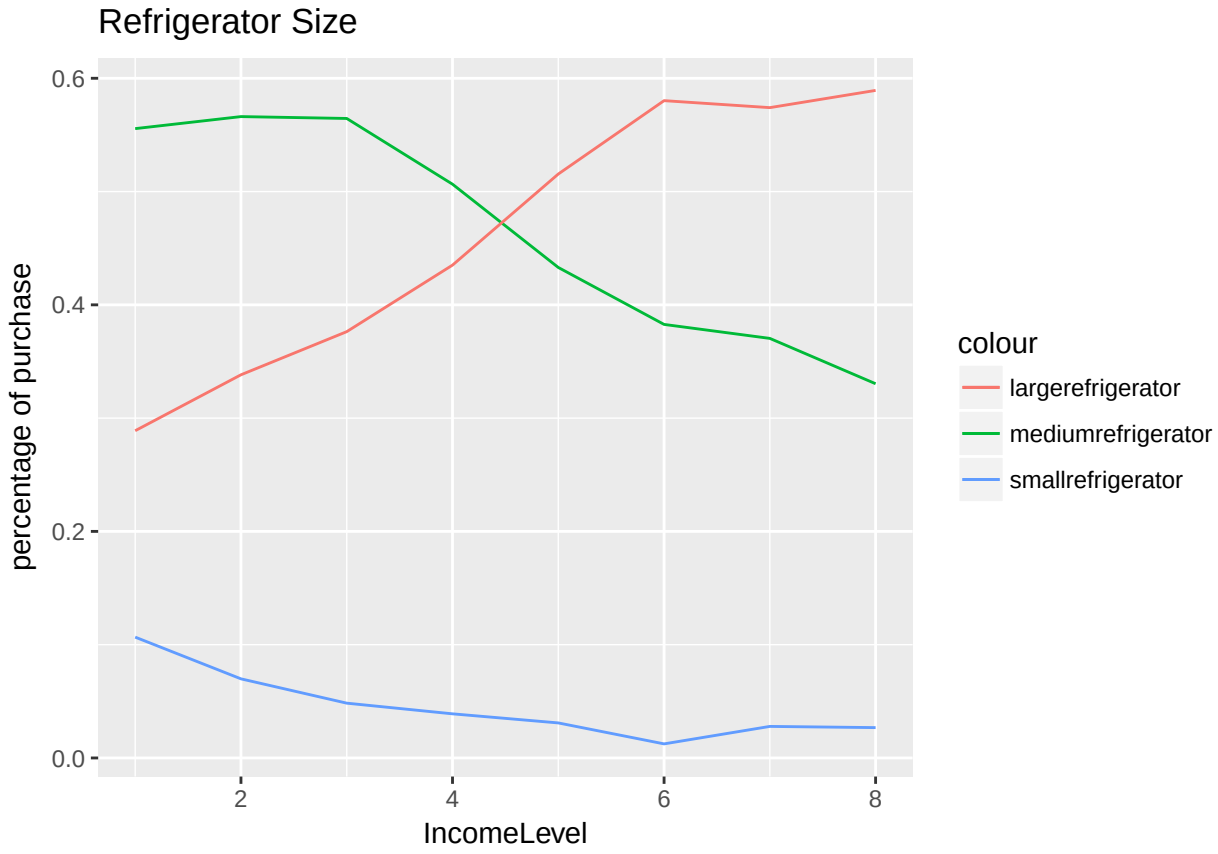
```
data1<-read.csv("income.csv")
ggplot(data = data1, aes(x = incomelevel)) +
  geom_line(aes(y = onef, colour = "onerefrigerator")) +
  geom_line(aes(y = twof, colour = "tworefrigerator")) +
  xlab("income level ") +
  scale_y_continuous("percentage of purchase",limit=c(0,1)) +
  labs(title="Refrigerator Purchased")
```



we use qqplot to draw the line for different income level of people who choose to buy one refrigerator or two refrigerators. we found most of people would choose to buy one refrigerator. But with income level increases, it is more likely that people will buy the second refrigerator. So it's a good idea to target at the high-income customers when selling the second refrigerator.

2.2.2 Size of Frige by Income Level

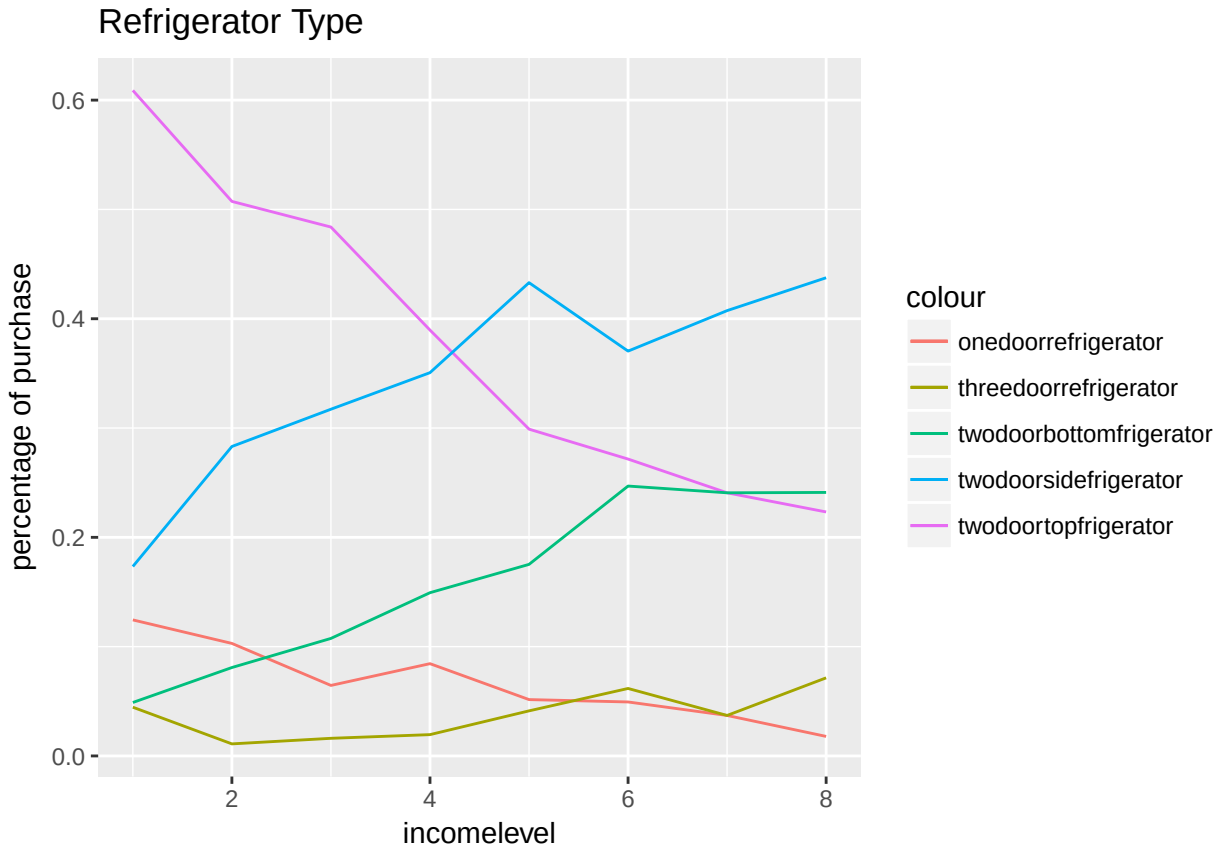
```
ggplot(data = data1, aes(x = incomelevel)) +
  geom_line(aes(y = small, colour = "smallrefrigerator")) +
  geom_line(aes(y = Mediumrate, colour = "mediumrefrigerator")) +
  geom_line(aes(y = large, colour = "largerefrigerator"))+
  xlab("IncomeLevel")+ylab("percentage of purchase")+ggtitle("Refrigerator Size")
```



we hypothesized that there is a relationship between the income level and size of refrigerator. According to the graph we concluded that the small refrigerator is the least preferred size. And medium-sized refrigerators are more popular in the low-income population, while large-sized refrigerators are more popular in the high-income population. When people have medium income, there's no significant difference in their preference between the size of medium and large. Our suggestion is that the manufacturer should produce more medium and large size refrigerator and less small refrigerator.

2.2.3 Type of Frige by Income Level

```
ggplot(data = data1, aes(x = incomelevel)) +
  geom_line(aes(y = onedoor, colour = "onedoorrefrigerator")) +
  geom_line(aes(y = twodoortop, colour = "twodoortopfrigerator")) +
  geom_line(aes(y = twodoorbottom, colour = "twodoorbottomfrigerator"))+
  geom_line(aes(y = twodoorside, colour = "twodoorsidefrigerator")) +
  geom_line(aes(y = threedoor, colour = "threedoorrefrigerator"))+
  ylab("percentage of purchase")+
  ggtitle("Refrigerator Type")
```

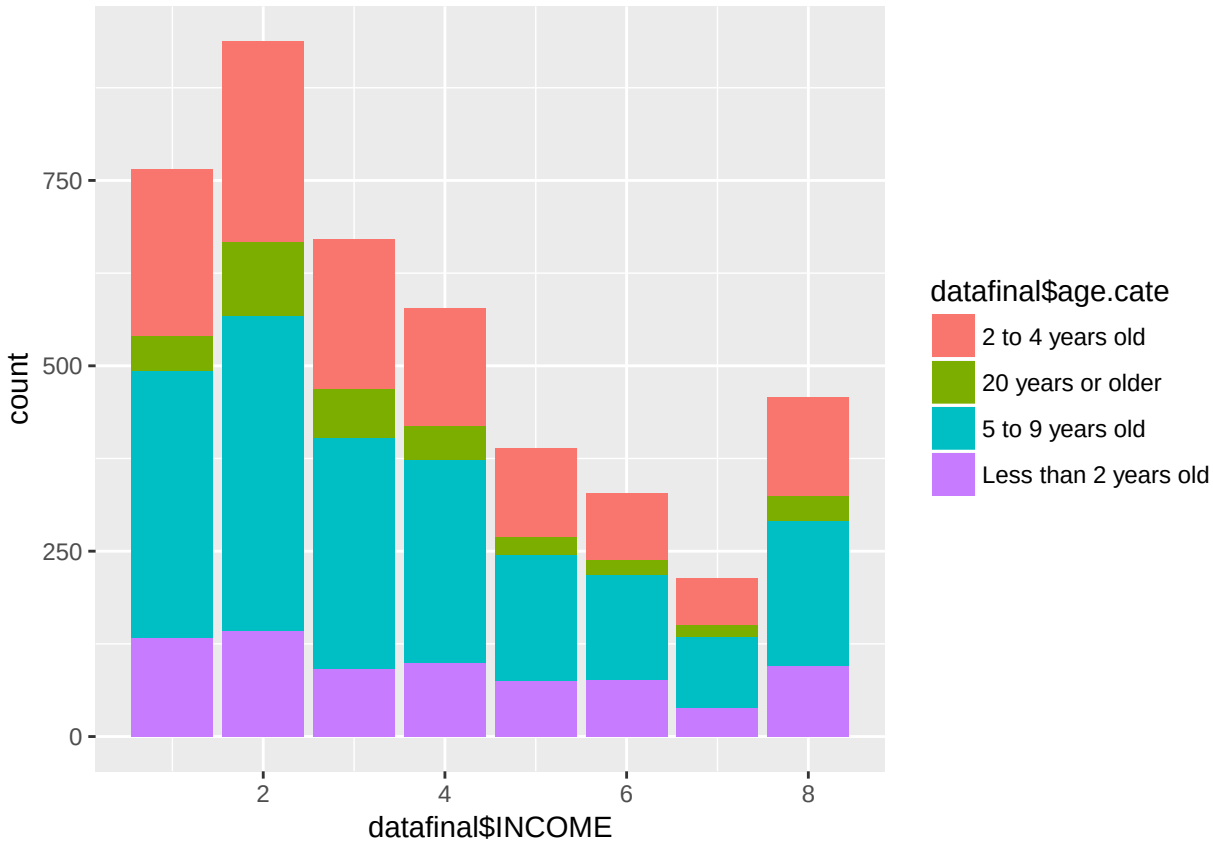


We made a graph to explore the relationship between refrigerator types and income levels of people. We included five types of refrigerator: The graph showed that few people would choose one door or three door refrigerator. Most people prefer the two door type. In the three subtypes of two door refrigerator, the top freezer refrigerator and the side by side refrigerator are more popular. What's more, with the income level increases, the preference of side by side refrigerator exceeds that of the top freezer type.

2.3 Age used of Refrigerator

2.3.1 Age used of Refrigerator For Different Income level

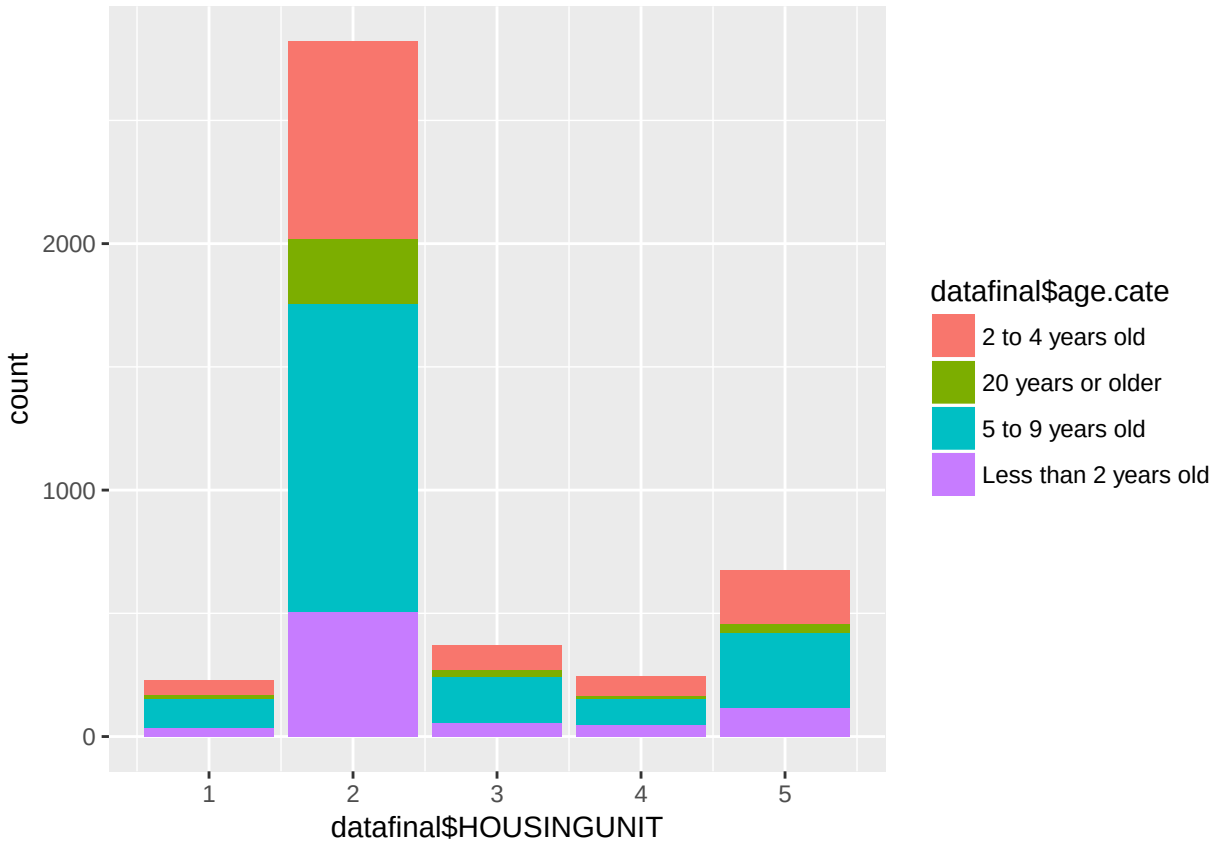
```
ggplot(data=datafinal)+
  geom_bar(mapping=aes(x=datafinal$INCOME, fill=datafinal$age.cate),  data= datafinal)
```



From the plot, we can find out that for every income level, people are more likely to use their refrigerator for 5-9 years and 2-4 years. And for the income level 8, which is that the income is higher than $140,000, people have a relatively high proportion of using a refrigerator for less than two years.

2.3.2 Age used of Refrigerator For Different Housing unit

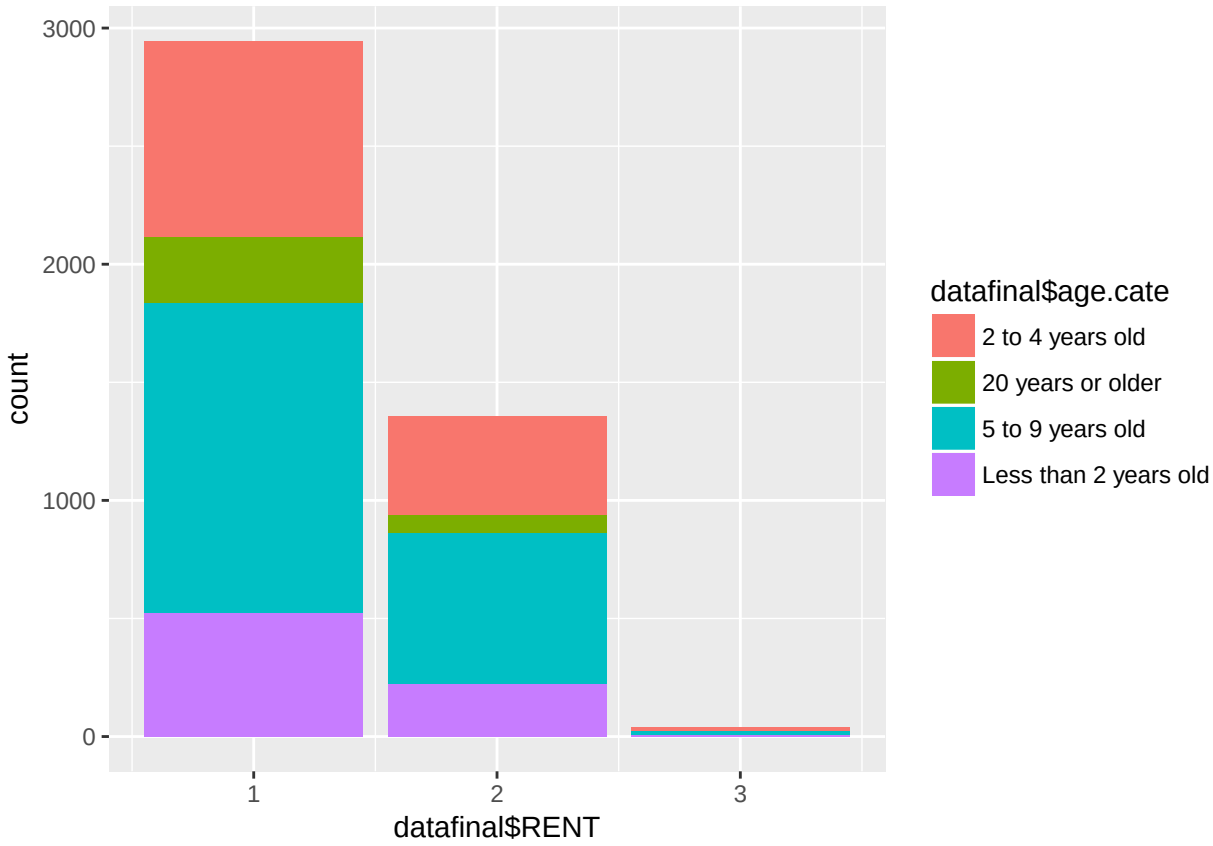
```
ggplot(data=datafinal)+
  geom_bar(mapping=aes(x=datafinal$HOUSINGUNIT, fill=datafinal$age.cate),  data= datafinal)
```



For this plot, 1—Mobile home; 2—Single-family detached house; 3—Single-family attached house; 4—Apartment in a building with 2 to 4 units; 5—Apartment in a building with 5 or more units.

2.3.3 Age used of Refrigerator For Different Rent type

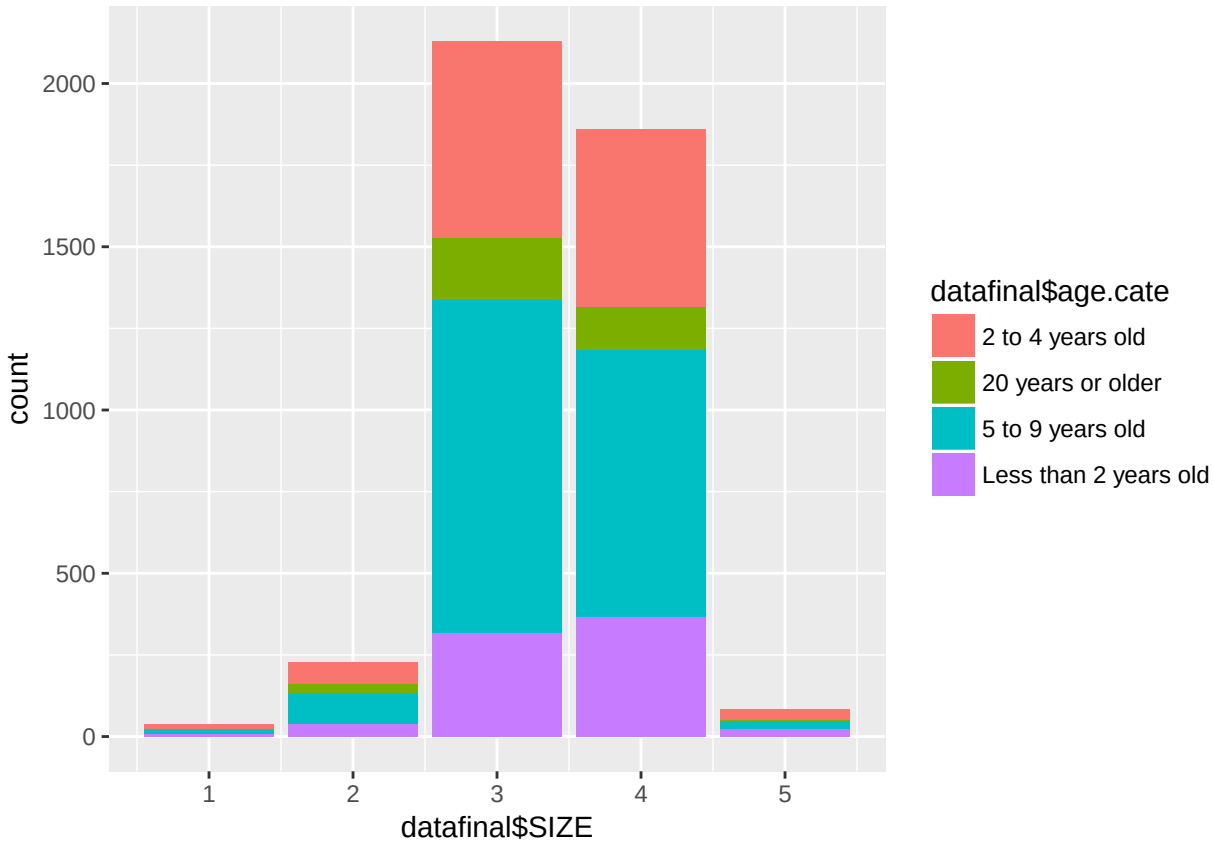
```
ggplot(data=datafinal)+
  geom_bar(mapping=aes(x=datafinal$RENT, fill=datafinal$age.cate),  data= datafinal)
```



For this plot, 1—Owned or being bought by someone in your household; 2—Rented; 3—Occupied without payment of rent.

2.3.4 Age used of Refrigerator For Different Refrigerator size

```
ggplot(data=datafinal)+
  geom_bar(mapping=aes(x=datafinal$SIZE, fill=datafinal$age.cate),  data= datafinal)
```



For this plot, 1—Half-size or compact; 2—Small (17.5 cubic feet or less); 3—Medium (17.6 to 22.5 cubic feet); 4—Large (22.6 to 29.5 cubic feet); 5—Very large (bigger than 29.5 cubic feet).

From 4 plots above, we can see that for every type of rent, every type of housing unit and every size of refrigerator, people are all more likely to use the regrigerator for 5-9 years or 2-4 years before change to a new one. So it seems that size, housing unit and rent type are not relevant factors for people's behavior on how long will they use for their refrigerators, and income is a relatively critial variable here.

3. Conclusion

From the analysis above, we find that people would like to choose two doors fridges at most time. Therefore, the company should produce and design more two doors fridges with large proportion of freezer next to or above fridge and small proportion of freezer below fridge to make more profit. Considering the income levels, high income customers would more likely to but a second refrigerator and they prefer two door side by side frige more than other types.While customers with lower income prefer medium sized refrigerators with two doors and the freezers on the top. Also, when distrubuting sales and refrigerators, the company should consider to distribute more to region in Midwest or Northeast, region with more heating days, or region with higher income level. Because from our findings, people in these regions are more likely to own more fridges.