Associations Between Refrigerators' Size, Type and Number of Household Members

Jinghan Cui, Siran Zhao, Sijian Li, Mengyun Li

10/9/2017

Refrigerater's Size and Household Memeber

data cleaning

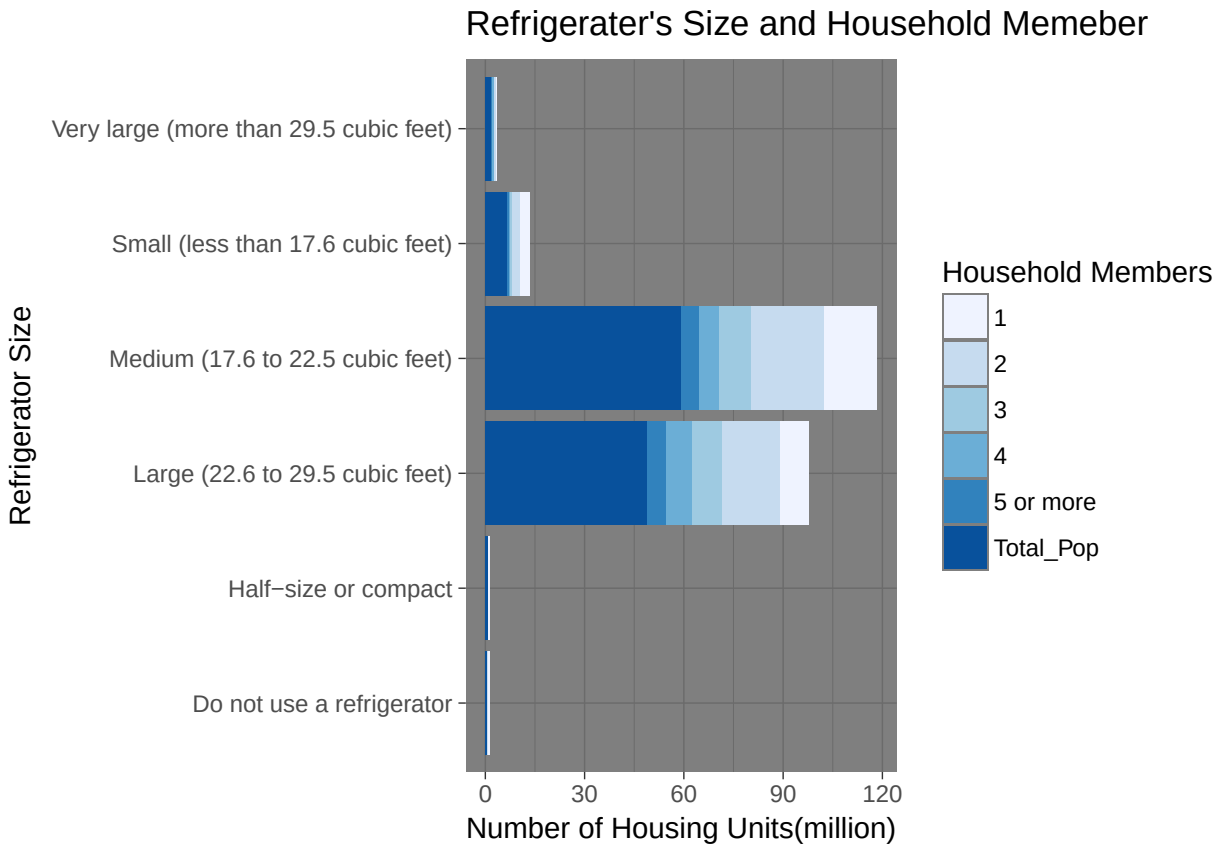
```
#import data
df <- read.csv("size.csv")
#mark insignificant as NA
df[df == "Q"] <- NA
df[df == "N"] <- NA
#set data as tibble
df <- as.tibble(df)
#rename varaibles
df <- rename(df, size=X, '1'=X1.member, '2'=X2.members, '3'=X3.members, '4'=X4.members, '5 or more'=X5.
#gather data for data clenning
df.tidy <- df %>%
  gather(`1`,`2`, `3`, `4`, `5 or more`,`Total_Pop`, key = "household", value = "units")
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

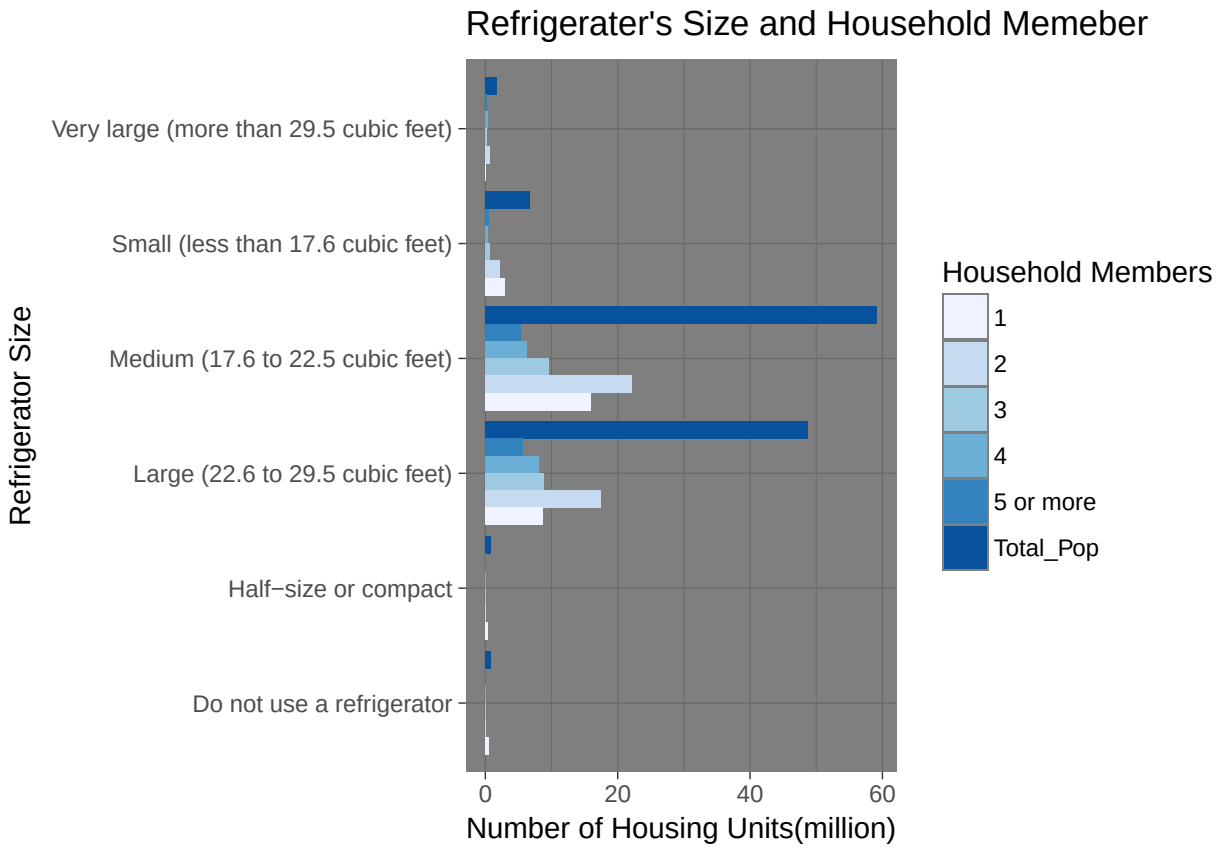
```
#set NA to 0 for graphing
df.tidy[is.na(df.tidy)] <- 0
#change clumns' properties
df.tidy[,1] <- sapply(df.tidy[,1],as.factor)
df.tidy[,2] <- sapply(df.tidy[,2],as.factor)
df.tidy[,3] <- sapply(df.tidy[,3],as.numeric)
```

EDA

```
ggplot()+
  geom_bar(aes(x=size,y=units,fill=household),data=df.tidy,stat="identity") +
  coord_flip() +
  scale_fill_brewer((pallete="Household Members")) +
  theme_dark() +
  labs(x = "Refrigerator Size", y = "Number of Housing Units(million)", title = "Refrigerater's Size an
```




```
ggplot()+
  geom_bar(aes(x=size,y=units,fill=household),data=df.tidy,stat="identity",position = "dodge") +
  coord_flip() + #set graph to horizontal
  theme(legend.position = "top") +
  scale_fill_brewer(palette="Blues") + #change color
  theme_dark() + #change background color
  labs(x = "Refrigerator Size", y = "Number of Housing Units(million)",fill="Household Members", title =
```



Refrigerator's Type and Household Member

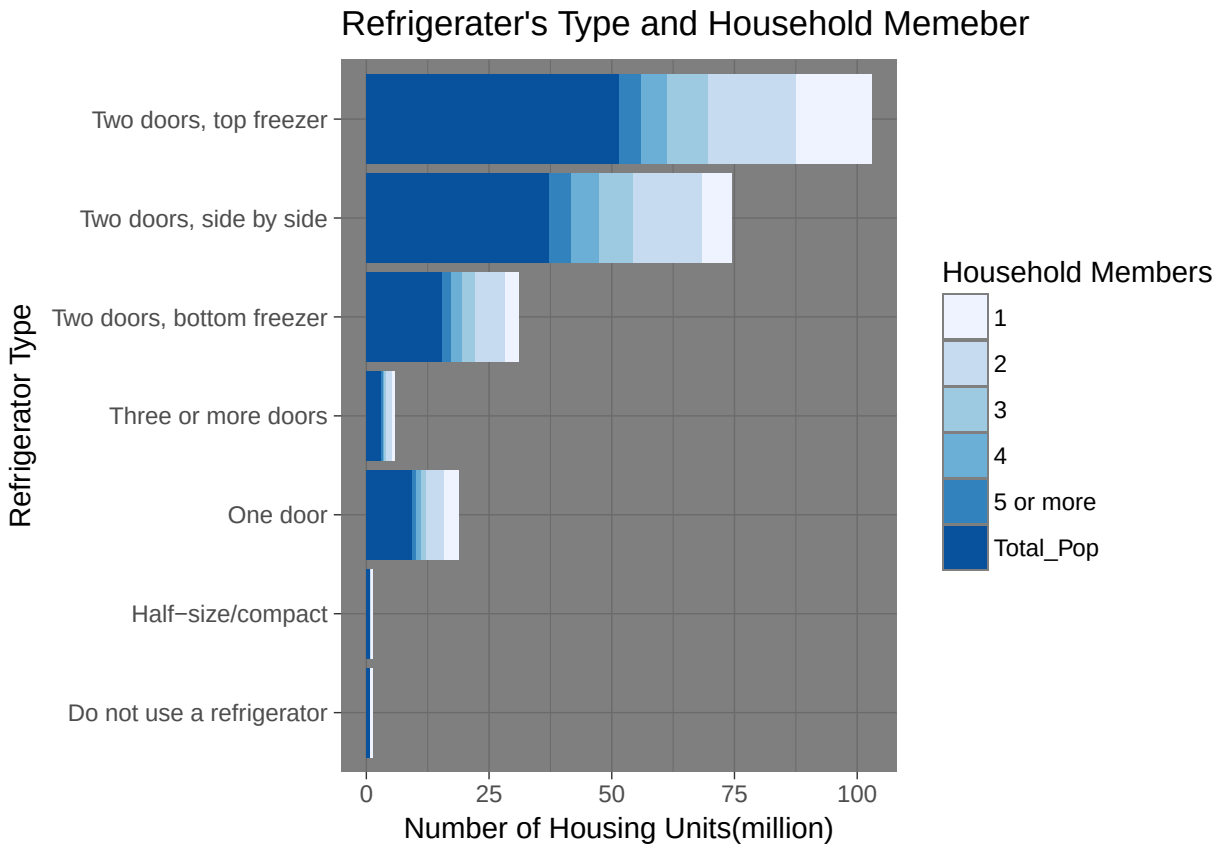
Data Cleaning

```
typmen <- read.csv("type.csv")
typmen <- rename(typmen,type=X,'1'=X1.member,'2'=X2.members,'3'=X3.members,'4'=X4.members, '5 or more'=
typmen.tidy <-typmen %>%
  gather('1','2','3','4','5 or more','Total_Pop', key="member", value= "household")#tidy the data
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

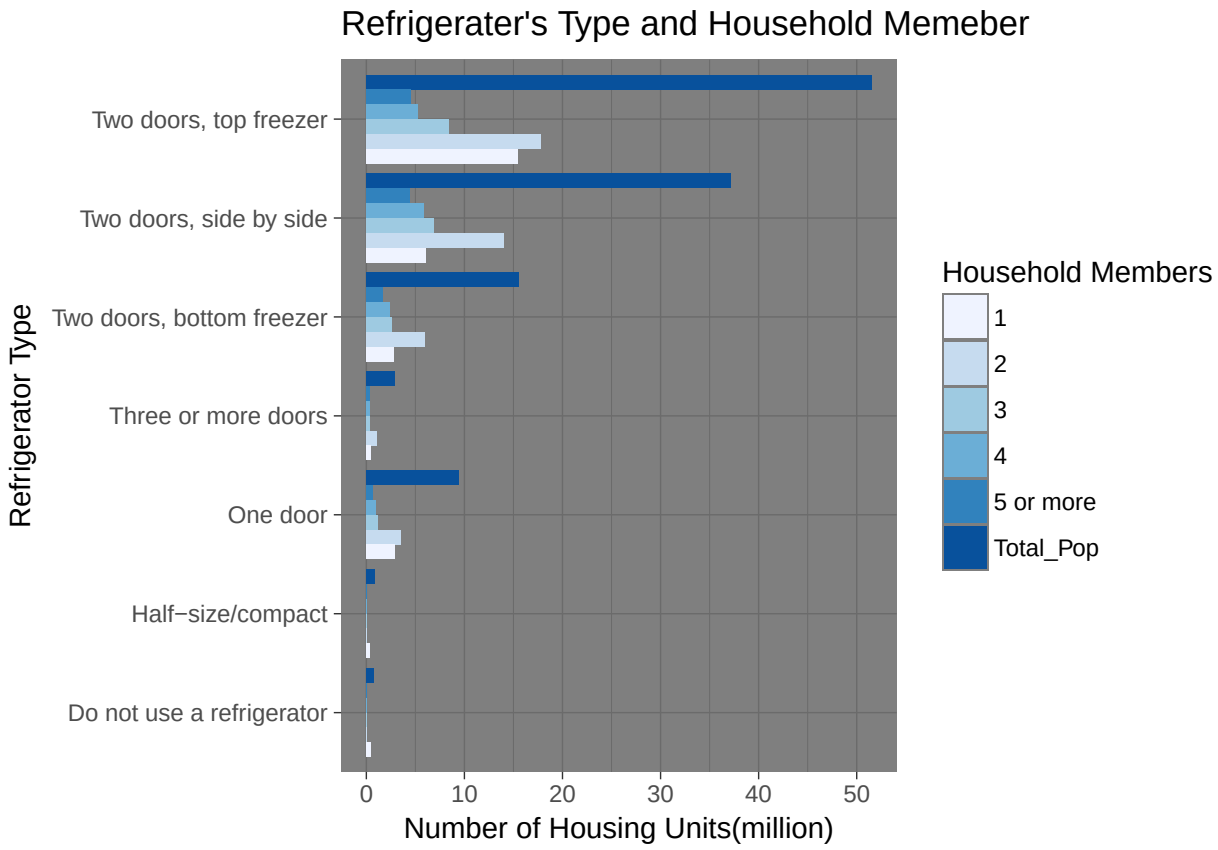
```
typmen.tidy$Total_Pop<-NULL #drop the column of useless data
typmen.tidy[typmen.tidy== 'Q']<-0 # change the not significant data into 0
typmen.tidy[typmen.tidy== 'N']<-0 # change the NA data into 0
typmen.tidy[,1] <- sapply(typmen.tidy[,1],as.factor) # transform type into factor
typmen.tidy[,2] <- sapply(typmen.tidy[,2],as.factor) # transform type into factor
typmen.tidy[,3] <- sapply(typmen.tidy[,3],as.numeric) # transform type into nymeric
```

```
ggplot()+
  geom_bar(aes(x=type,y=household,fill=member),data=typmen.tidy,stat='identity')+
  coord_flip() + #plot the bar plot of data
  scale_fill_brewer((pallete="Household Members")) +
  theme_dark() +
  labs(x = "Refrigerator Type", y = "Number of Housing Units(million)", title = "Refrigerater's Type an
```



EDA

```
ggplot(data=typmen.tidy)+
  geom_bar(mapping=aes(x=type,y=household,fill=member),stat = "identity",position = "dodge")+
  coord_flip() +
  theme(legend.position = "top") +
  scale_fill_brewer(palette="Blues") + #change color
  theme_dark() +
  labs(x = "Refrigerator Type", y = "Number of Housing Units(million)", fill = "Household Members",titl
```


Refrigerater's Type and Household Memeber
Refrigerator Type
Two doors, top freezer
Two doors, side by side
Two doors, bottom freezer
Three or more doors
One door
Half−size/compact
Do not use a refrigerator
0
10
20
30
40
50
Number of Housing Units(million)
Household Members
1
2
3
4
5 or more
Total_Pop