MA615 Project 1

Liwen Zhang, Luying Cao, Jiazhou Liu

October 8, 2017

Introduction

Nowadays, refrigerators play an essential role in people's daily lives. Our following report analyzes determining the target customer group and marketing approaches. We used 'Appliances in U.S. homes by number of household members' from 2015 as our data set and did some explorations using tibble, ggplot2, dplyr, etc.

```
#Data Input
d1 <- read_csv("hc3.4.csv")
```

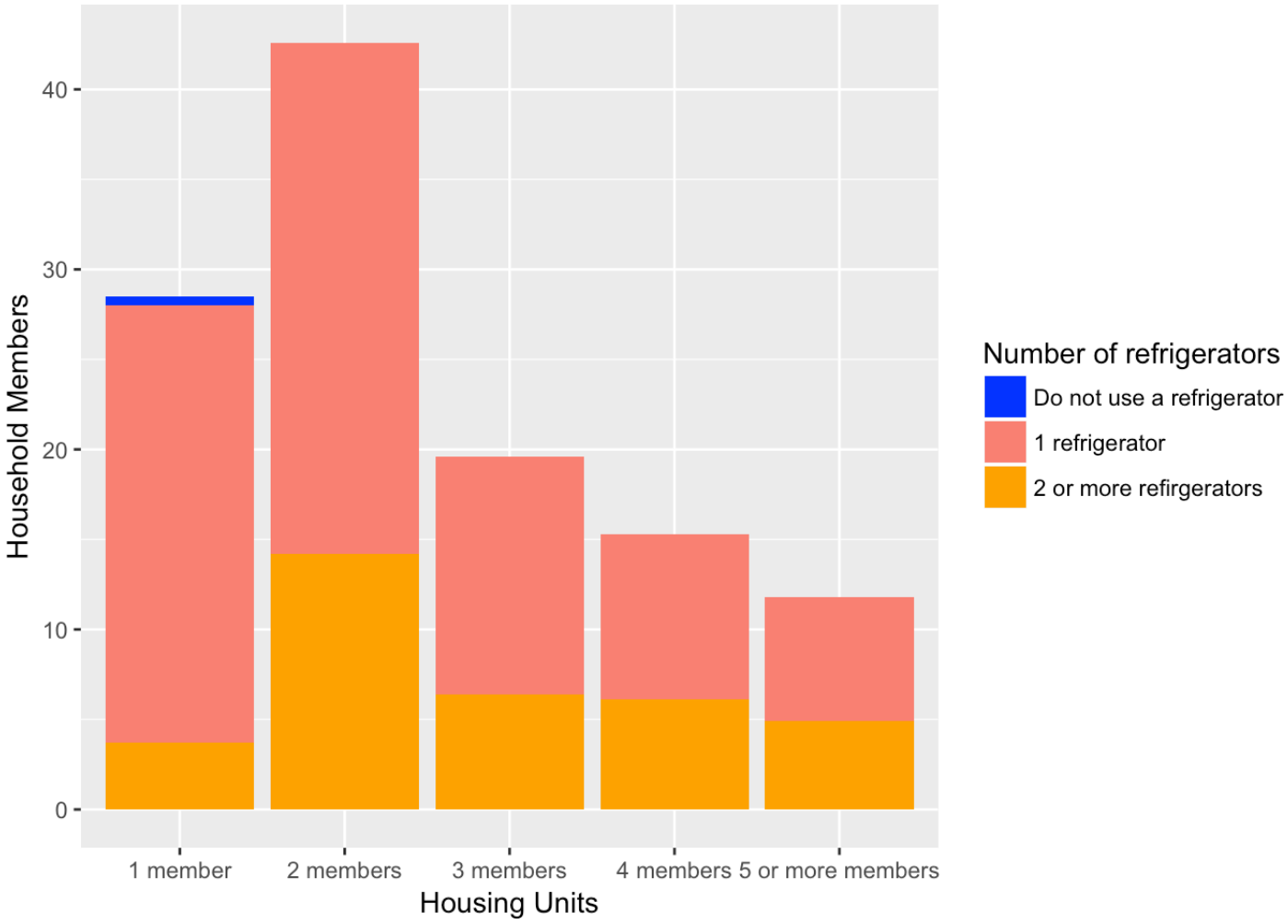
```
## Parsed with column specification:
## cols(
##   Appliances = col_character(),
##   Frequency = col_character(),
##   `Total U.S.2` = col_character(),
##   `1 member` = col_character(),
##   `2 members` = col_character(),
##   `3 members` = col_character(),
##   `4 members` = col_character(),
##   `5 or more members` = col_character()
## )
```

```
#Data Cleaning
d1[ d1 == "Q" | d1 == "N" ] <- NA
d1[is.na(d1)] <- 0
d1$`1 member` <- as.numeric(d1$`1 member`)
d1$`2 members` <- as.numeric(d1$`2 members`)
d1$`3 members` <- as.numeric(d1$`3 members`)
d1$`4 members` <- as.numeric(d1$`4 members`)
d1$`5 or more members` <- as.numeric(d1$`5 or more members`)

Appliances = rep(d1$Appliances, 5)
Frequency = rep(d1$Frequency, 5)
Household_members = c(d1$`1 member`, d1$`2 members`, d1$`3 members`, d1$`4 members`,
d1$`5 or more members`)
Housing_units = rep(c("1 member", "2 members", "3 members", "4 members", "5 or more m
embers"), each = 89)
d2 <- tibble(Appliances, Frequency, Household_members, Housing_units)
```

```
sd1 <- d2 %>%
  filter(d2$Appliances == "Refrigerators" & !d2$Frequency == "Use a refrigerator")
sd1$Frequency <- factor(sd1$Frequency, levels = c("Do not use a refrigerator", "1", "
2 or more"))[]

sd1 %>%
  group_by(Frequency) %>%
  summarise(Avg_Household_Members=mean(Household_members))
```

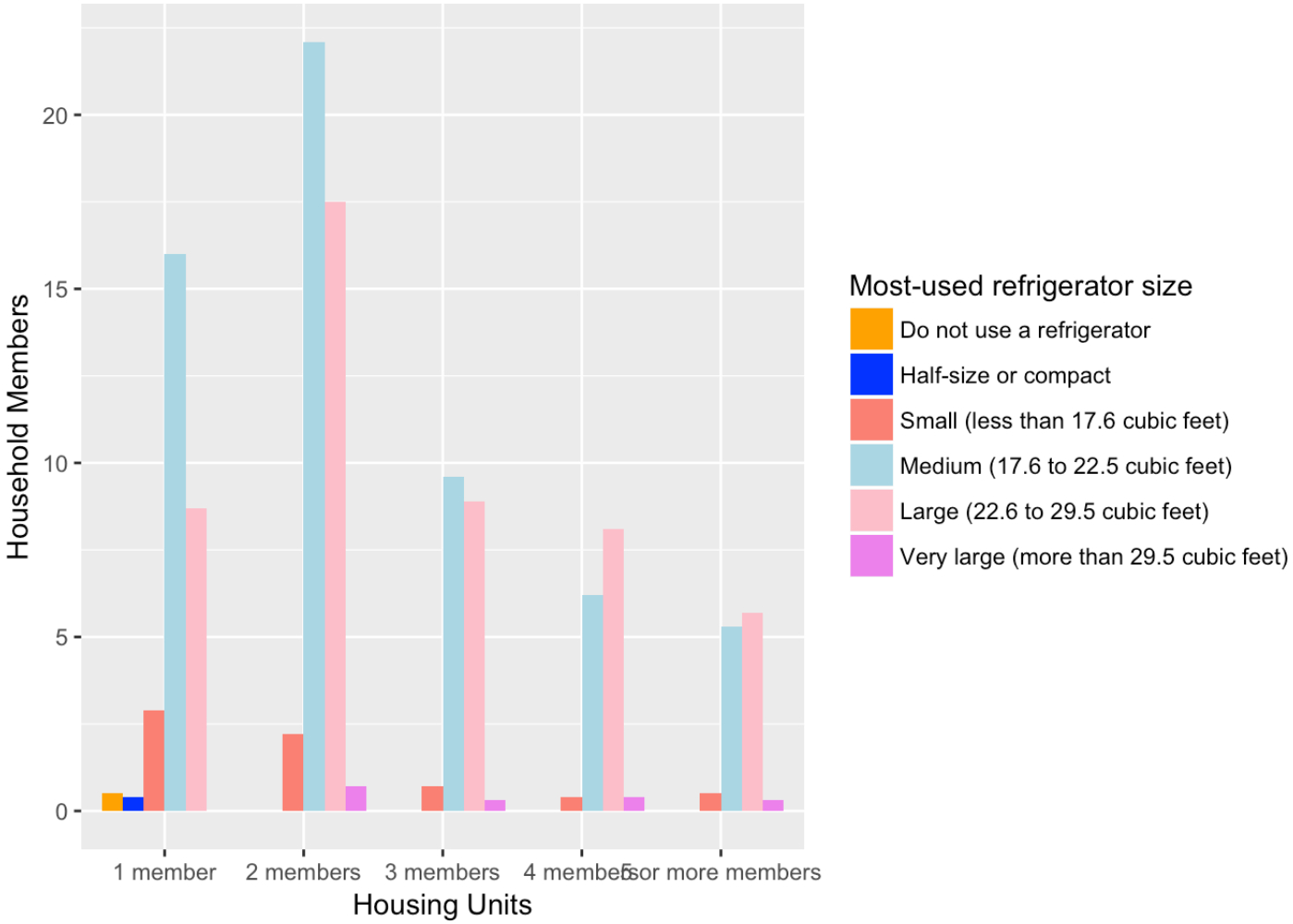
```
## # A tibble: 3 x 2
##                   Frequency Avg_Household_Members
##                      <fctr>                 <dbl>
## 1 Do not use a refrigerator                  0.10
## 2                         1                 16.40
## 3                 2 or more                  7.06
```

```
ggplot(sd1, aes(x=sd1$Housing_units, y=sd1$Household_members, fill = sd1$Frequency))
+
  geom_bar(stat="identity") +
  scale_fill_manual(values=c("blue", "salmon", "orange", "pink"),
                    name="Number of refrigerators",
                    labels=c("Do not use a refrigerator", "1 refrigerator", "2 or mor
e refirgerators")) +
  xlab("Housing Units") +
  ylab("Household Members")
```



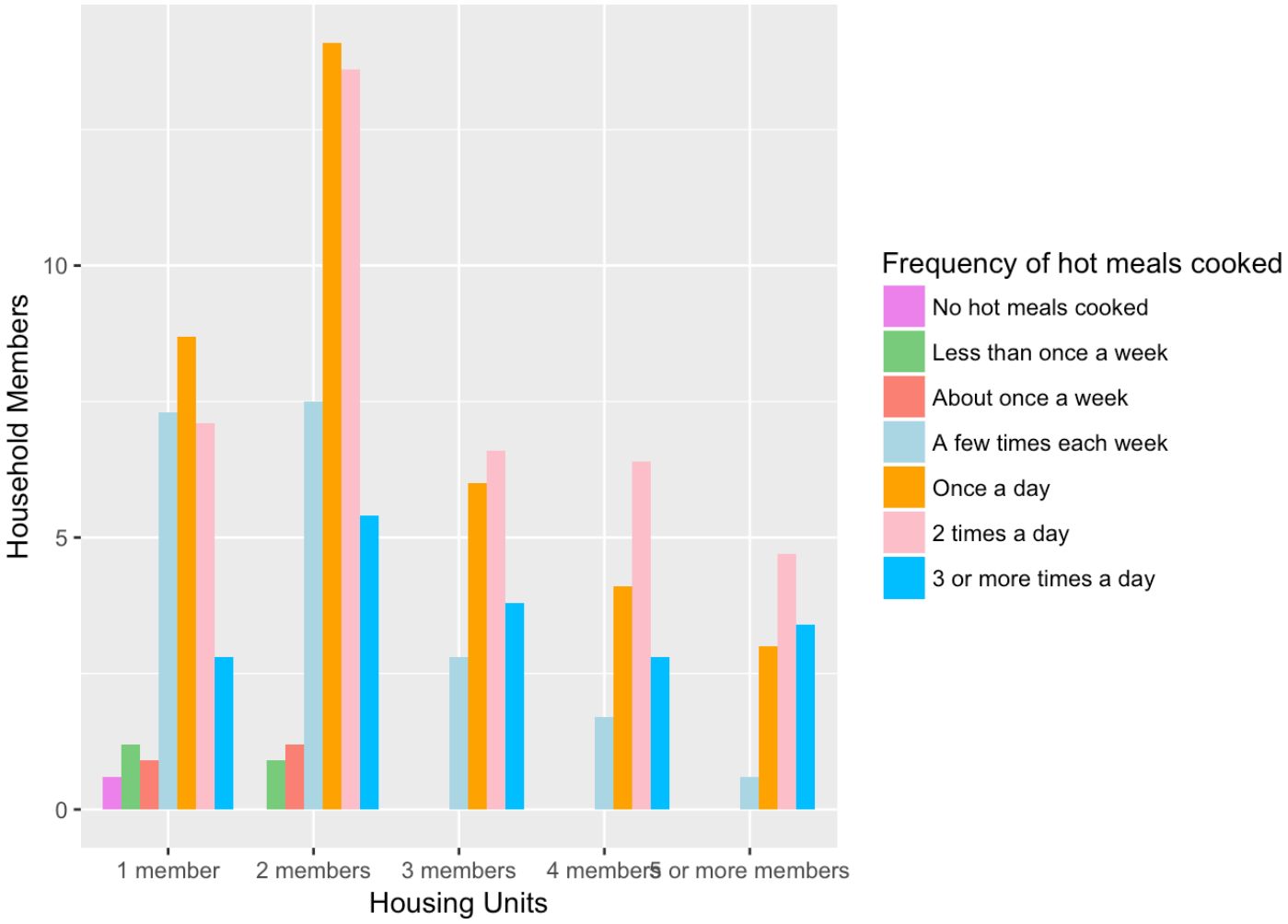
According to the bar plot above, we found out that as the number of household members increases, the proportion of using 2 or more refrigerators also increases. The probabiltiy of households with greater than or equal to 2 members to have a second refrigerator is higher than that with only 1 member. Thus, these families would be our target customer group becaue there could be a potential increase in selling a second refrigerator to them.

```
sd2 <- d2 %>%
  filter(d2$Appliances == "Most-used refrigerator size")
sd2$Frequency <- factor(sd2$Frequency, levels = c("Do not use a refrigerator", "Half-
size or compact", "Small (less than 17.6 cubic feet)", "Medium (17.6 to 22.5 cubic fe
et)", "Large (22.6 to 29.5 cubic feet)", "Very large (more than 29.5 cubic feet)"))

ggplot(sd2, aes(x=sd2$Housing_units, y=sd2$Household_members, fill = sd2$Frequency))
+
  geom_bar(stat="identity",position='dodge') +
  scale_fill_manual(values=c("orange", "blue", "salmon", "lightblue", "pink", "violet
"),
                    name="Most-used refrigerator size",
                    labels=c("Do not use a refrigerator", "Half-size or compact", "Sm
all (less than 17.6 cubic feet)", "Medium (17.6 to 22.5 cubic feet)", "Large (22.6 to
29.5 cubic feet)", "Very large (more than 29.5 cubic feet)"))+
  xlab("Housing Units") +
  ylab("Household Members")
```



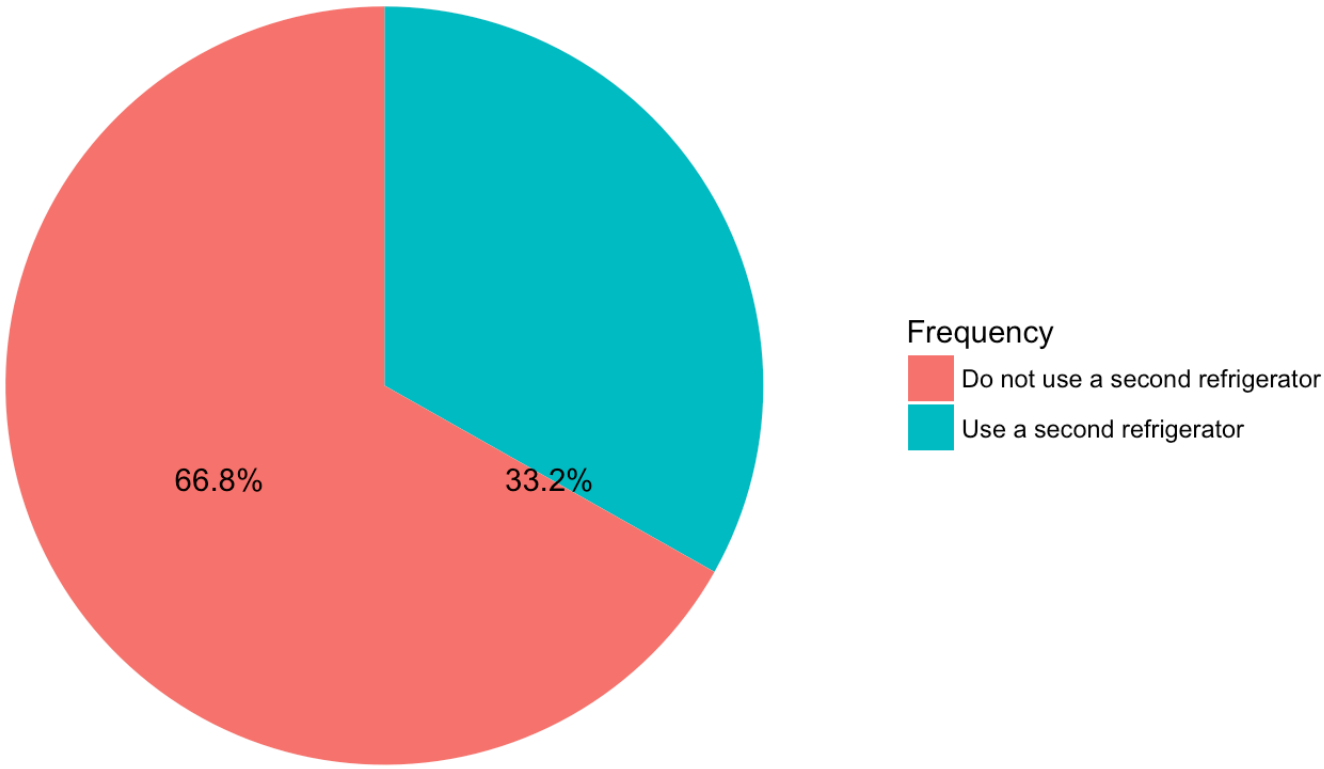
Based on the graph above, we would recommend to market the medium size refrigerators to households that have less than or equal to 3 members. For household households that have more than 3 members, we would market large size refrigerators instead. In addition, since there is no market potential for half-size and very large size refrigerators, we would suggest spend less money promoting them.

```
sd3 <- d2 %>%
  filter(d2$Appliances == "Frequency of hot meals cooked")
sd3$Frequency <- factor(sd3$Frequency, levels = c("No hot meals cooked", "Less than o
nce a week", "About once a week", "A few times each week", "Once a day", "2 times a d
ay", "3 or more times a day"))

ggplot(sd3, aes(x=sd3$Housing_units, y=sd3$Household_members, fill = sd3$Frequency))
+
  geom_bar(stat="identity",position='dodge') +
  scale_fill_manual(values=c("violet", "palegreen3", "salmon", "lightblue", "orange",
"pink", "deepskyblue"),
                    name="Frequency of hot meals cooked",
                    labels=c("No hot meals cooked", "Less than once a week", "About o
nce a week", "A few times each week", "Once a day", "2 times a day", "3 or more times
a day")) +
  xlab("Housing Units") +
  ylab("Household Members")
```



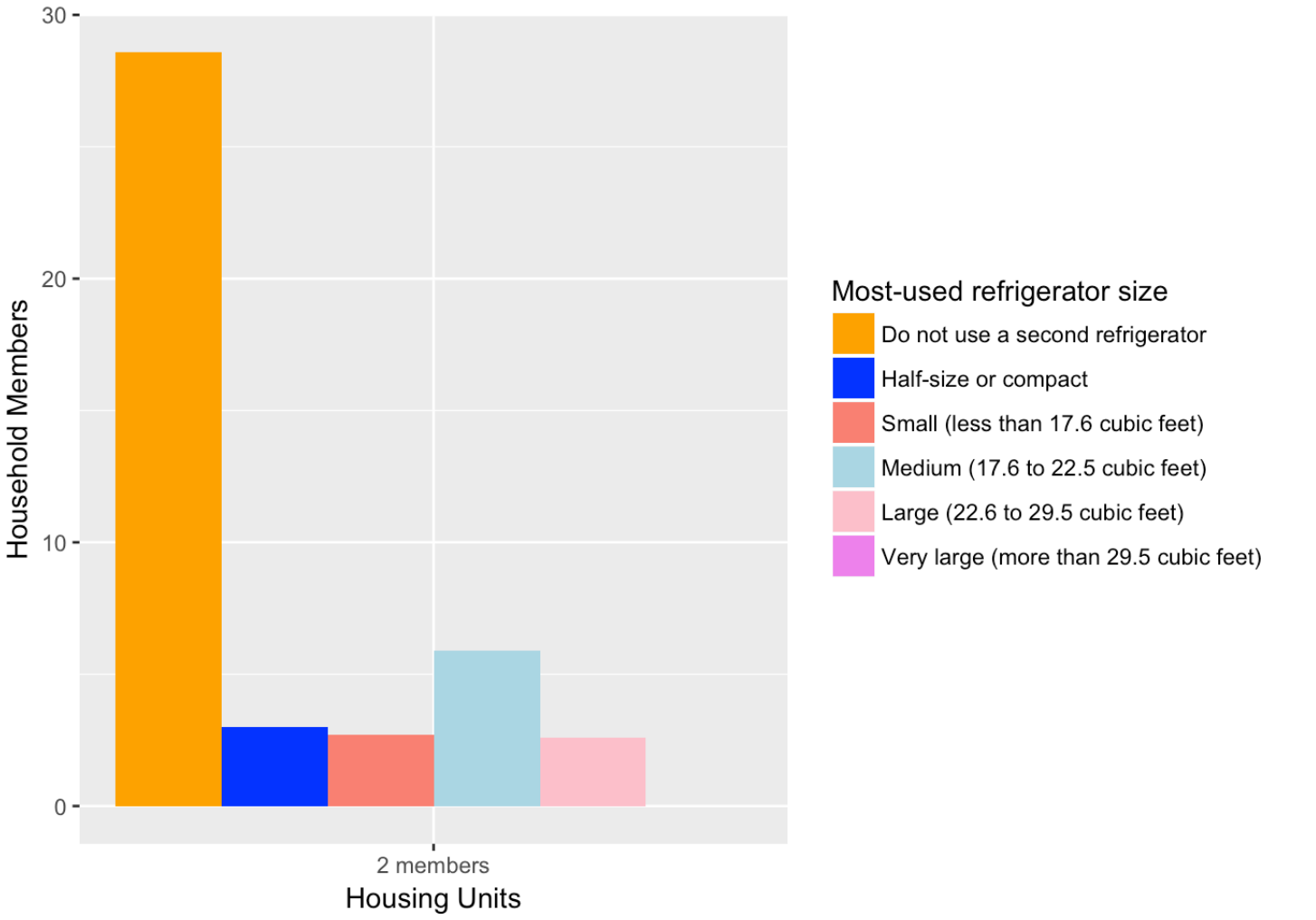
According to the plot above, households with only 1 member tend to cook hot meals less than once a week or no hot meals at all. For households with 2 members, they cooked relatively frequent than households with more than 2 members. In most cases, the cooking materials for hot meals need to be stored in refrigerators, which means people who cook hot meals more, would use refrigerators more often. This corresponds to the result we have: households with 2 members are the biggest group of people owning refrigerators. Therefore, the company should still focus on 2 members household group.

```
sd5 <- d2 %>%
  filter(Appliances == "Second refrigerator" & Housing_units == "2 members")

sd5 %>%
  ggplot(aes(x=Housing_units, y=Household_members, fill=Frequency)) +
  geom_bar(width = 1, stat="identity") +
  coord_polar("y") +
  theme_void() +
  geom_text(aes(label = percent(Household_members/sum(Household_members))))
```



Since all of the results from previous plots have showed us that 2 members households have the highest frequency of using refrigerators, we decide to take a closer look at their proportion of whether they use a second refrigerators or not. Based on the pie chart, one third of them do own a second refrigerator.

```
sd4 <- d2 %>%
  filter(d2$Appliances == "Second refrigerator size" & Housing_units == "2 members")
sd4$Frequency <- factor(sd4$Frequency, levels = c("Do not use a second refrigerator",
"Half-size or compact", "Small (less than 17.6 cubic feet)", "Medium (17.6 to 22.5 cu
bic feet)", "Large (22.6 to 29.5 cubic feet)", "Very large (more than 29.5 cubic feet
)"))

ggplot(sd4, aes(x=sd4$Housing_units, y=sd4$Household_members, fill = sd4$Frequency))
+
  geom_bar(stat="identity",position='dodge') +
  scale_fill_manual(values=c("orange", "blue", "salmon", "lightblue", "pink", "violet
"),
                    name="Most-used refrigerator size",
                    labels=c("Do not use a second refrigerator", "Half-size or compac
t", "Small (less than 17.6 cubic feet)", "Medium (17.6 to 22.5 cubic feet)", "Large (
22.6 to 29.5 cubic feet)", "Very large (more than 29.5 cubic feet)")) +
  xlab("Housing Units") +
  ylab("Household Members")
```



Among the one third of the population who own a second refrigerator, the most popular size is medium. Therefore, we recommend medium size refrigerator to 2 members households.

Conclusion

To sum up, 2 members households will be our targer customer groups, which is consistent throughout our analyses. Moreover, since medium size refrigerator is the most popular type among 2 members households, we would recommend market this type to them, regardless of how many refrigerators they have.