MA615 EDA

Dongyuan Zhou, Jiawei Li, Chushu Wu

October 10, 2017

Assignment

This assignment is about refrigerators. Imagine that you work for a company the makes and sells refrigerators in the US - or a company that provides consulting on "white goods" – air conditioners, dishwashers, clothes dryers, drying cabinets, freezers, refrigerators, kitchen stoves, water heaters, washing machines, trash compactors, microwave ovens, and so on.

The statistics for refrigerators in the US and for relevant household demographics can be found on The U.S. Energy Information Administration (EIA) website.

https://www.eia.gov/consumption/residential/data/2015/

Your job is to explore the data and prepare a written report that describes and discusses your findings related to planning and marketing refrigerators. Use the EIA data, but don't hesitate to include other data such as refrigerator pricing.

Goals

Explore the data and give the refrigerator marketing plan for different region according to the EIA data.

In the markrting plan, the refrigerator's type, size and quantity trend are specified.

Finish a written report to illustrate our result.

Key questions

For different regions, we consider the following aspect:

Which type/size used most in this region?

What about the refrigerator's age? Does it have effect on the quantity of refrigerator?

For different years, we consider the follwing changes:

Whether there is changes on prefered type/size due to the years changes.

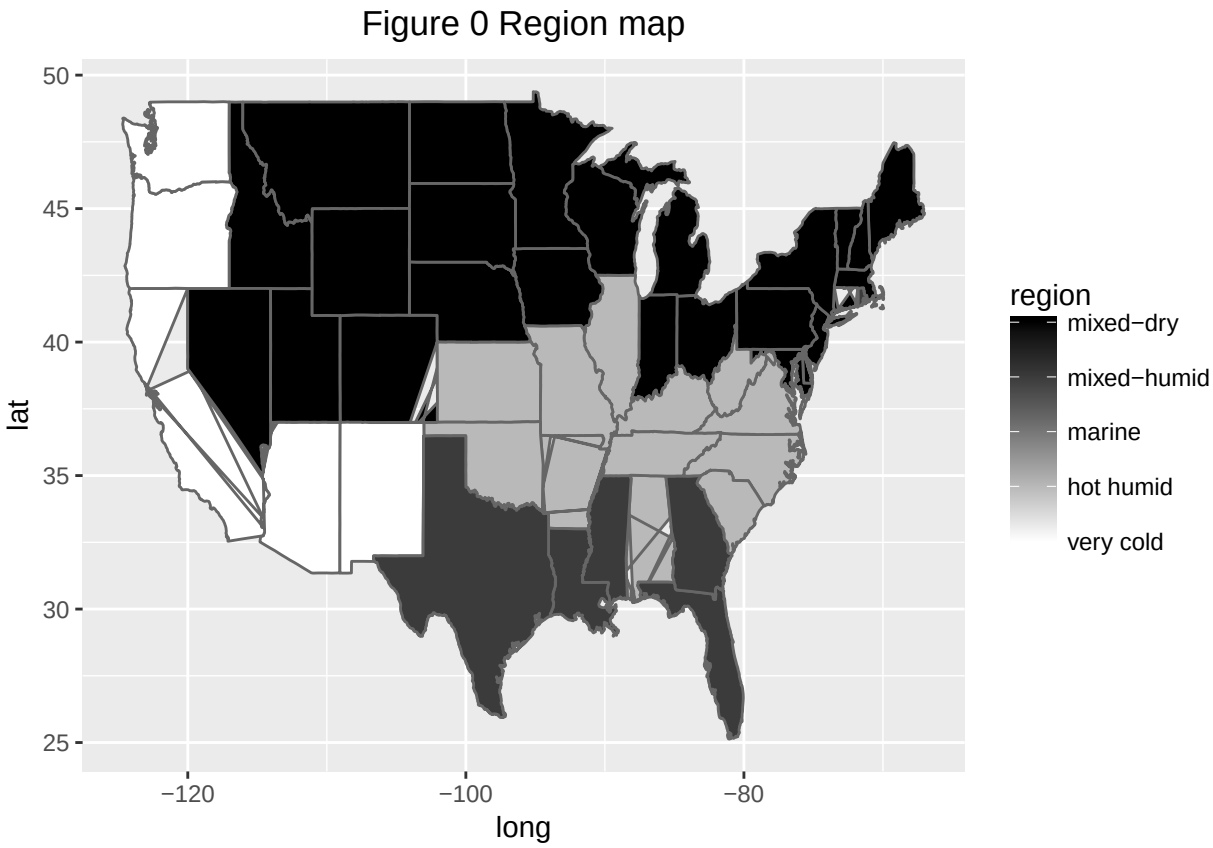
Solving the question above, we can give appropriate marketing plan for different region.

```
data2015 <- read.table("2015.csv", header = TRUE, sep = ",")
regionmap <- read.table("regionmap.csv", header = FALSE, sep = ",")

states <- map_data("state")
names(USArrests) <- tolower(names(USArrests))
regionmap$region <-tolower(rownames(USArrests))
regionmap <- merge(states, regionmap, by = "region")

plot2015type <- ggplot(data2015[1:25,], aes(x = region, y = refrigerator, fill = type))+
  geom_bar(stat = "identity")+
  labs(title = "Figure 1-1 2015 type - region")+
  theme(plot.title = element_text(hjust = 0.5))
```

```
plot2009type <- ggplot(data2015[26:50,], aes(x = region, y = refrigerator, fill = type))+
  geom_bar(stat = "identity")+
  labs(title = "Figure 1-2 2009 type - region")+
  theme(plot.title = element_text(hjust = 0.5))

plotregion <- ggplot(regionmap, aes(x = long, y = lat, group = group, fill =V2))+
  geom_polygon(colour = "grey40")+
  scale_fill_gradient(low="white",high="black",name = 'region',
           labels=c("very cold","hot humid","marine","mixed-humid","mixed-dry"))+
  labs(title = "Figure 0 Region map")+
  theme(plot.title = element_text(hjust = 0.5))
```

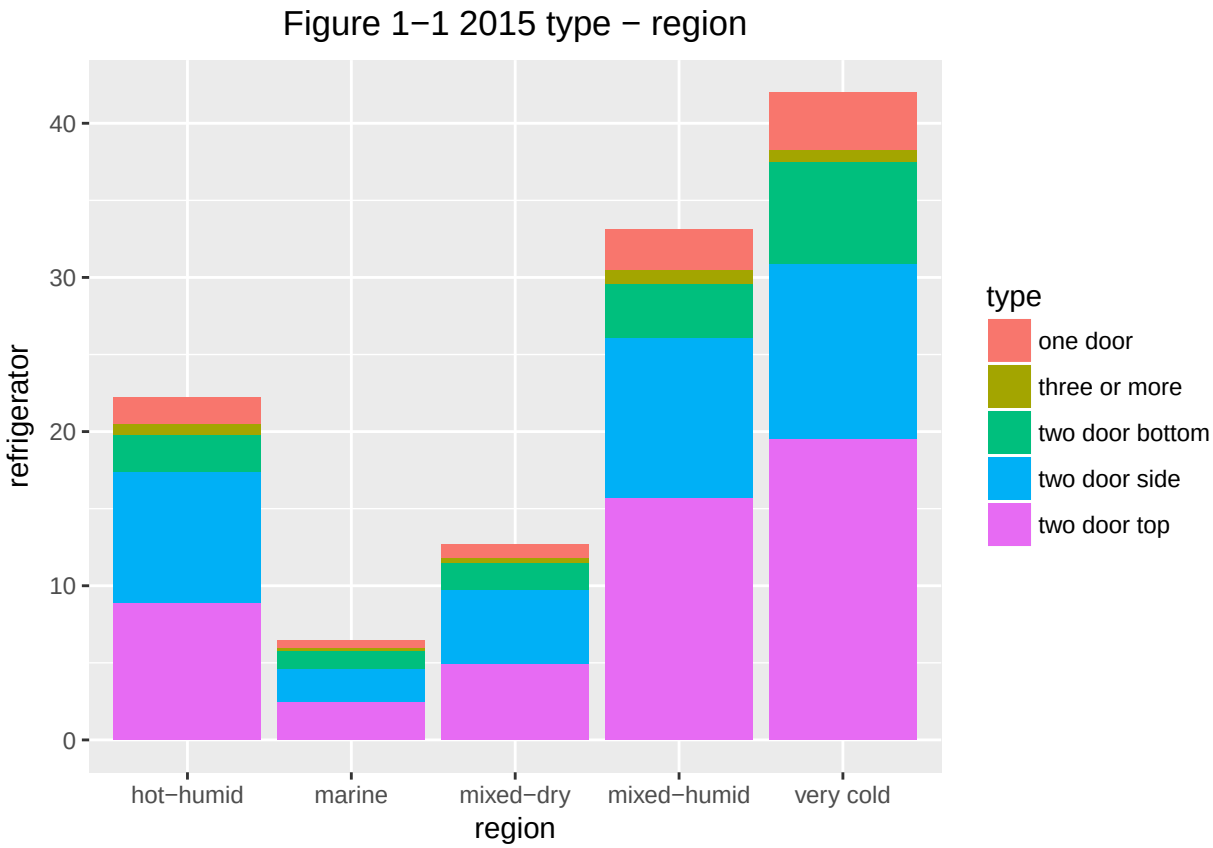
In our report, we divided the whole market in US into 5 regions according to the climate and state, which is shown below. In this way, it is more convinient to prepare goods for delivery and easier to understand.

```
plotregion
```



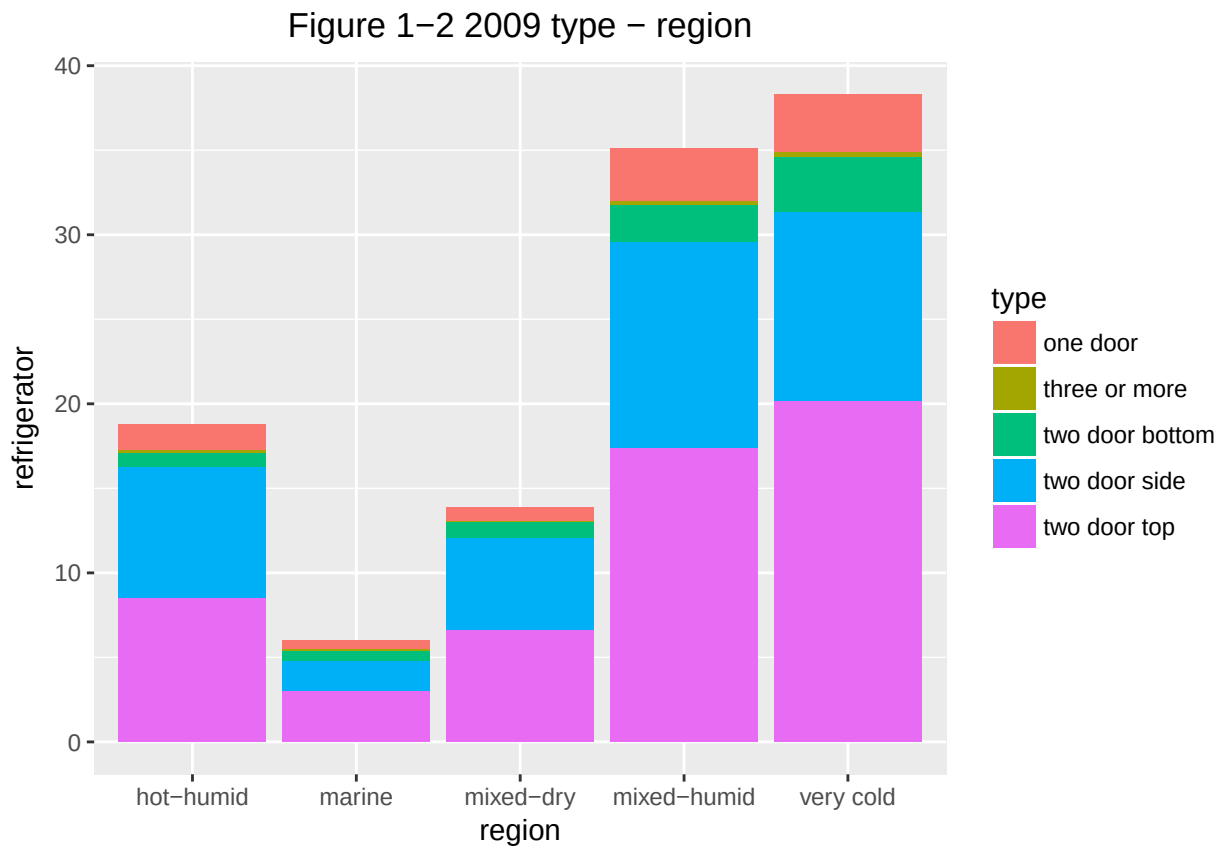
Plan related to Type

```
plot2015type
```

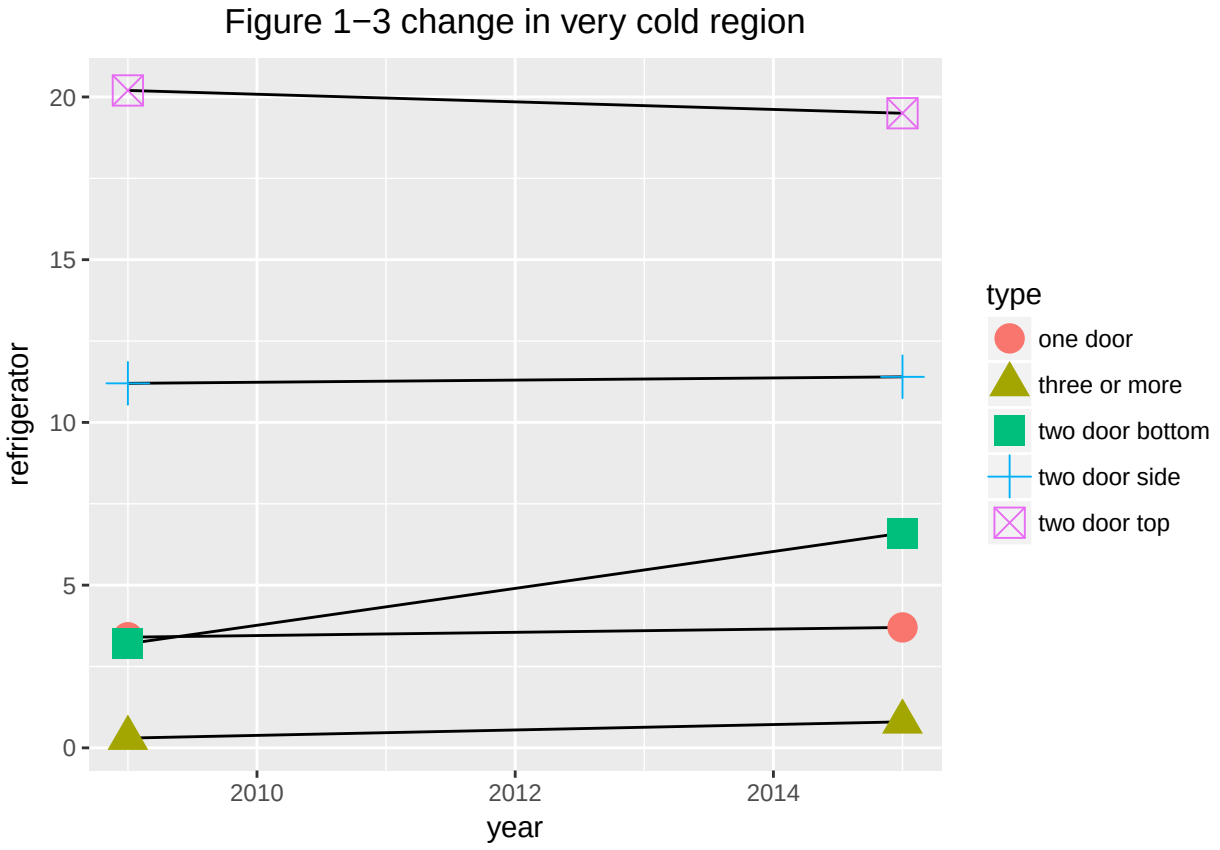


As can be seen from the picture above, “Two doors, top freezer” and “Two doors, side freezer” is used more in each region, compared to other types. What’s more, in the hot humid region, “Two doors, side freezer” has great percentage, compared to other regions. Therefore, we can get the suggestion that put more “Two door, side freezer” in this region’s market.

```
plot2009type
```

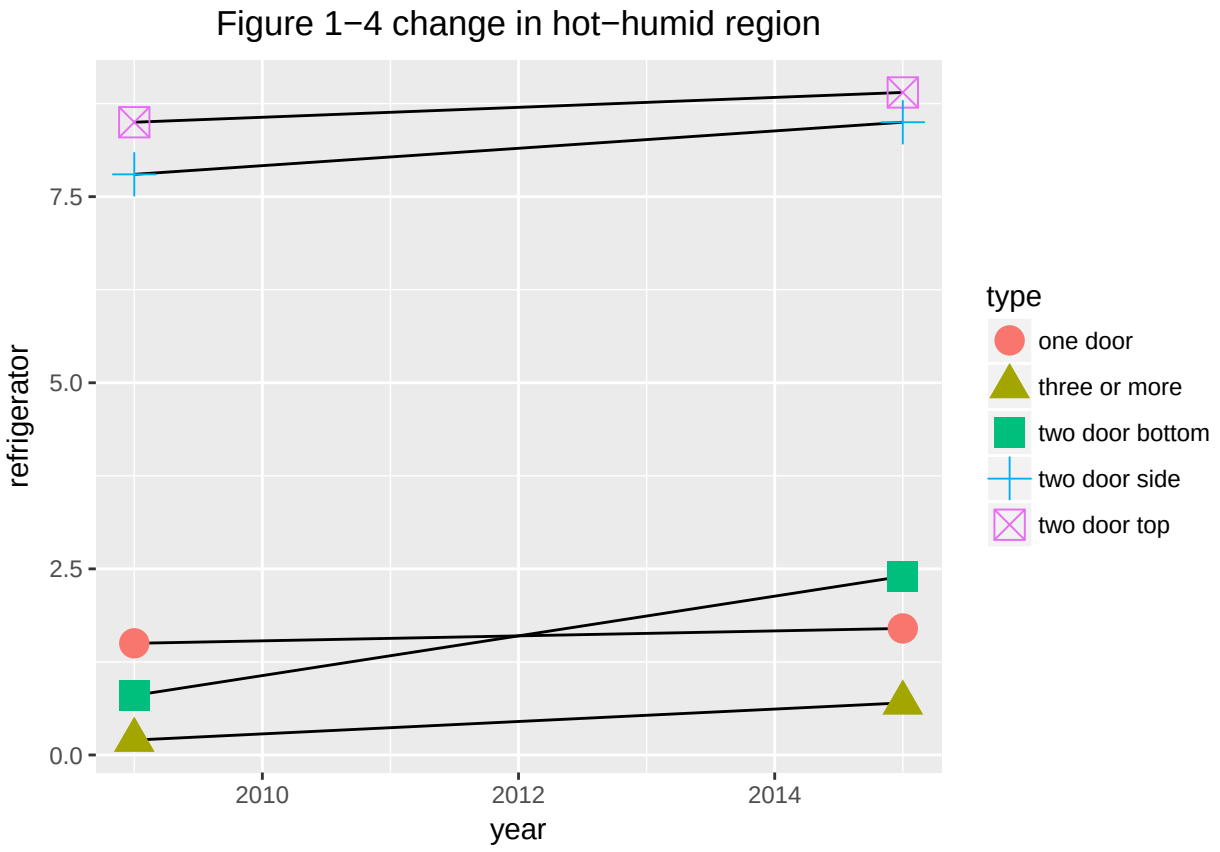



```
ggplot(data2015[data2015$regionnum == 5,], aes(x = year, y = refrigerator, group = type))+
  geom_line()+
  geom_point(aes(shape = type, colour = type),size = 5)+
  labs(title = "Figure 1-3 change in very cold region")+
  theme(plot.title = element_text(hjust = 0.5))
```



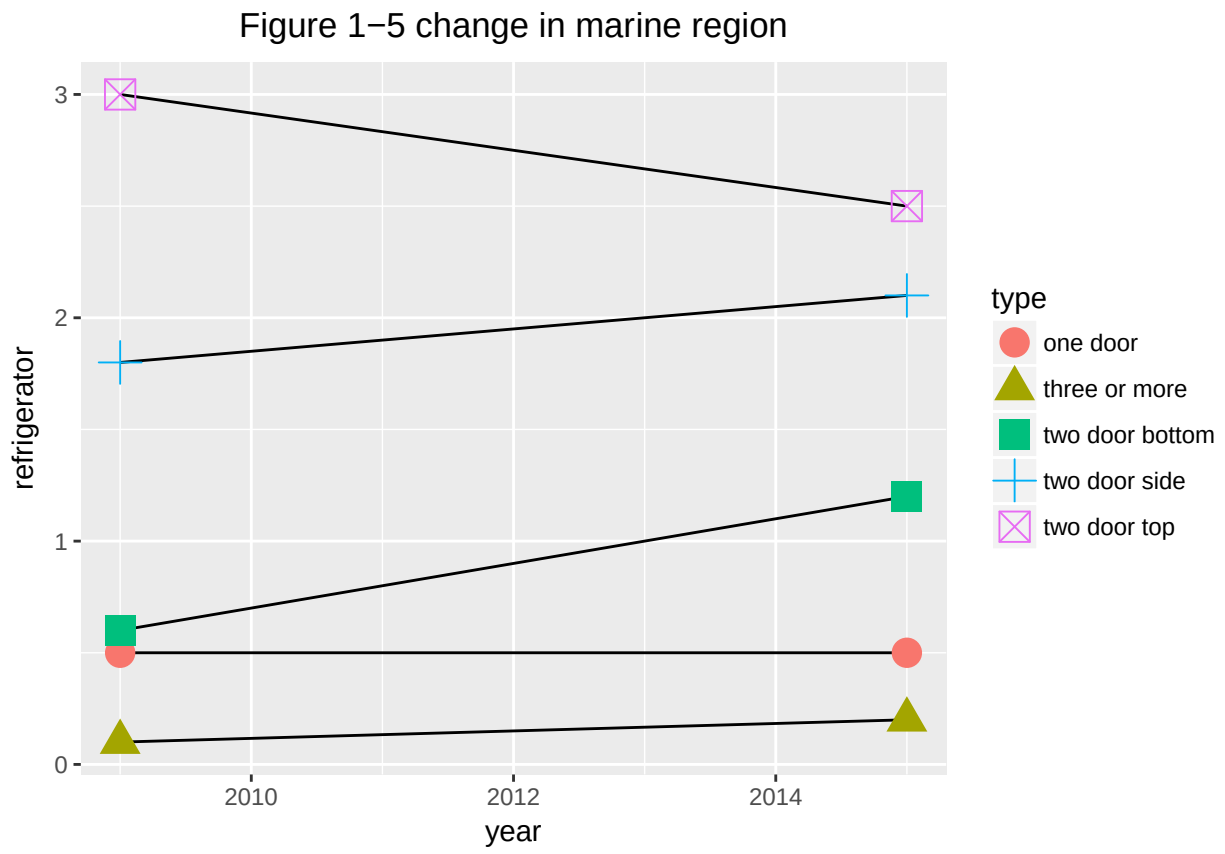
As shown from Figure 1-3 above, although there is a upper trend in "Two doors, bottom freezer", the most widely used type of refrigerators is "Two doors,top freezers". Therefore, we make the plan to suggest put more in very cold region's market.

```
ggplot(data2015[data2015$regionnum == 4,], aes(x = year, y = refrigerator, group = type))+
  geom_line()+
  geom_point(aes(shape = type, colour = type),size = 5)+
  labs(title = "Figure 1-4 change in hot-humid region")+
  theme(plot.title = element_text(hjust = 0.5))
```



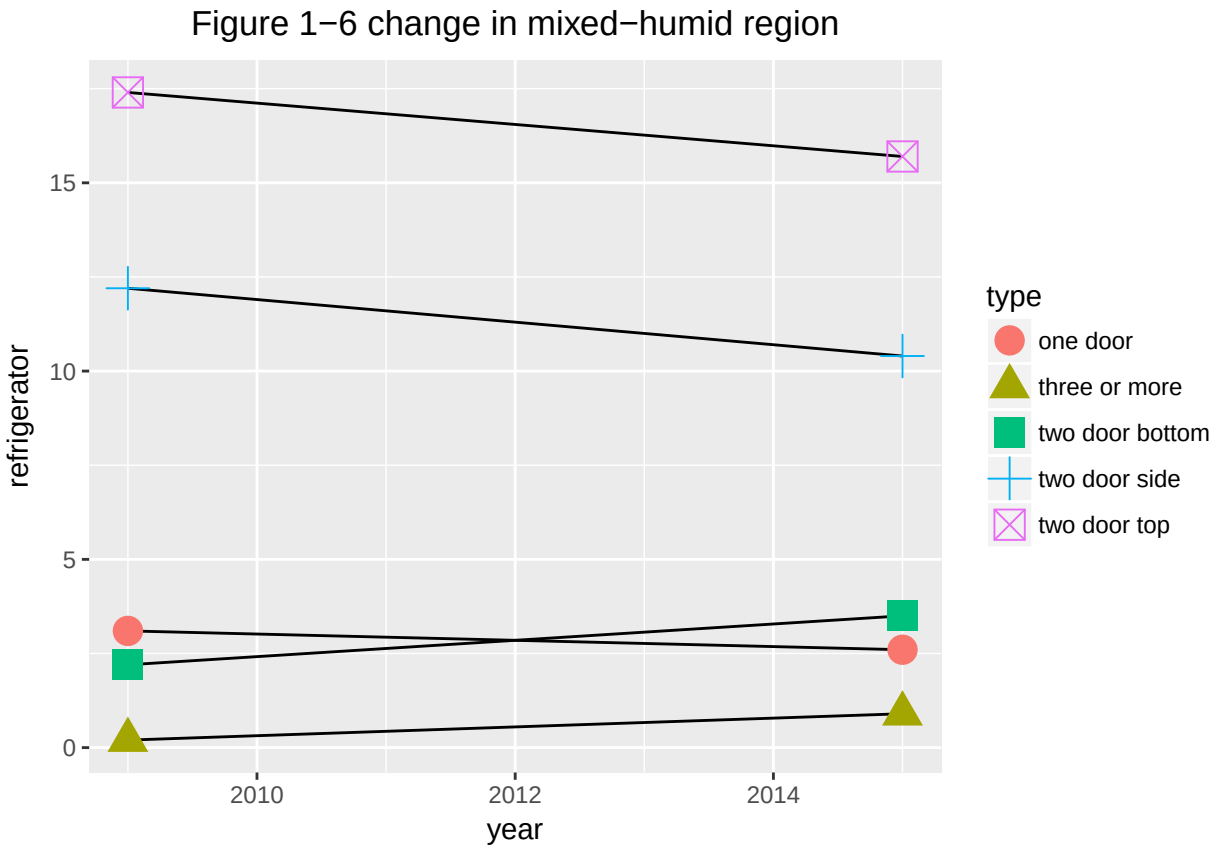
As can be seen from Figure 1-4, "Two doors, side freezer" has an upper trend, which is consistent with the result before.

```
ggplot(data2015[data2015$regionnum == 3,], aes(x = year, y = refrigerator, group = type))+
  geom_line()+
  geom_point(aes(shape = type, colour = type),size = 5)+
  labs(title = "Figure 1-5 change in marine region")+
  theme(plot.title = element_text(hjust = 0.5))
```



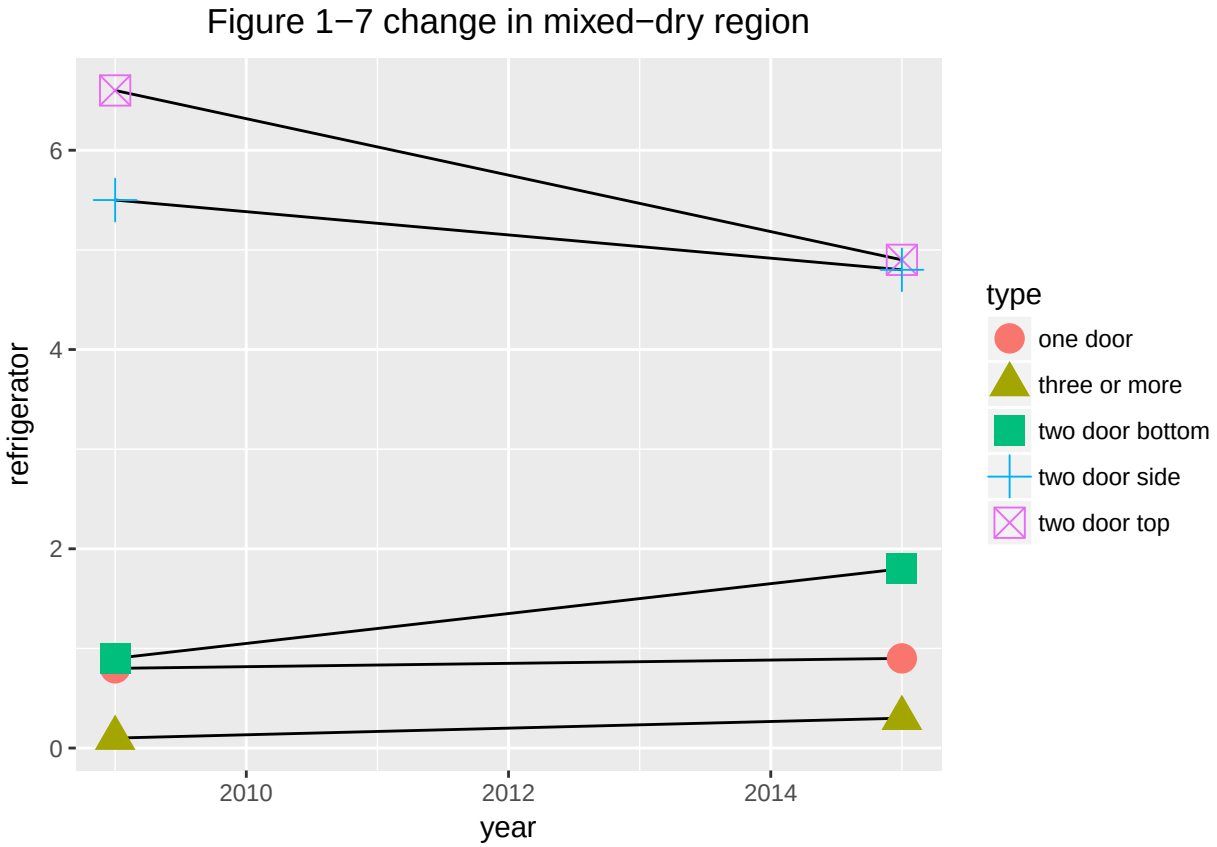
From Figure 1-5, we plan to introduce "Two doors, bottom freezer" to the market in the marine region, because of ite upper trend and significant quantity in used.

```
ggplot(data2015[data2015$regionnum == 2,], aes(x = year, y = refrigerator, group = type))+
  geom_line()+
  geom_point(aes(shape = type, colour = type),size = 5)+
  labs(title = "Figure 1-6 change in mixed-humid region")+
  theme(plot.title = element_text(hjust = 0.5))
```



As shown in Figure 1-6, “Two doors, top freezer” is still have most percentage in use, and therefore, we give the suggestion to mainly put this type in the mixed-humid region.

```
ggplot(data2015[data2015$regionnum == 1,], aes(x = year, y = refrigerator, group = type))+
  geom_line()+
  geom_point(aes(shape = type, colour = type),size = 5)+
  labs(title = "Figure 1-7 change in mixed-dry region")+
  theme(plot.title = element_text(hjust = 0.5))
```



From Figure 1-7, the best plan for mixed-dry region is put "Two doors, side freezer" in the market due to its large used in daily life.

From the result above, we can get to the conclusion that the "Two doors, bottom freezer" is used wider in 2015 compared to 2009, although it was still less used compared to "Two doors, top freezer". Therefore, we make markerting plans for the "type" part as follows:

Table 1: Type plan for different region

region	type
very cold	Two doors, top freezer
hot humid	Two door,side freezer
marine	Two doors, bottom freezer
mixed-humid	Two doors, top freezer
mixed-dry	Two doors, side freezer

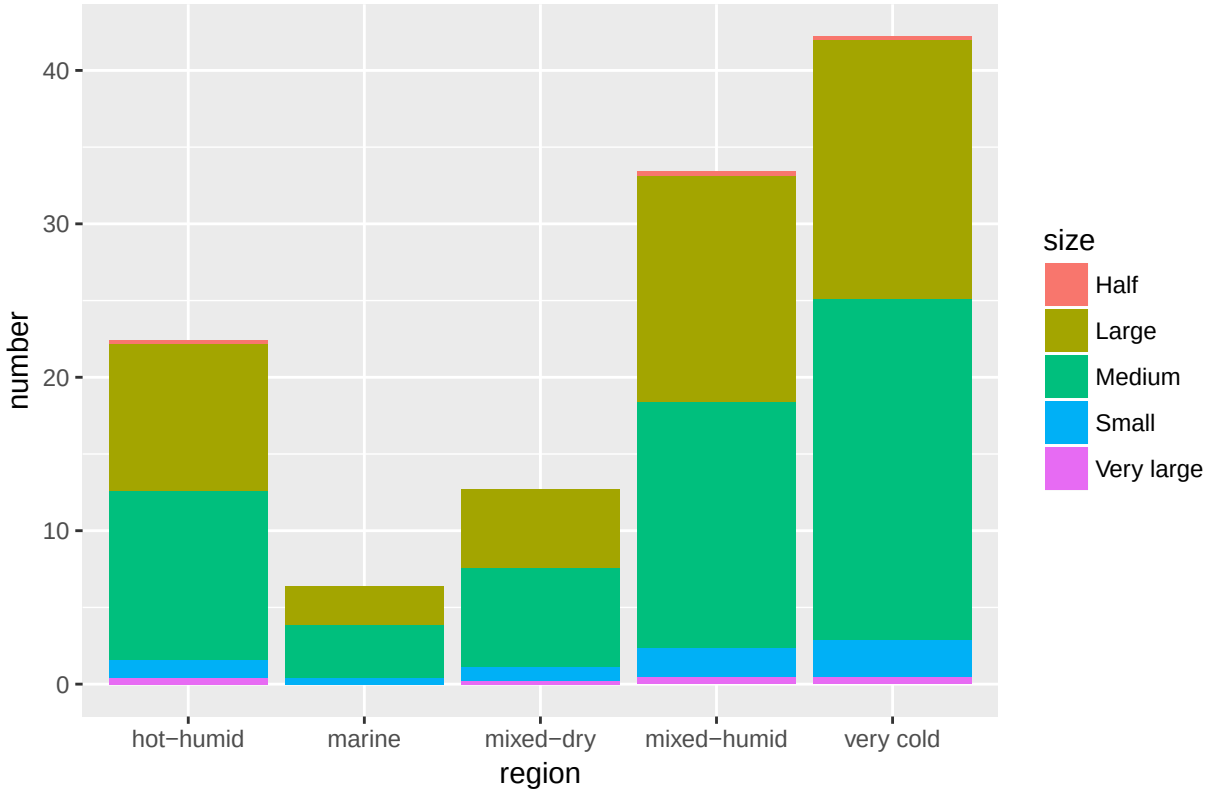
Plan related to Size

```
data <- read.table("2015_2.csv", header = TRUE, sep = ",")
size2015 <- ggplot(data[1:25,], aes(x = region, y = number, fill = size ))+
geom_bar(stat = "identity")+
labs(title = "Figure 2-1 2015 size - region") + theme(plot.title = element_text(hjust = 0.5))
size2009 <- ggplot(data[26:50,], aes(x = region, y = number, fill = size))+
  geom_bar(stat = "identity")+
labs(title = "Figure 2-2 2009 size - region") +
```

```
theme(plot.title = element_text(hjust = 0.5))
size2015
```

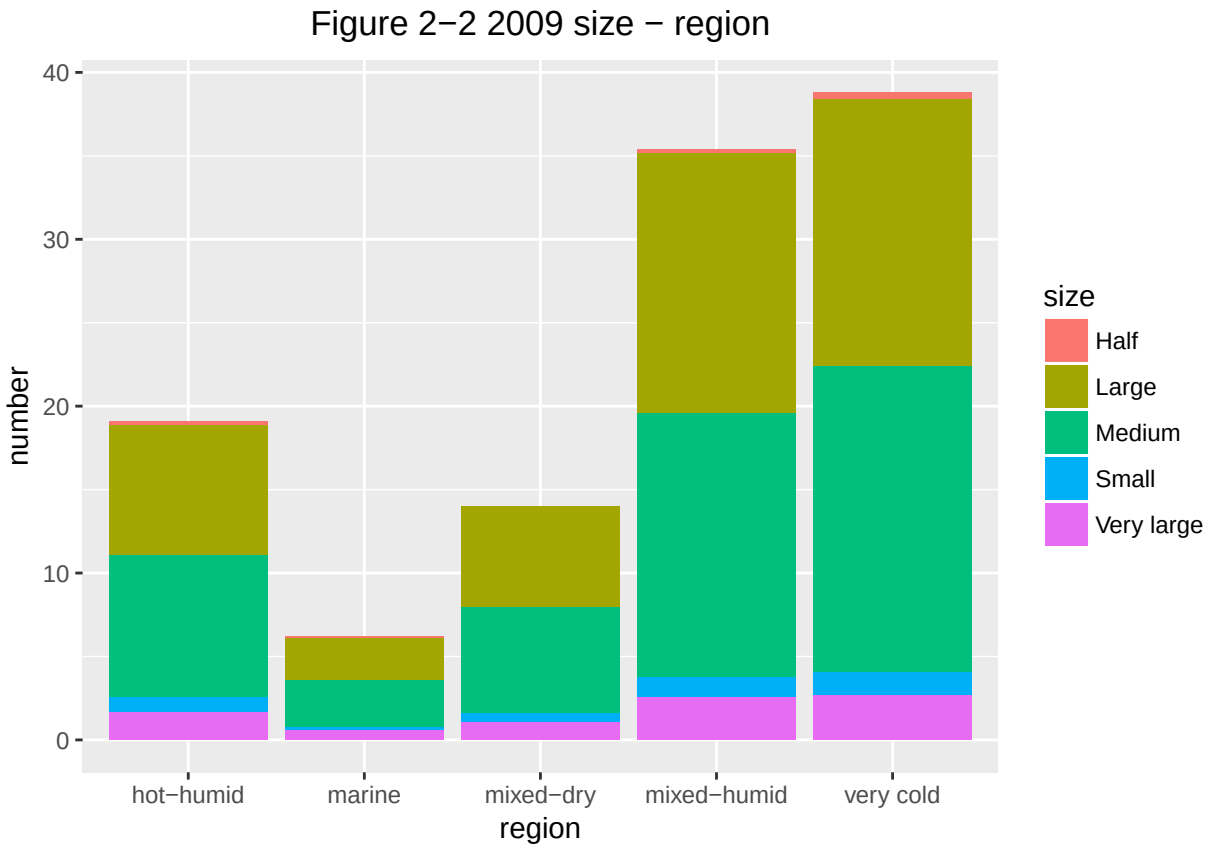
```
## Warning: Removed 3 rows containing missing values (position_stack).
```

Figure 2−1 2015 size − region



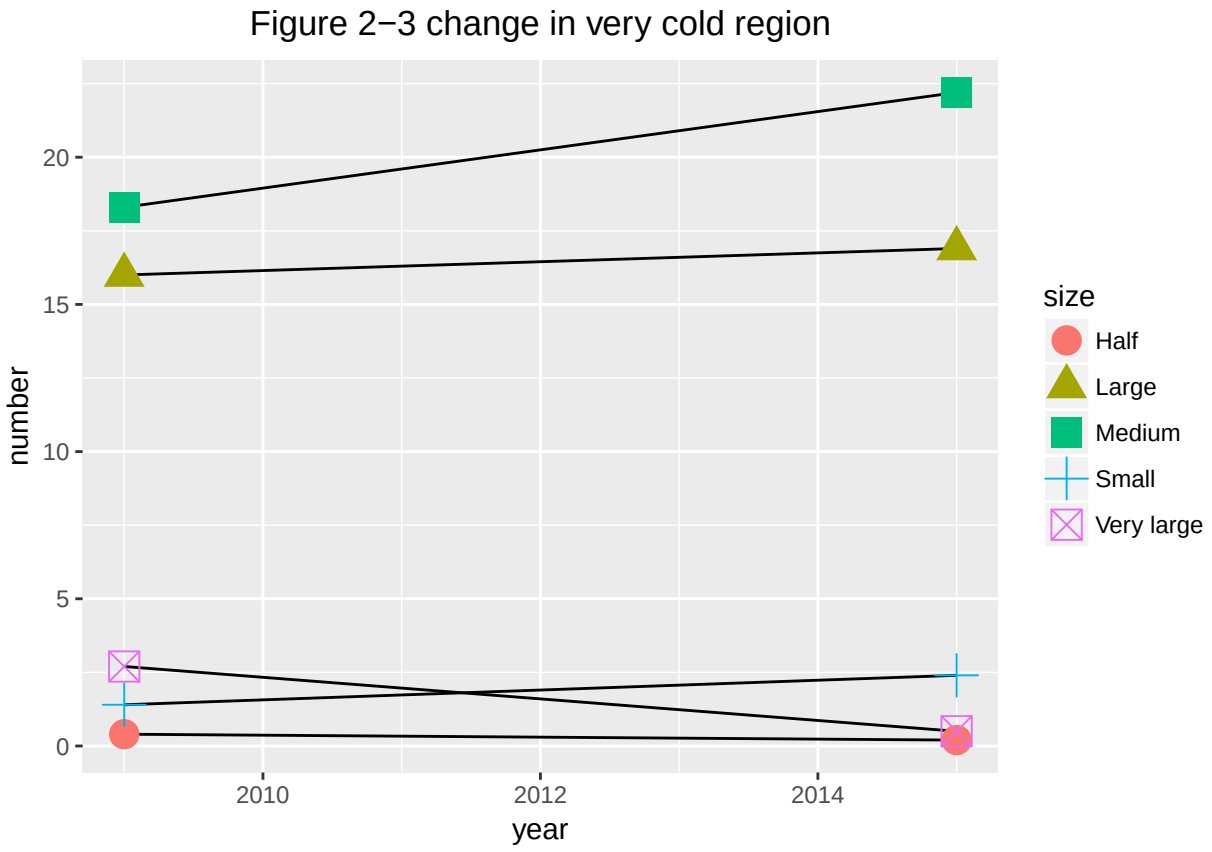

```
size2009
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```



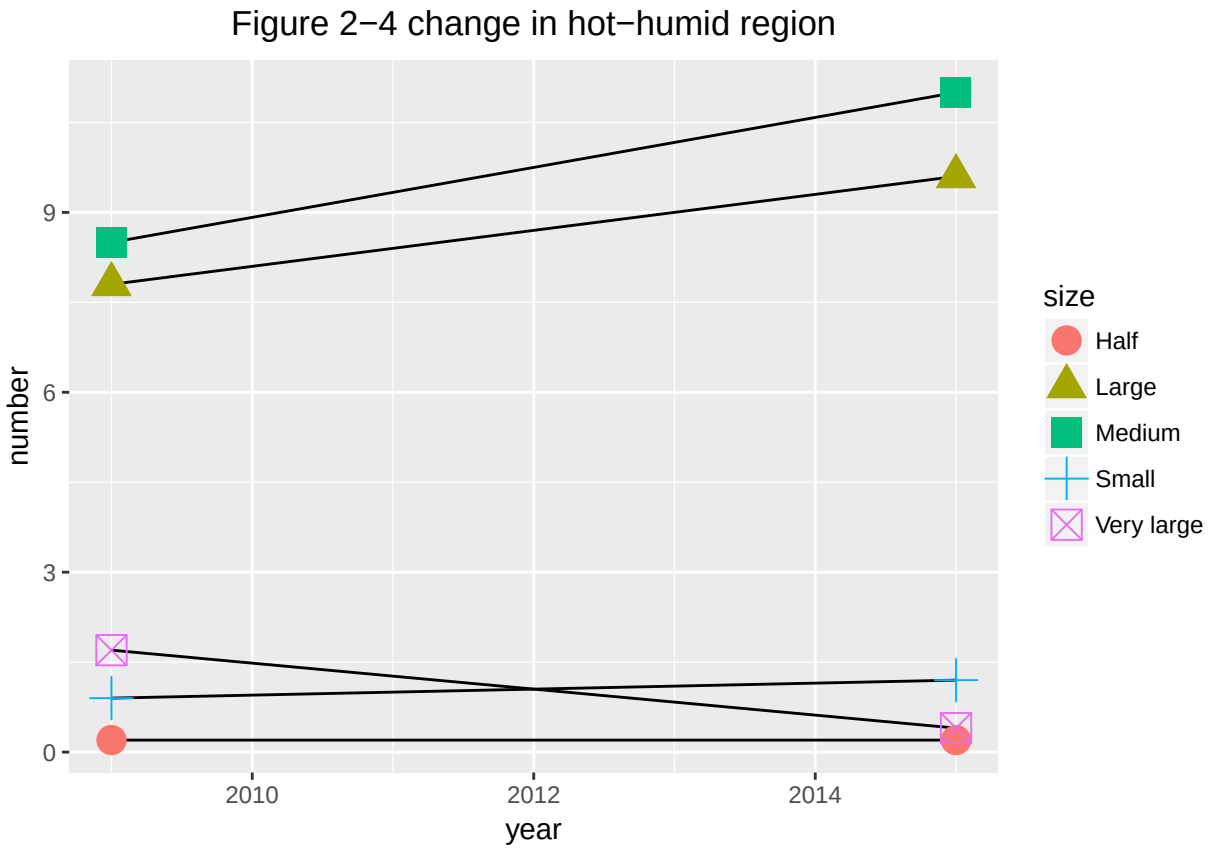
Accoding to figure 2-1 and figure 2-2,the number of housing units that used refrigerators are different in different regions,and the refrigerartors of different size have similar proportion in different regions.

```
ggplot(data[data$regionnum == 5,], aes(x = year, y = number, group = size))+ geom_line()+
geom_point(aes(shape = size, colour = size),size = 5)+
labs(title = "Figure 2-3 change in very cold region")+
theme(plot.title = element_text(hjust = 0.5))
```



According to the figure 2-3,the number of Mediumn size changes the most,but all of the sizes have not changed much.The number of Medium,Large,Small size slightly increased,and the number of Very large,Half size slightly decreased.

```
ggplot(data[data$regionnum == 4,], aes(x = year, y = number, group = size))+ geom_line()+
geom_point(aes(shape = size, colour = size),size = 5)+
labs(title = "Figure 2-4 change in hot-humid region")+
theme(plot.title = element_text(hjust = 0.5))
```

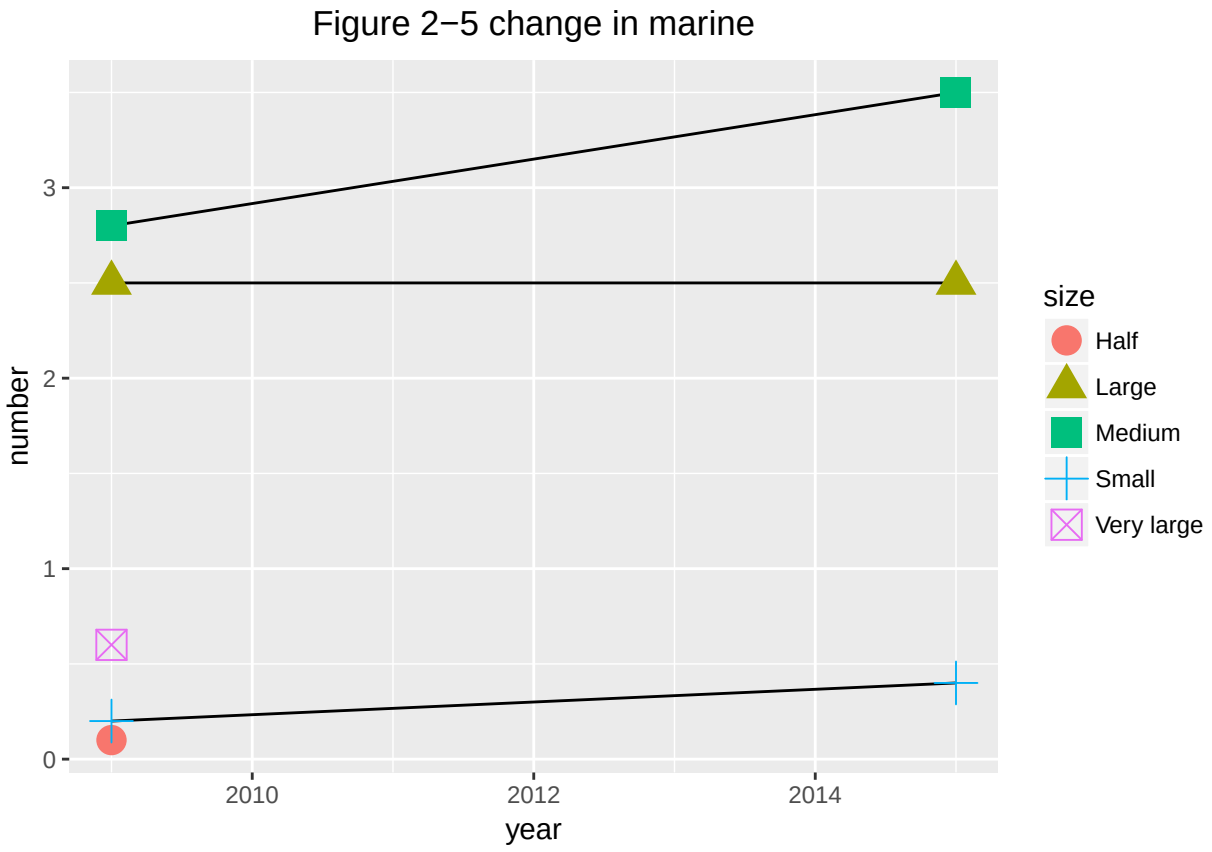


According to the figure 2-4,the number of Mediumn size changes the most,but all of the sizes have not changed much.The number of Medium,Large,Small size slightly increased,and the number of Very large,Half size slightly decreased.

```
ggplot(data[data$regionnum == 3,], aes(x = year, y = number, group = size))+ geom_line()+
geom_point(aes(shape = size, colour = size),size = 5)+
labs(title = "Figure 2-5 change in marine")+
theme(plot.title = element_text(hjust = 0.5))
```

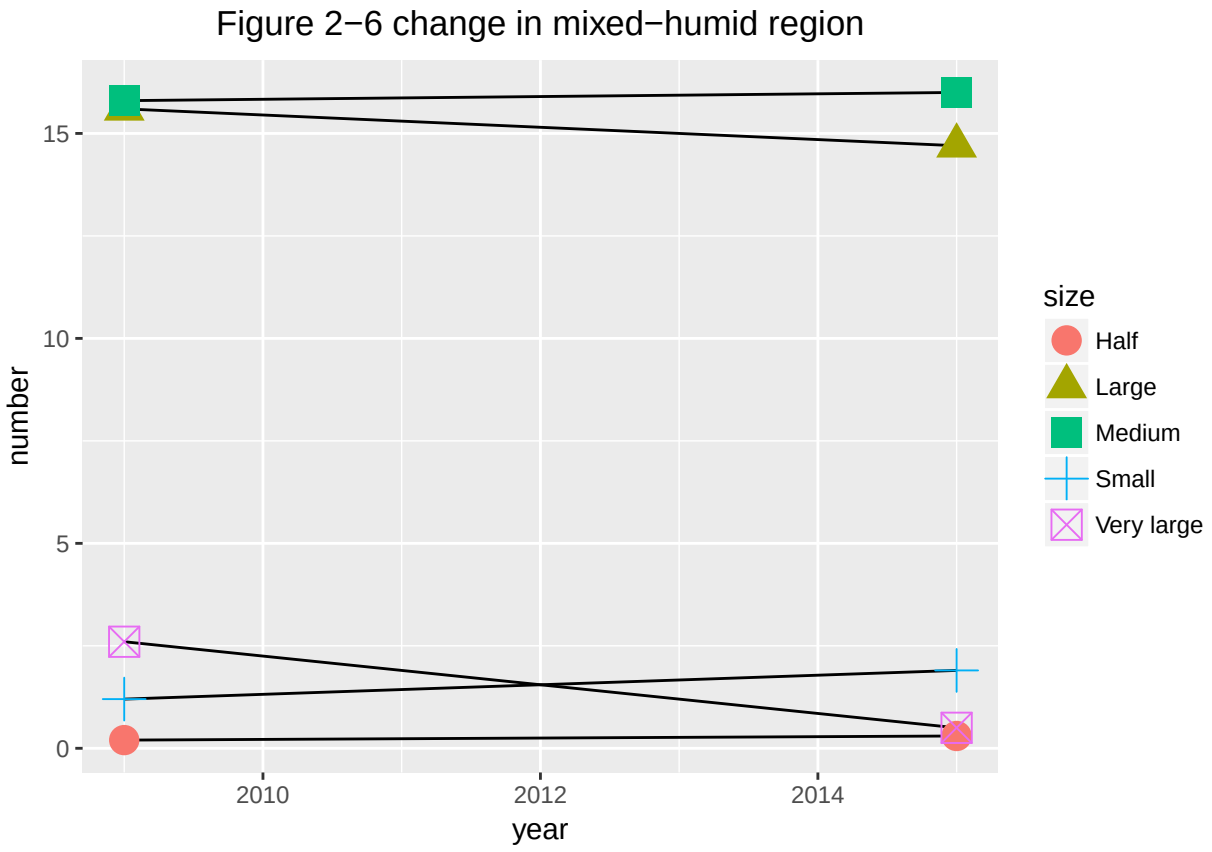
```
## Warning: Removed 2 rows containing missing values (geom_path).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



According to the figure 2-5,the number of Mediumn size changes the most,but all of the sizes have not changed much.The number of Medium,Large,Small size slightly increased,and the number of Very large,Half size missed.

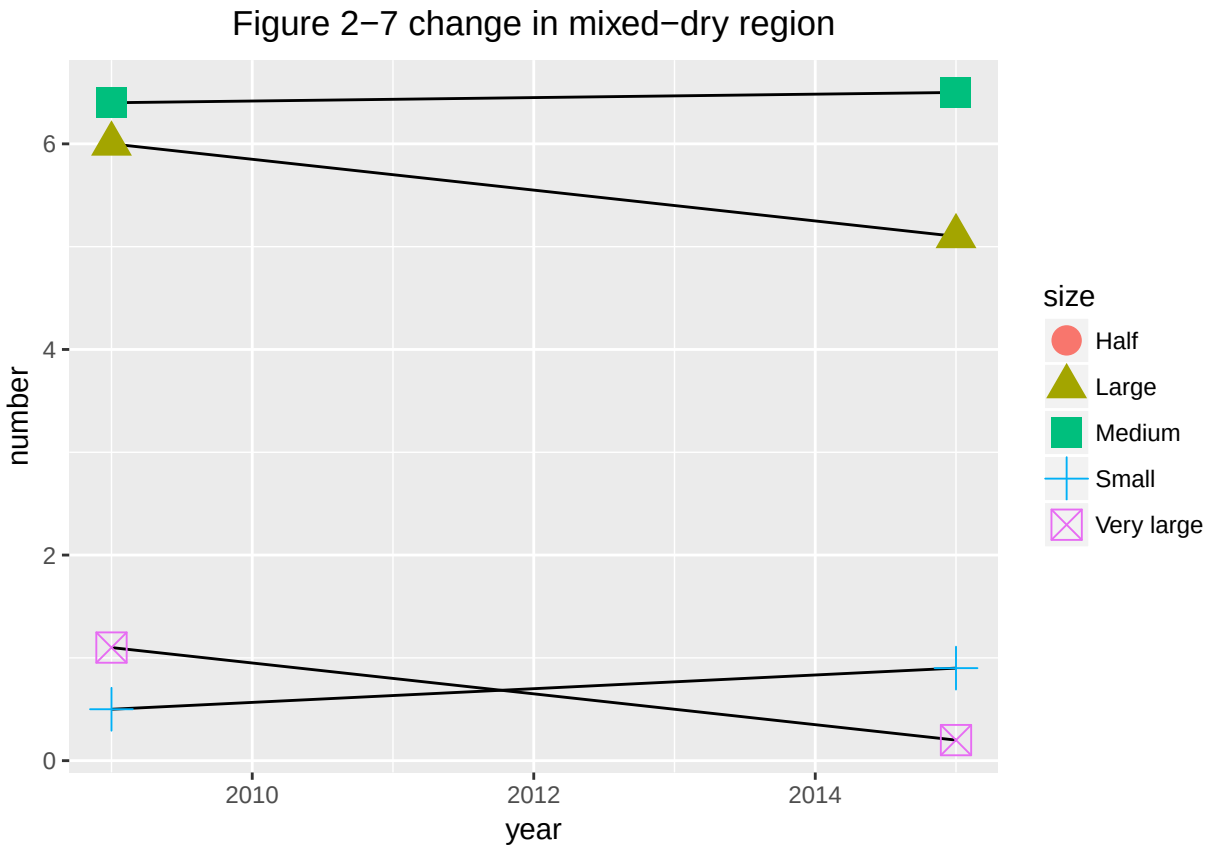
```
ggplot(data[data$regionnum == 2,], aes(x = year, y = number, group = size))+ geom_line()+
geom_point(aes(shape = size, colour = size),size = 5)+
labs(title = "Figure 2-6 change in mixed-humid region")+
theme(plot.title = element_text(hjust = 0.5))
```



According to the figure 2-6, all of the sizes have not changed much.The number of Medium,Half,Small size slightly increased,and the number of Very large,Large size slightly decreased,which indicates the housing units in mixed-humid region tend to use smaller size.

```
ggplot(data[data$regionnum == 1,], aes(x = year, y = number, group = size))+ geom_line()+
geom_point(aes(shape = size, colour = size),size = 5)+
labs(title = "Figure 2-7 change in mixed-dry region")+
theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



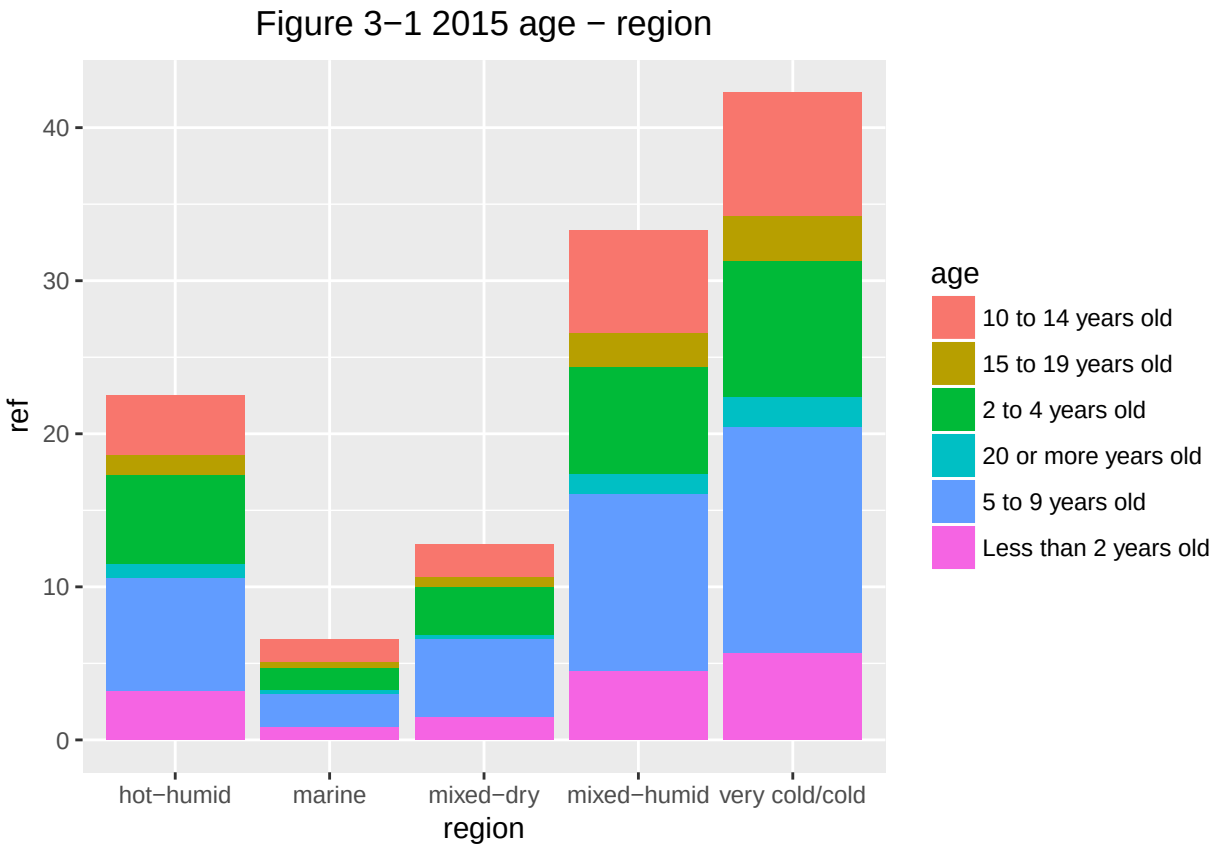
According to the figure 2-7,the number of Large size changes the most,but all of the sizes have not changed much.The number of Medium and Small size slightly increased,and the number of Very large and Large size slightly decreased,which indicates the housing units in mixed-dry region tend to use smaller size.

In conclusion,the Medium and Large size are used in most housing units,and the Medium size has remained the high quantity,the Large size has decrease a little in some region.So we can reduce the number of Large size in mixed-dry and mixed-humid regions.The Half,Small and Very large size are used in much less housing units,and the number of Small size has remained the level or increased a little,the number of Very large size has decreased across the crountry,the number of Half size missed a lot but it has basically remain the level.So we should increase the number of Small size a little,and decrease the number of Very large size.

Quantity

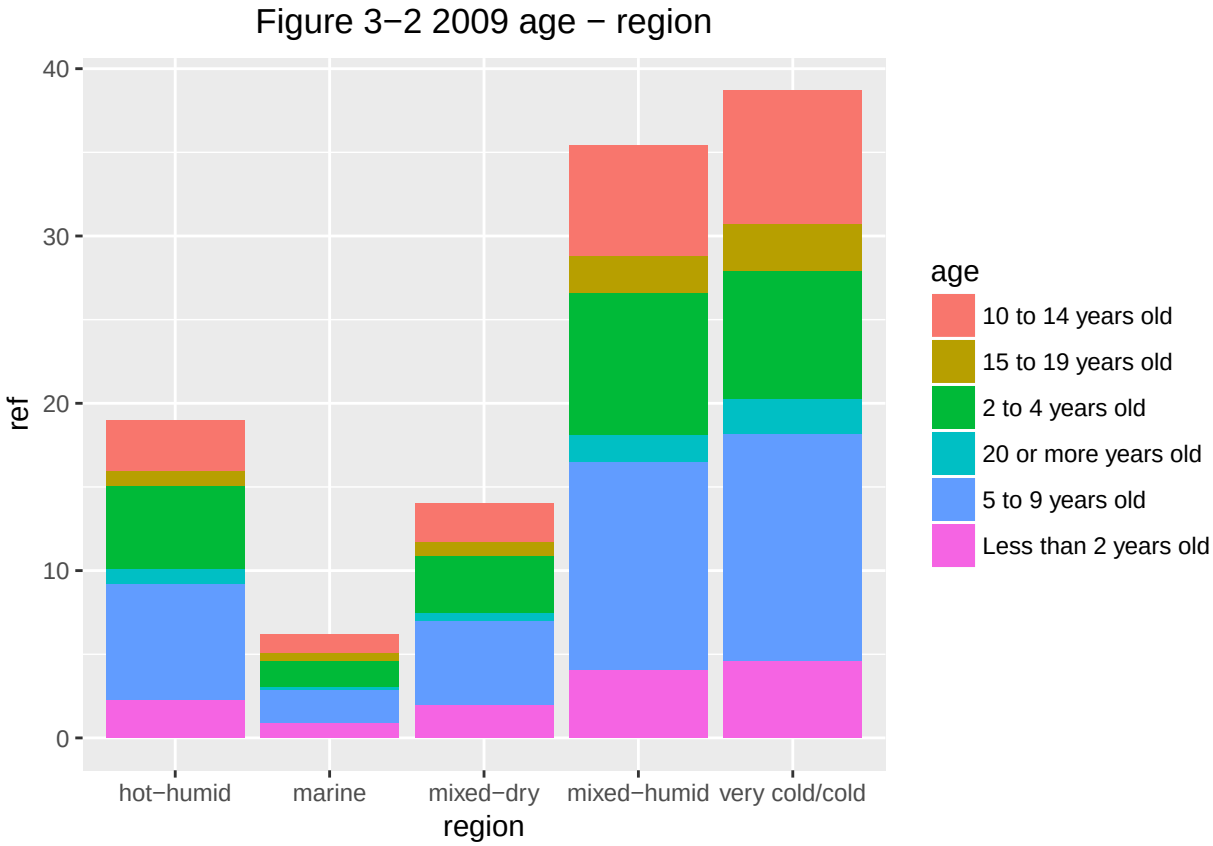
The graph of refrigerators' age in 2015 and 2009.

```
library(ggplot2)
data_1<-read.table("2015_1.csv", header = TRUE, sep = ",")
plot2015age <-ggplot(data_1[1:30, ], aes(x =region , y = ref, fill = age))+
geom_bar(stat = "identity")+
labs(title = "Figure 3-1 2015 age - region")+
theme(plot.title = element_text(hjust = 0.5))
plot2015age
```



As the figure show above, we can see that in different region, the biggest part of refrigerators' age is 5 to 9years old. And the smallest part of refrigeratos' age is 20 or more years old.

```
plot2009age <- ggplot(data_1[31:60,], aes(x = region, y = ref, fill = age))+
 geom_bar(stat = "identity")+
 labs(title = "Figure 3-2 2009 age - region")+
 theme(plot.title = element_text(hjust = 0.5))
plot2009age
```



The situations are similar in 2009. No matter what regions are, the biggest part of refrigerators' age is 5 to 9years old. And the smallest part of refrigeratos' age is 20 or more years old.

Also, from the two graphs above, we can see directly that in different region, the distribution of refrigerators' age are similar, which means that the usage of the refrigerator is stable.

The barplot of age in different region.

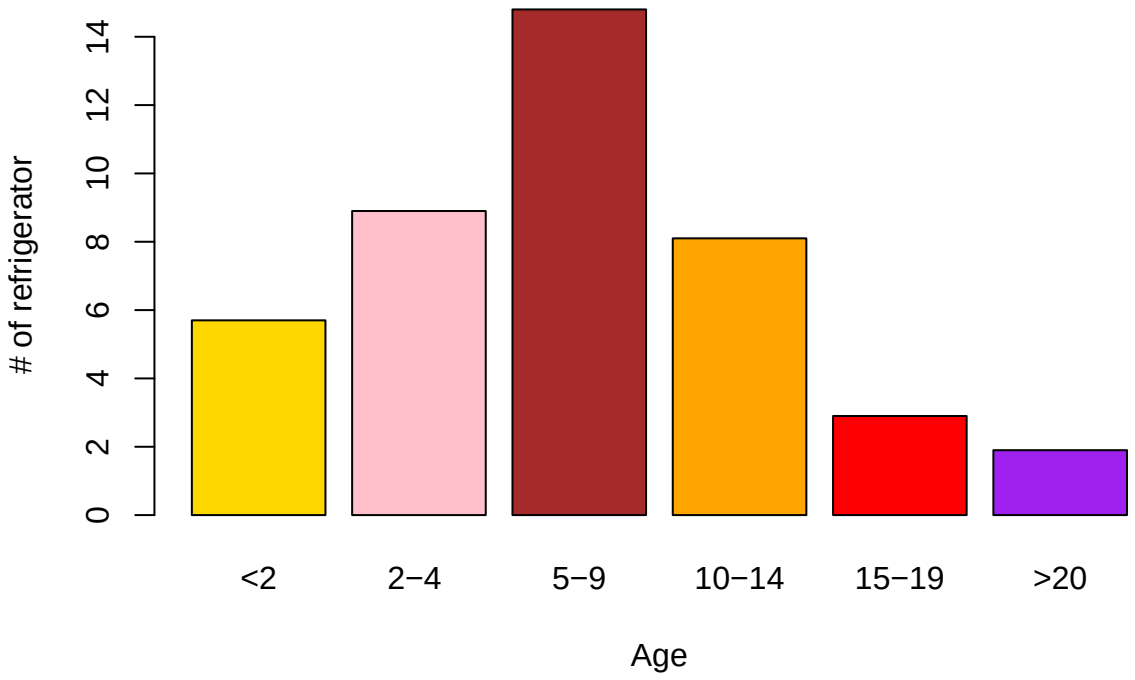
Considering the age of refrigerators, we thought that the datas collected in 2009 had little influence in our decision, so we would focus on the datas in 2015. Based on the data in 2015, we would like to find whether some region would have more requirments in buying new refrigerators.

We assume that refrigerators with age less than 4 years old are almost new refrigerators, and those used more than 10 years are almost old refrigerators.

The age of refrigerators in Very Cold/Cold Region.

```
barplot(data_1[data_1$region == "very cold/cold" & data_1$year == "2015",]$ref,
        names.arg  = c("<2","2-4 ","5-9","10-14","15-19",">20"),
        xlab = "Age", ylab = "# of refrigerator",
        col=rep(c("gold","pink","brown","orange","red","purple")),
        main = "Figure 3-3 The age of refrigerators in Very Cold/Cold Region")
```

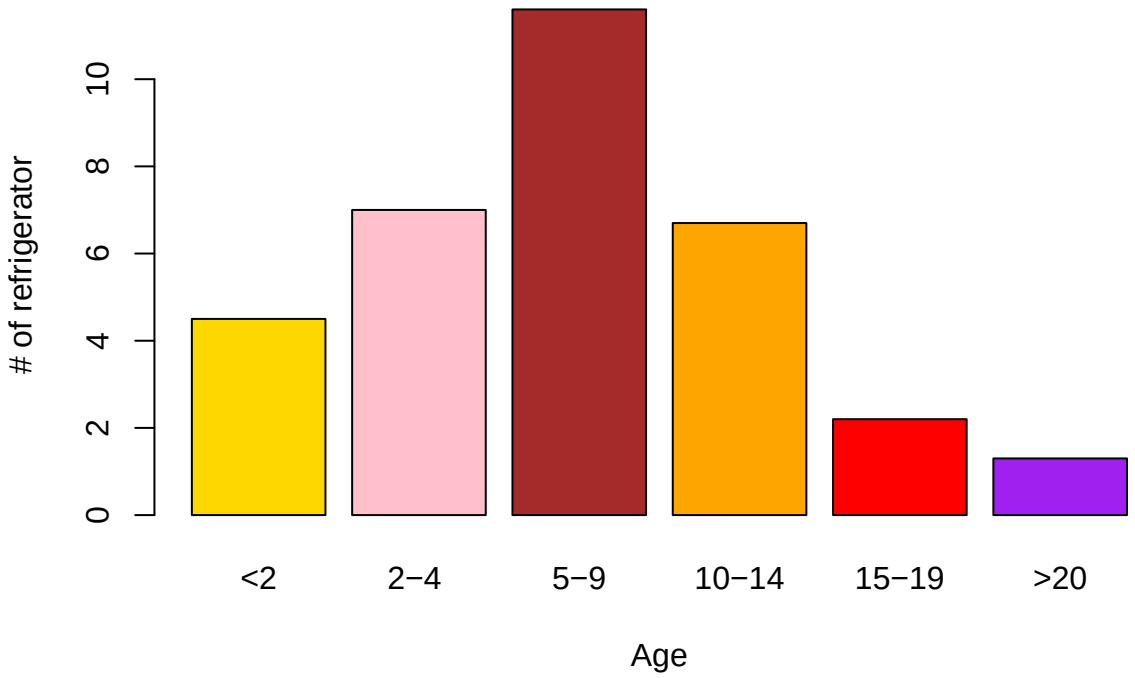
Figure 3−3 The age of refrigerators in Very Cold/Cold Region



As the figure shows, most of the regrigerator in very cold region have been used 5-9 years and then is 2-4 years. However, there is only small difference between the number of 2-4 years old refrigerators and the number of 10-14 years old refrigerators. And the number of 15 or more years old refrigerators in this area is very close to the number of less than 2 years old refrigerators. Since the total amount in this ares is the biggest, which means that there may have more requirement in changing the refrigerators, and the difference between the number of new and old refrigerators in this area is not obvious, we would like to supply more refrigerators in very cold region.

The age of refrigerators in Mixed-humid Region

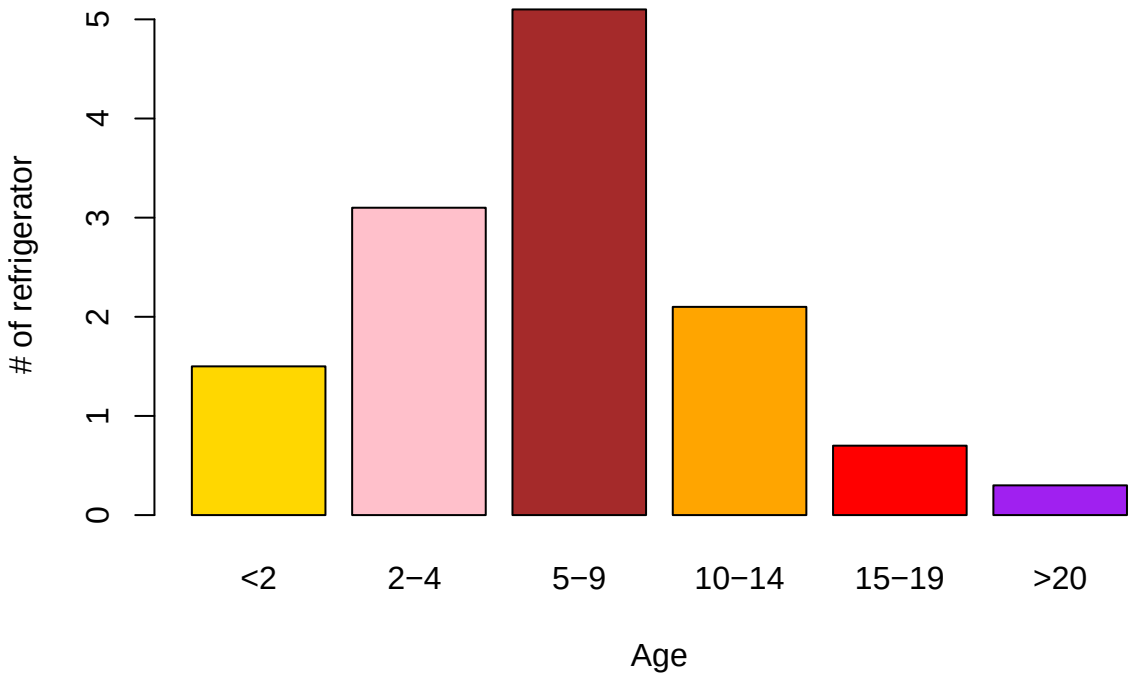
```
barplot(data_1[data_1$region == "mixed-humid" & data_1$year == "2015",]$ref,
        names.arg  = c("<2","2-4 ","5-9","10-14","15-19",">20"),
        xlab = "Age", ylab = "# of refrigerator",
        col=rep(c("gold","pink","brown","orange","red","purple")),
        main = "Figure 3-4 The age of refrigerators in Mixed-humid Region")
```



Mixed-humid region has similar situation as very cold region has, and we would like to supply more refrigerators in mixed-humid region.

The age of refrigerators in Mixed-dry Region

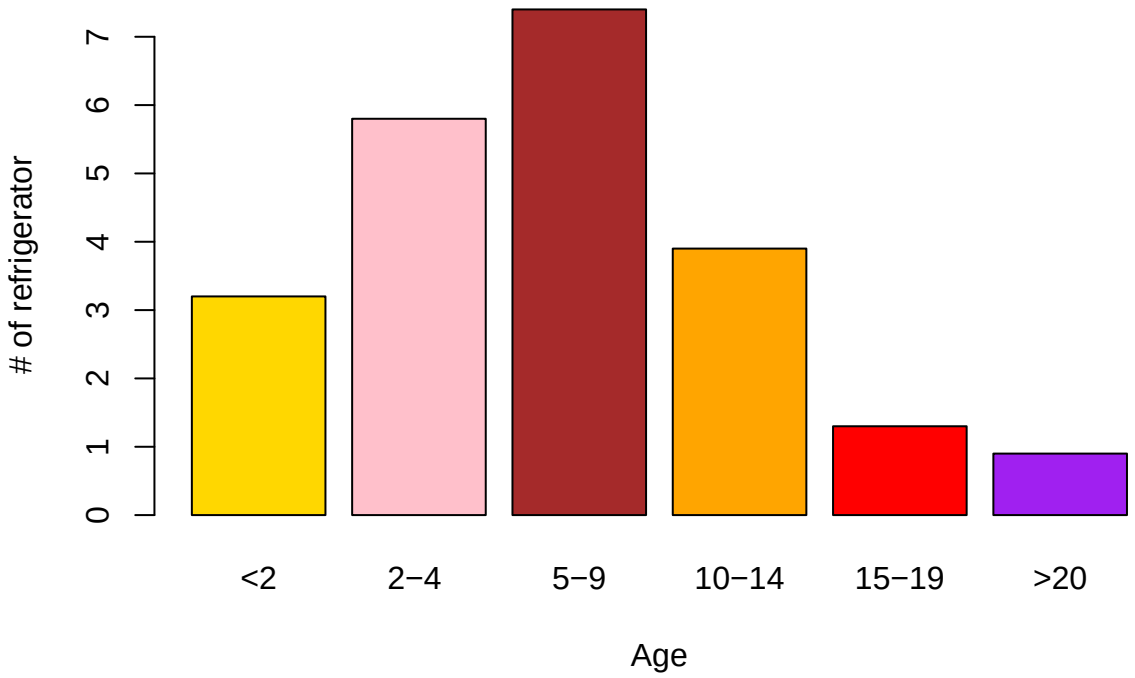
```
barplot(data_1[data_1$region == "mixed-dry" & data_1$year == "2015",]$ref,
        names.arg  = c("<2","2-4 ","5-9","10-14","15-19",">20"),
        xlab = "Age", ylab = "# of refrigerator",
        col=rep(c("gold","pink","brown","orange","red","purple")),
        main = "Figure 3-5 The age of refrigerators in Mixed-dry Region")
```



Based on the figure above, most of the refrigerators in mixed-dry area haven't been used more than 10 years. We would not consider to supply more new refrigerators in this area.

The age of refrigerators in Hot-humid Region

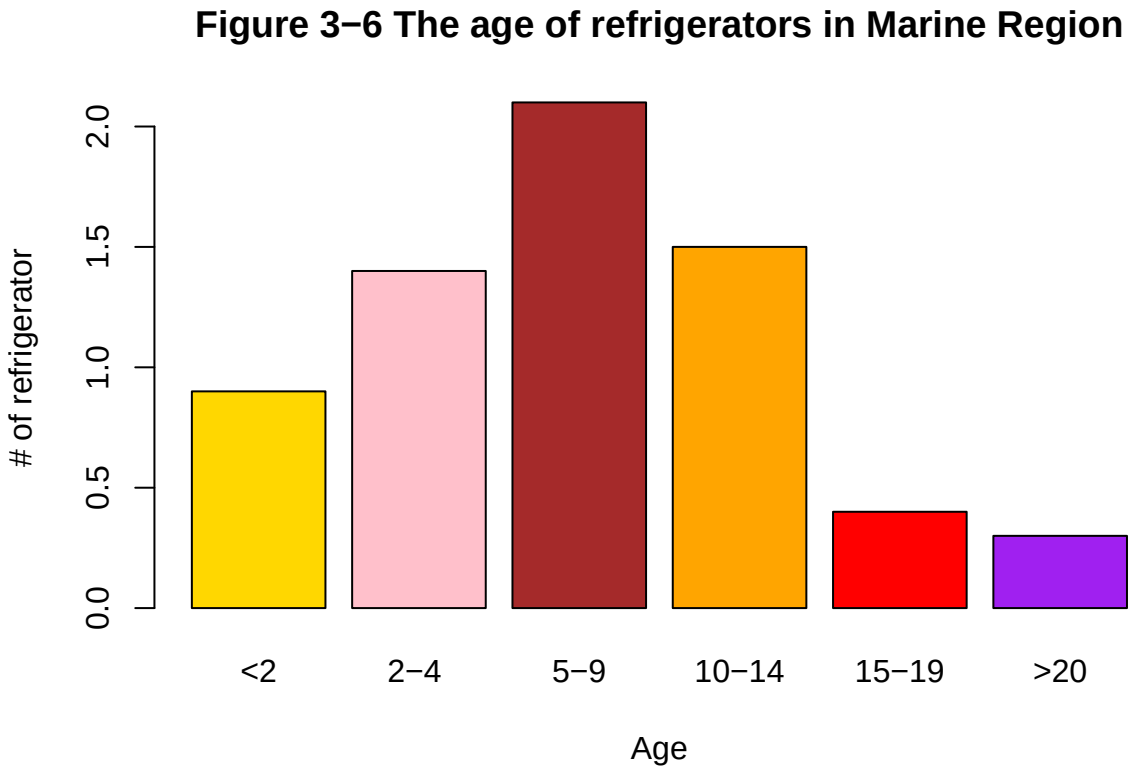
```
barplot(data_1[data_1$region == "hot-humid" & data_1$year == "2015",]$ref,
        names.arg  = c("<2","2-4 ","5-9","10-14","15-19",">20"),
        xlab = "Age", ylab = "# of refrigerator",
        col=rep(c("gold","pink","brown","orange","red","purple")),
        main = "Figure 3-6 The age of refrigerators in Hot-humid Region")
```



Based on the figure above, the situation in hot humid region is just like the situation in mixed-dry area. Most of the refrigerators haven't been used more than 10 years. And the difference between 2-4years and 5-9 years is not as large as the situation in other areas, which also means that most of the refrigerators in hot-humid area can still be used in a long time. We would not consider to supply more new refrigerators in this area.

The age of refrigerators in Marine Region

```
barplot(data_1[data_1$region == "marine" & data_1$year == "2015",]$ref,
        names.arg  = c("<2","2-4 ","5-9","10-14","15-19",">20"),
        xlab = "Age", ylab = "# of refrigerator",
        col=rep(c("gold","pink","brown","orange","red","purple")),
        main = "Figure 3-6 The age of refrigerators in Marine Region")
```



In the marine area, the situation is different. 5-9 years old refrigeratos are still the biggest part, however, the number of 10-14 years old refrigerators is higher than the number of 2-4 years old refrigerators. However, the total amount in this area is the smallest, so we would not consider supply extra refrigerators in this area.

Conclusion

By exploring the EIA data, we give the refrigerator marketing plan for different region, which is illustrated as follows:

Table 2: Marketing plan for different region

Region	Type	Size	Quantity
very cold	Two doors, top freezer	increase the number of medium size	more
hot humid	Two door,side freezer	increase the number of medium and large size	less
marine	Two doors, bottom freezer	increase the medium size	less
mixed-humid	Two doors, top freezer	reduce the number of large size	more
mixed-dry	Two doors, side freezer	reduce the number of large size	less