

A09 - CSC 285

Suthi de Silva

1/20/2024

Visualizing Proportions

Loading the data

```
shp <- read.csv(file="/srv/R/CSC285_public/Suthi /Data/saratoga-house-prices.txt",
                sep="\t", quote="", comment.char="")
```

Scenario

The data set saratoga-house-prices.txt contains data from homes in Saratoga, NY. The columns are described below:

- **Price** - Price of the home in thousands of dollars
- **Size** - Size of the house in thousands of square feet
- **Baths** - Number of bathrooms
- **Bedrooms** - Number of bedrooms
- **Fireplace** - Fireplace? Yes (1) / No (0)
- **Acres** - Size of lot in acres
- **Age** - Age of the home in years

You have just started a project with a Saratoga Realtor named Regina. She wants to have visuals to help her clients find their dream home.

For each question below, create the necessary visuals and describe your design choices as instructed within an RMarkdown.

Here are the first 6 rows of original data for your reference.

```
head(shp)
```

##	Price	Size	Baths	Bedrooms	Fireplace	Acres	Age
## 1	142.212	1.982	1.0	3	0	2.00	133
## 2	134.865	1.676	1.5	3	1	0.38	14
## 3	118.007	1.694	2.0	3	1	0.96	15
## 4	138.297	1.800	1.0	2	1	0.48	49
## 5	129.470	2.088	1.0	3	1	1.84	29
## 6	206.512	1.456	2.0	3	0	0.98	10

Questions:

1. Realtor Regina has noticed that many clients have a certain number of bedrooms they want. She wants some sort of a visual to show her clients the proportion of homes with 'x' number of bedrooms.

- a. Create a visual to show the proportion of homes with the different number of bedrooms

b. As a pie chart

ii. As a bar chart

```
# Creating a separate data frame "shp2" to get frequency for bedrooms
table(shp$Bedrooms)
```

```
##
```

```
##  1  2  3  4  5  6  7
```

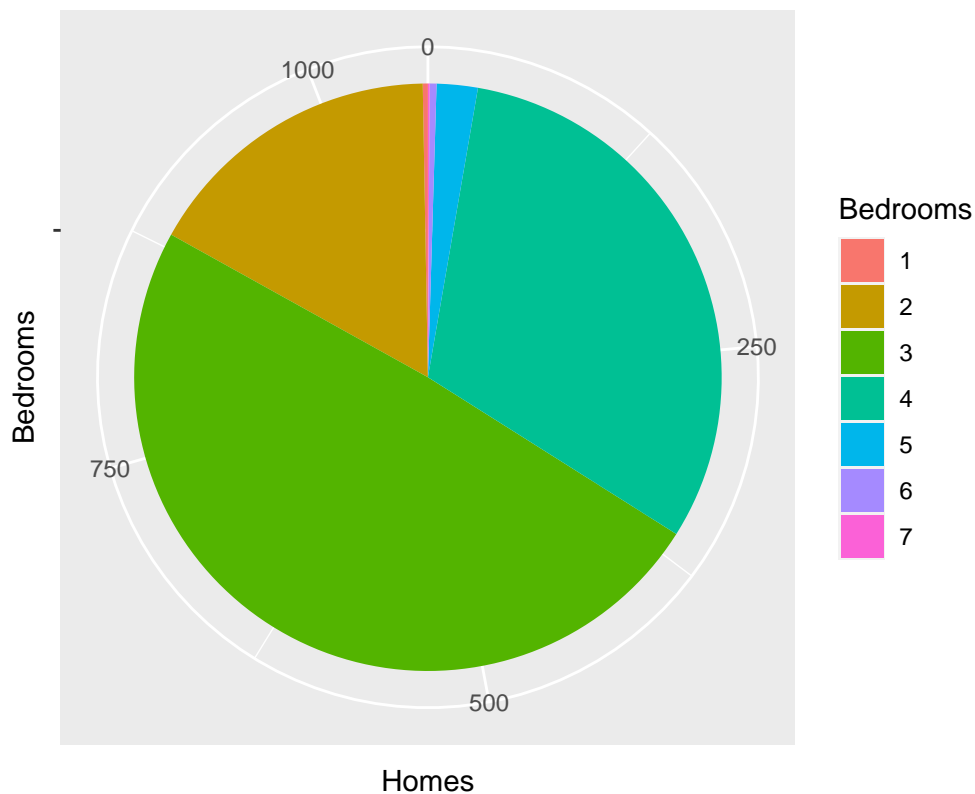
```
##  3 177 522 332  24  4  1
```

```
shp2 <- as.data.frame(table(shp$Bedrooms))
```

```
# As a pie chart
```

```
ggplot(shp2, aes(x="", y=Freq, fill= Var1)) +
  geom_bar(stat="identity", width=1) +
  labs( title = "Pie chart for homes with the different number of bedrooms",
        x = "Bedrooms",
        y = "Homes",
        fill = "Bedrooms" ) +
  coord_polar("y", start=0)
```

Pie chart for homes with the different number of bedrooms

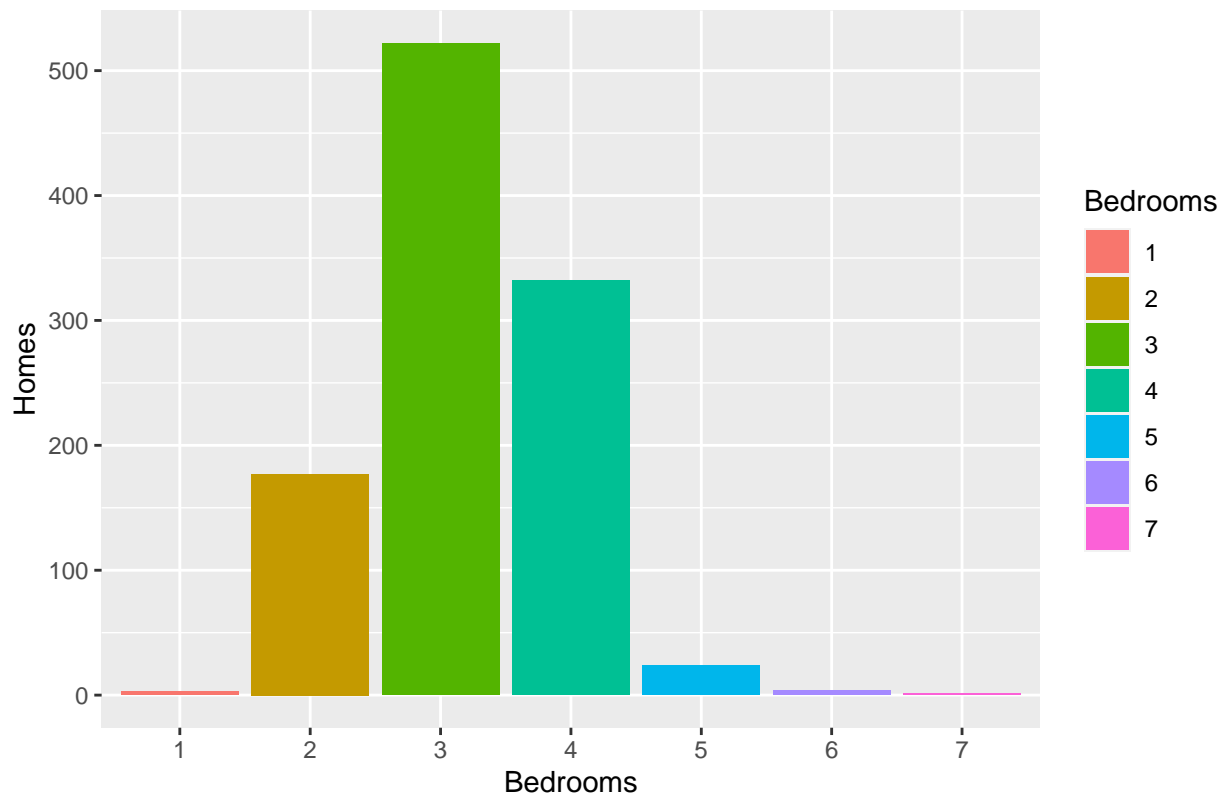


```
# As a bar chart
```

```
ggplot(shp2, aes(x= Var1, y = Freq, fill=as.factor(Var1))) +
  geom_bar(stat = "identity") +
  labs( title = "Barplot for homes with the different number of bedrooms",
        x = "Bedrooms",
        y = "Homes",
```

```
fill = "Bedrooms" )
```

Barplot for homes with the different number of bedrooms



b. Which visual do you recommend Regina use and why? *Bar chart, since it is easier to comprehend proportions with it.*

2. Realtor Regina wants another visual to show her clients that show the proportion of houses with all the possible bedroom/bath combinations.

- Create a visual that shows the proportion of homes with all the bedroom/bath combinations present in the data
- Multiple Pie Charts
 - Multiple stacked bars
 - Grouped bars

Here it does not make sense to have number of bathrooms with decimal values, so it would be rounded up.

```
# Rounding up the values
shp$Baths <- signif(shp$Baths, digits = 1)

# Creating a separate data frame "shp3" to get frequency for bedroom/bath combinations
table(shp$Bedrooms , shp$Baths)
```

```
##
##      1  2  3  4  5  6
##  1   3  0  0  0  0  0
##  2  65 110  2  0  0  0
##  3 114 405  3  0  0  0
```

```
## 4 14 280 16 21 0 1
## 5 1 17 3 3 0 0
## 6 1 2 0 1 0 0
## 7 0 0 0 0 1 0
```

```
shp3 <- as.data.frame(table(shp$Bedrooms , shp$Baths))
```

```
# As multiple pie charts
```

```
for (n in list("1","2","3","4","5","6")) {
```

```
  print(ggplot(shp3[c(which(shp3$Var2 == n)),], aes(x="", y= Freq, fill=as.factor(Var1))) + geom_bar(sta
```

```
    labs( title = glue("Pie chart for {n} Bath homes for each number of bedrooms"),
```

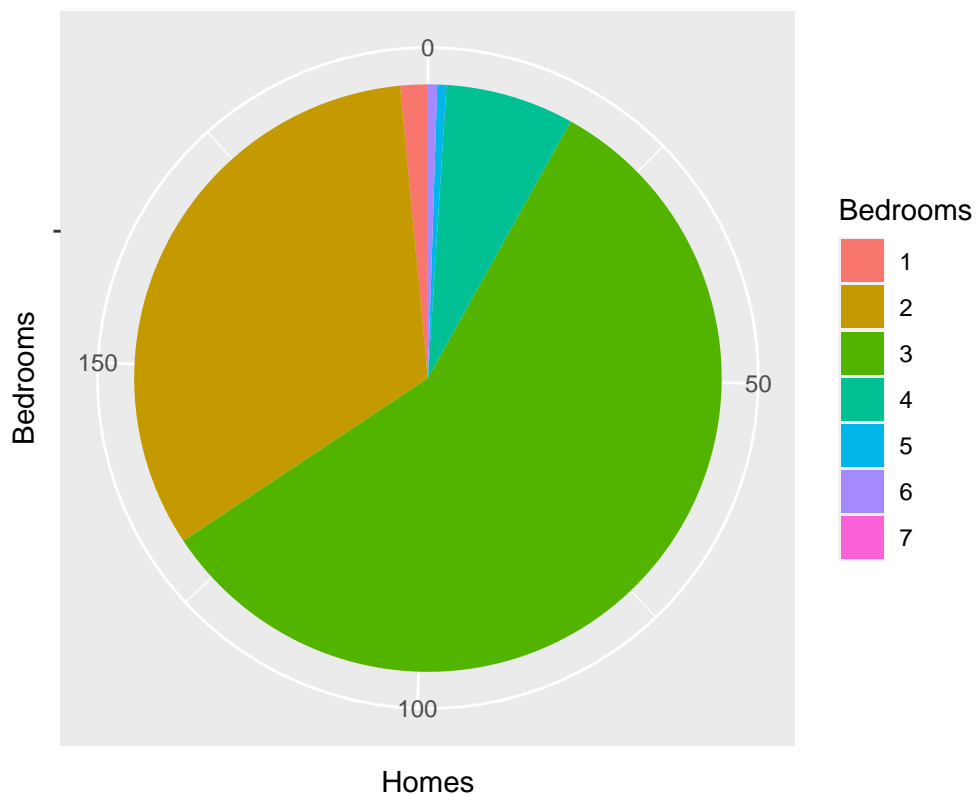
```
          x = "Bedrooms",
```

```
          y = "Homes",
```

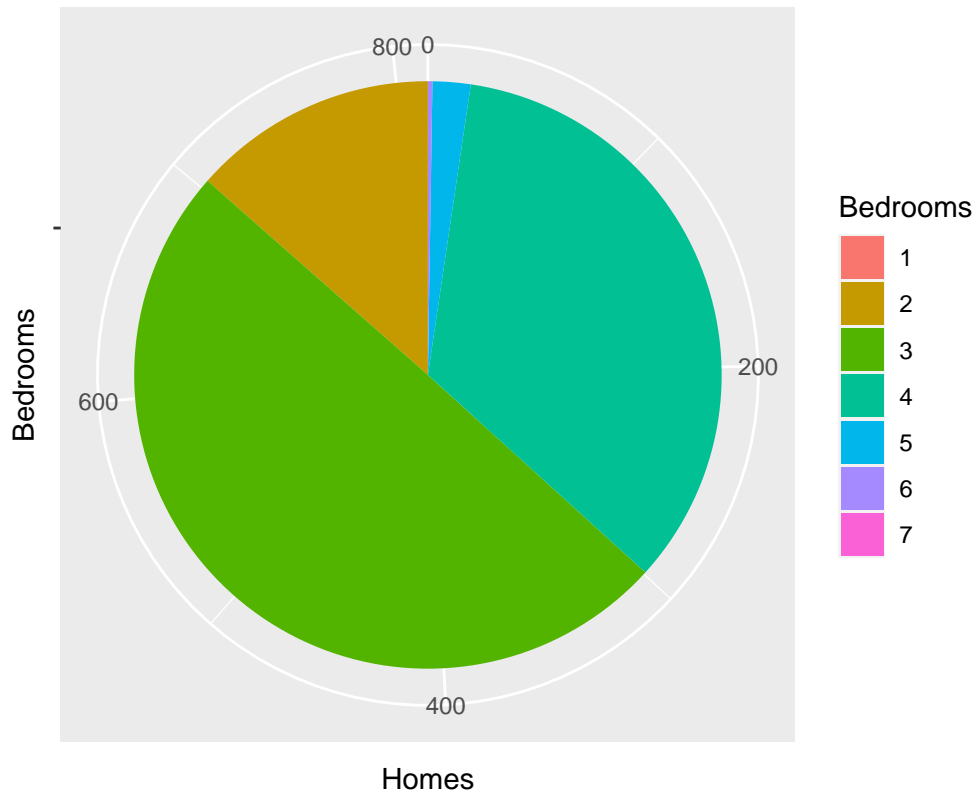
```
          fill = "Bedrooms" ) + coord_polar("y", start=0) )
```

```
}
```

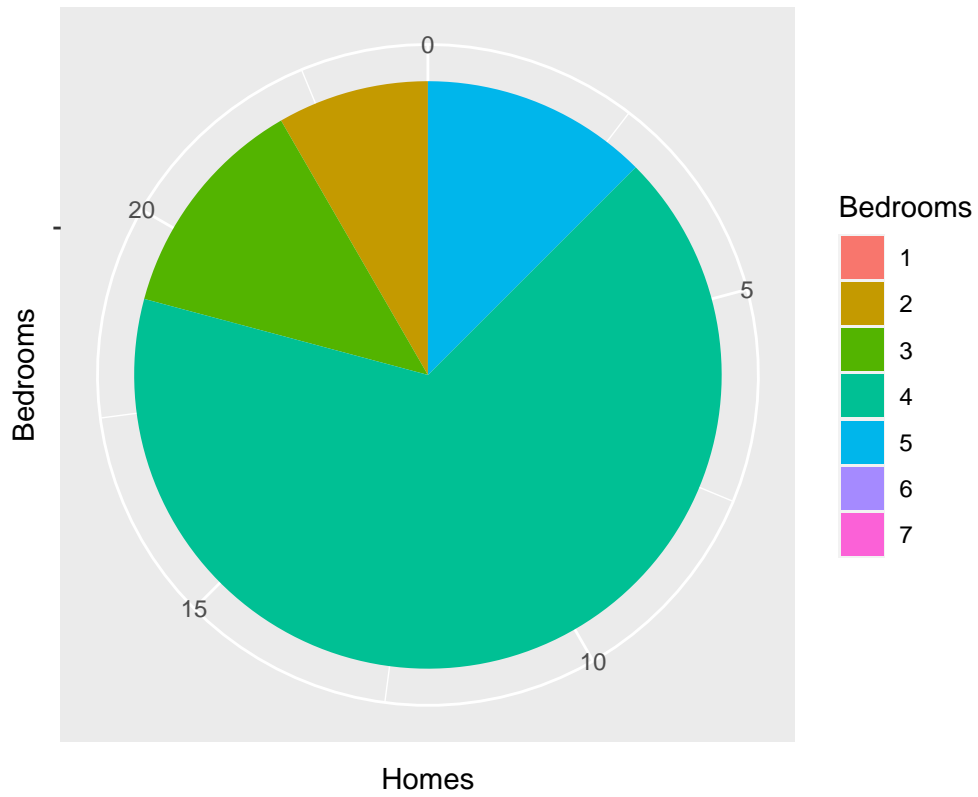
Pie chart for 1 Bath homes for each number of bedrooms



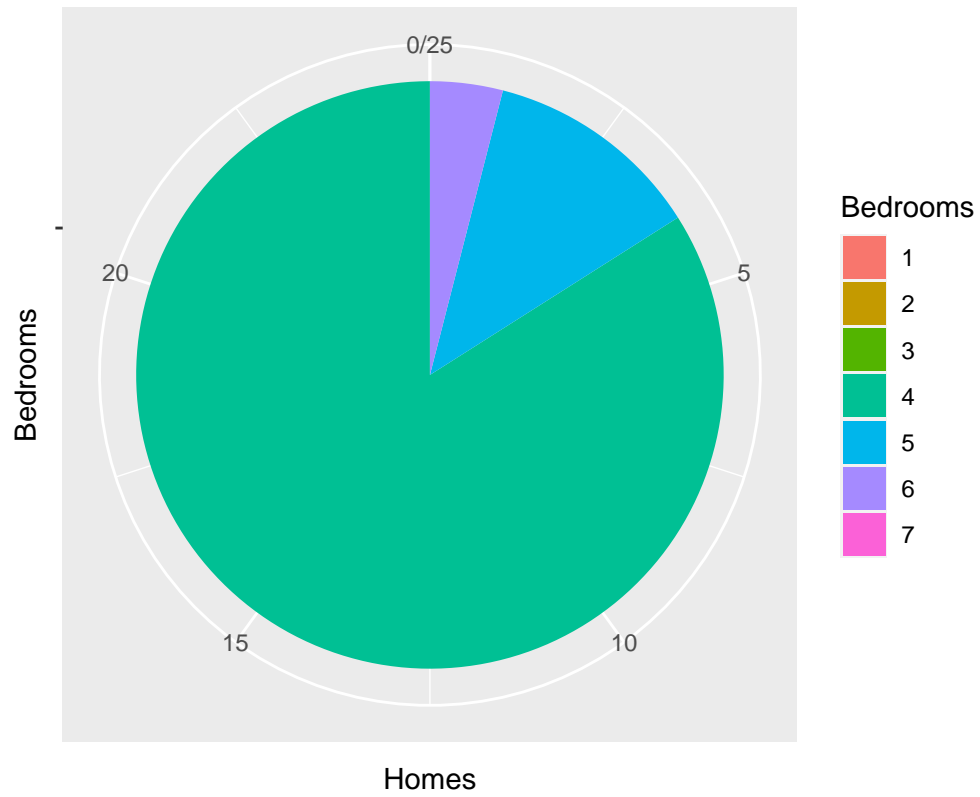
Pie chart for 2 Bath homes for each number of bedrooms



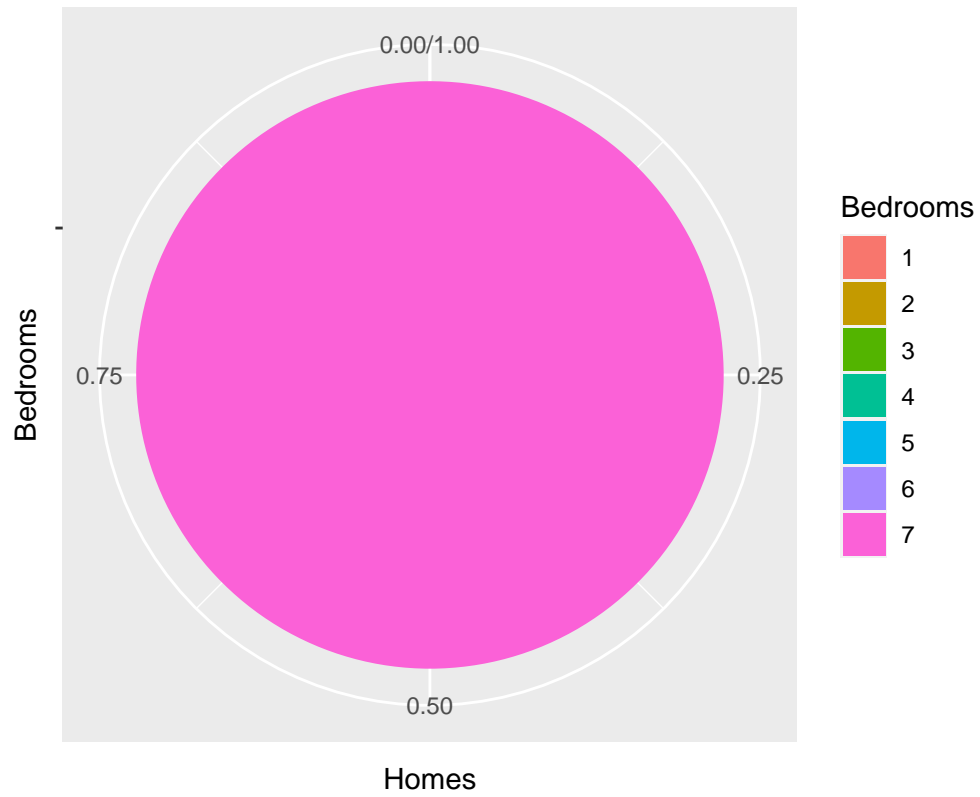
Pie chart for 3 Bath homes for each number of bedrooms



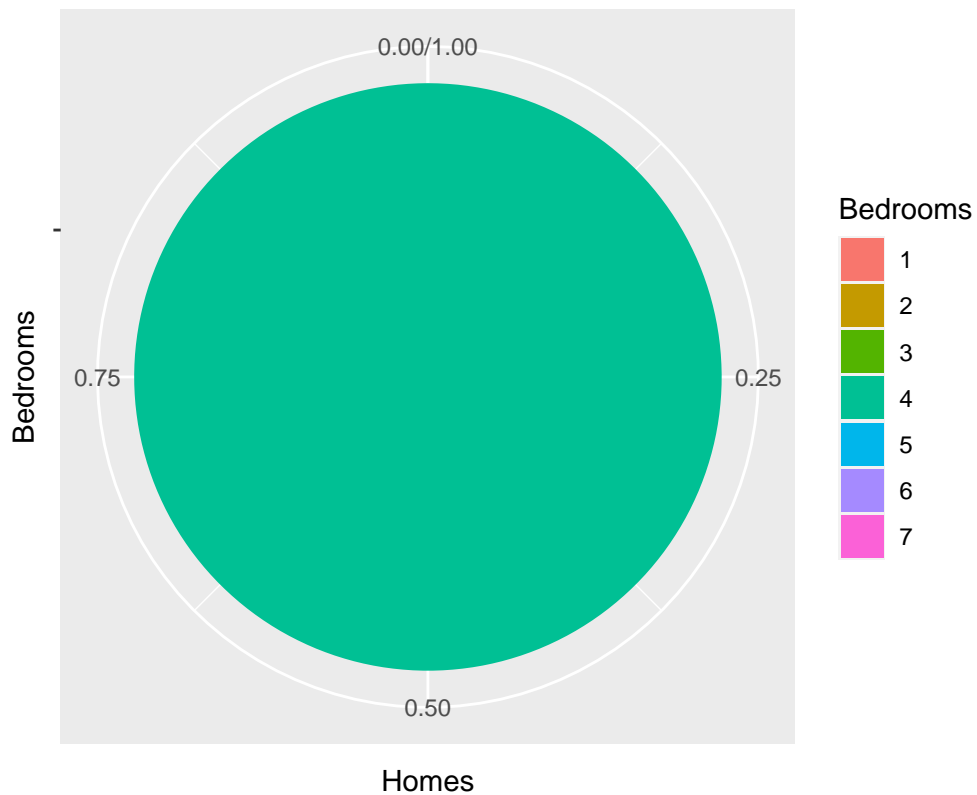
Pie chart for 4 Bath homes for each number of bedrooms



Pie chart for 5 Bath homes for each number of bedrooms

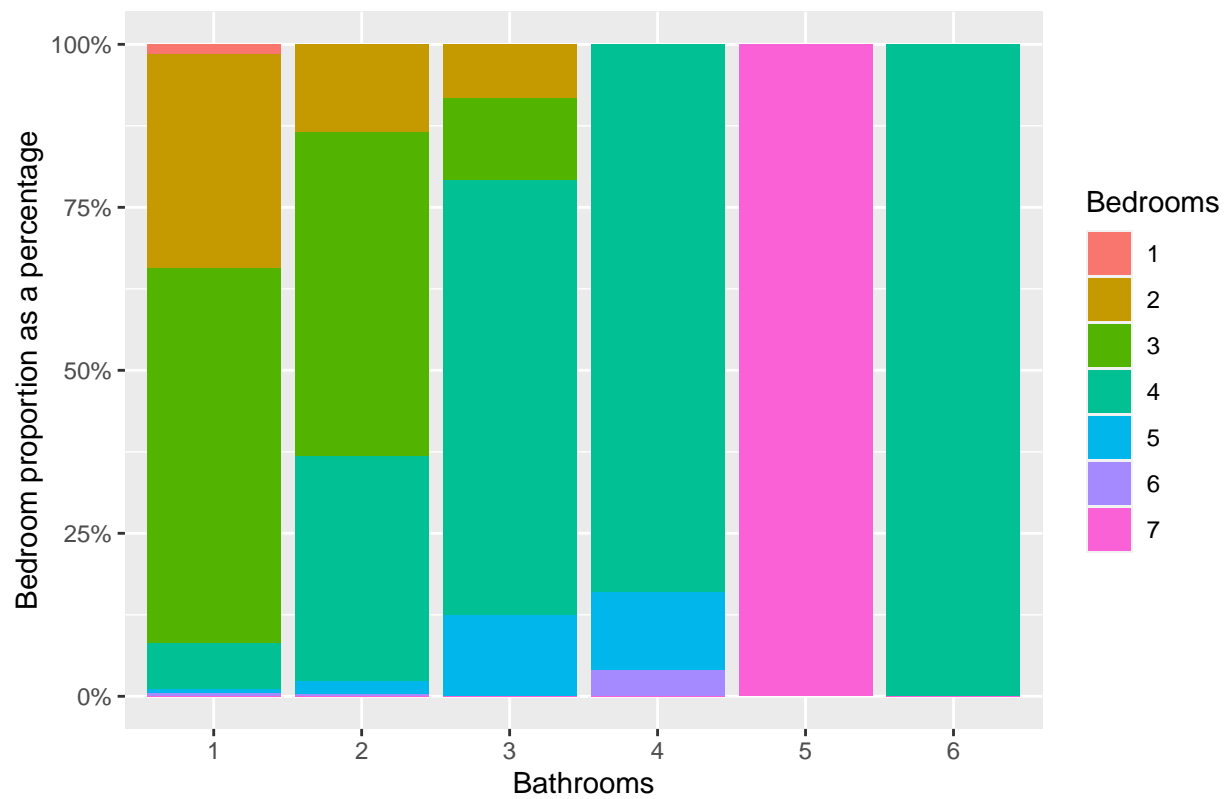


Pie chart for 6 Bath homes for each number of bedrooms



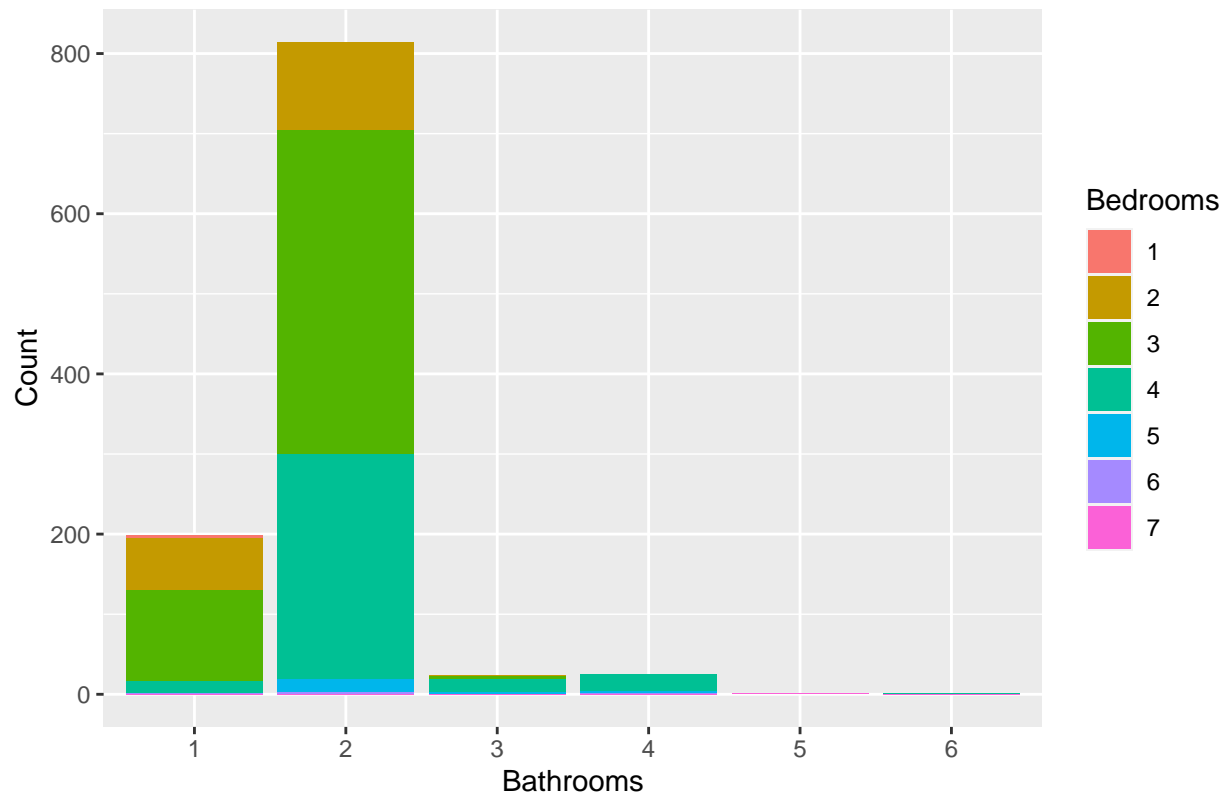
```
# As stacked bar plots
ggplot(shp3, aes(Var2, Freq, fill = Var1)) +
  geom_bar(position = "fill", stat = "identity") +
  scale_y_continuous(labels = percent) +
  labs( title = "Stacked bar plot for bedrooms/bathrooms combinations (As a percentage)",
        x = "Bathrooms",
        y = "Bedroom proportion as a percentage",
        fill = "Bedrooms" )
```

Stacked bar plot for bedrooms/bathrooms combinations (As a percentage)

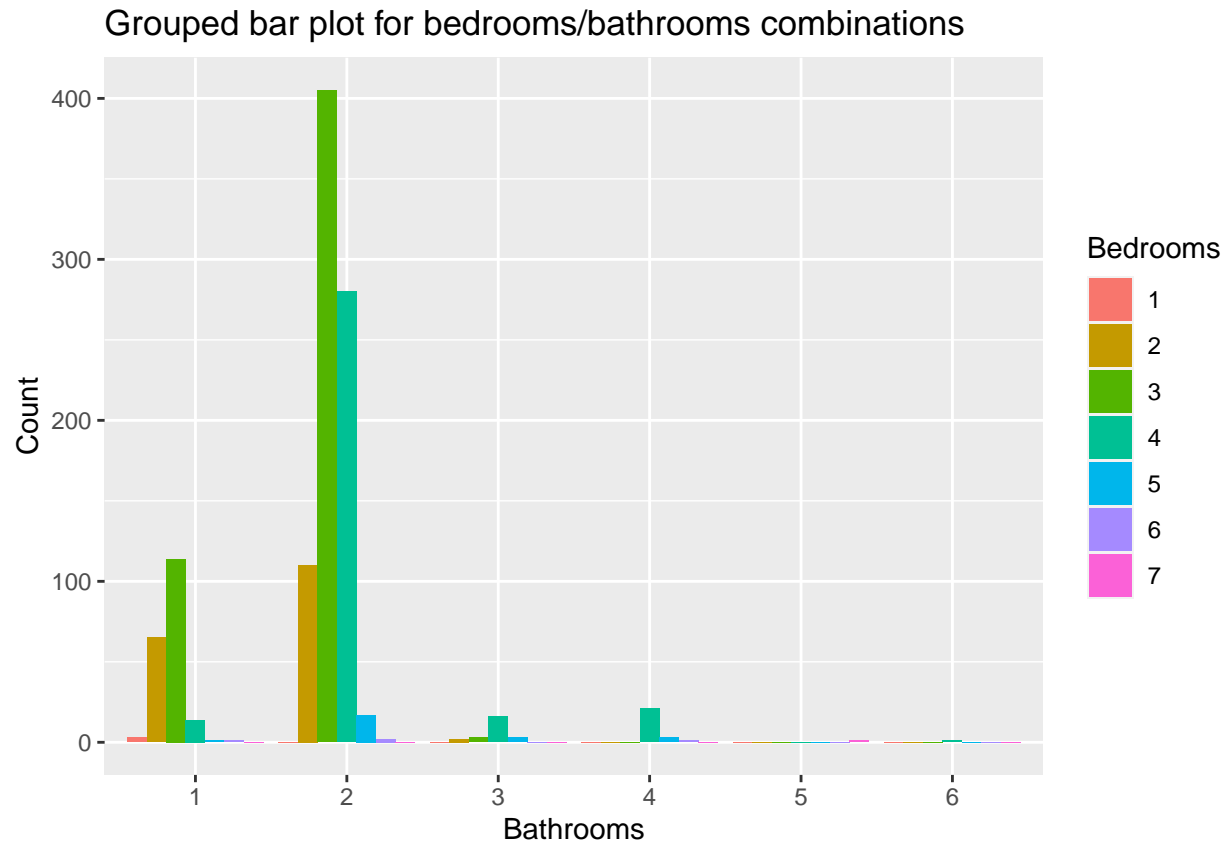


```
ggplot(shp3, aes(fill=Var1, y=Freq, x=Var2)) +
  geom_bar(position="stack", stat="identity") +
  labs( title = "Stacked bar plot for bedrooms/bathrooms combinations",
        x = "Bathrooms",
        y = "Count",
        fill = "Bedrooms" )
```

Stacked bar plot for bedrooms/bathrooms combinations



```
# As grouped bars
ggplot(shp3, aes(fill=Var1, y=Freq, x=Var2)) +
  geom_bar(position="dodge", stat="identity") +
  labs( title = "Grouped bar plot for bedrooms/bathrooms combinations",
        x = "Bathrooms",
        y = "Count",
        fill = "Bedrooms" )
```



b. Which would you bring forward – i, ii, or iii? Why? *iii grouped bars, since it represents correct proportions and count for all bedrooms/bathrooms combinations*

3. Realtor Regina also wants to show clients the proportion of homes that have a fireplace based on the age of the home.

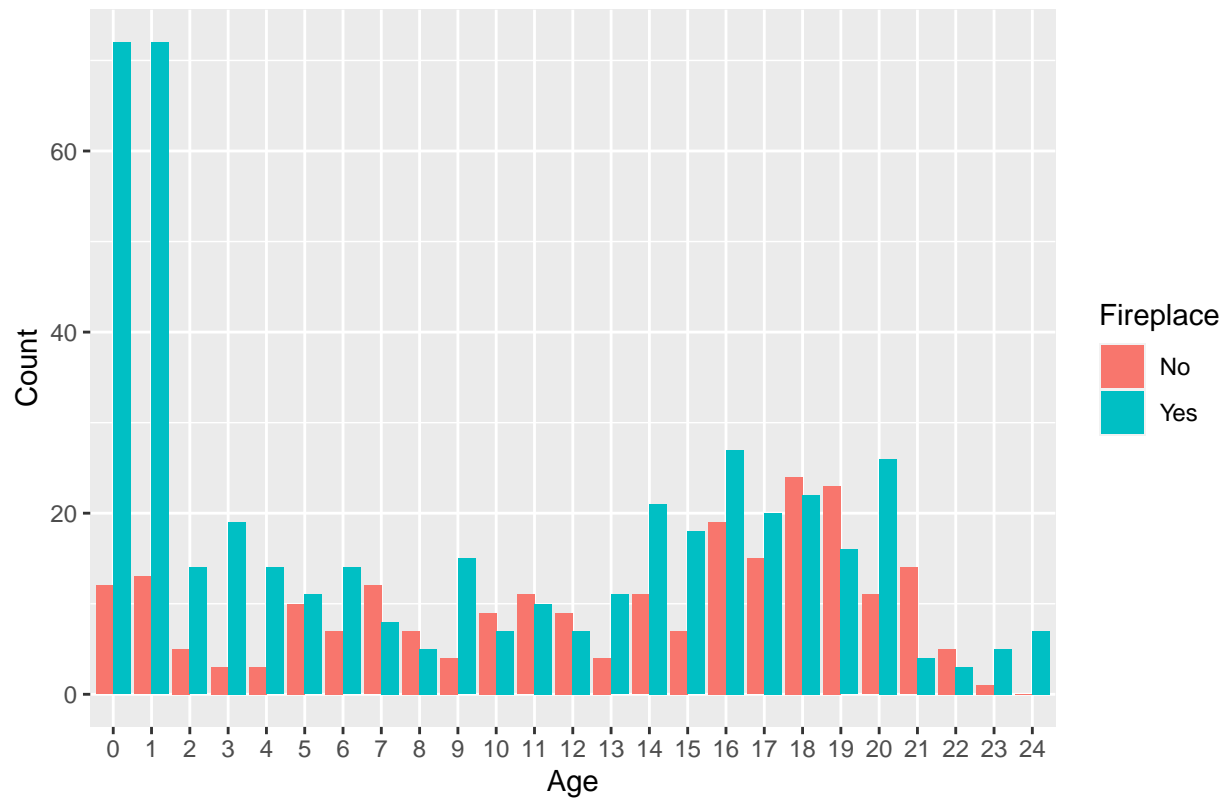
a. Create a visual (you choose the type!) to help Regina show this trend to her clients.

b. Describe your design choices.

```
# Creating a separate data frame "shp4" to get frequency for fireplace/age combinations
table(shp$Fireplace, shp$Age)
shp4 <- as.data.frame(table(shp$Fireplace, shp$Age))
```

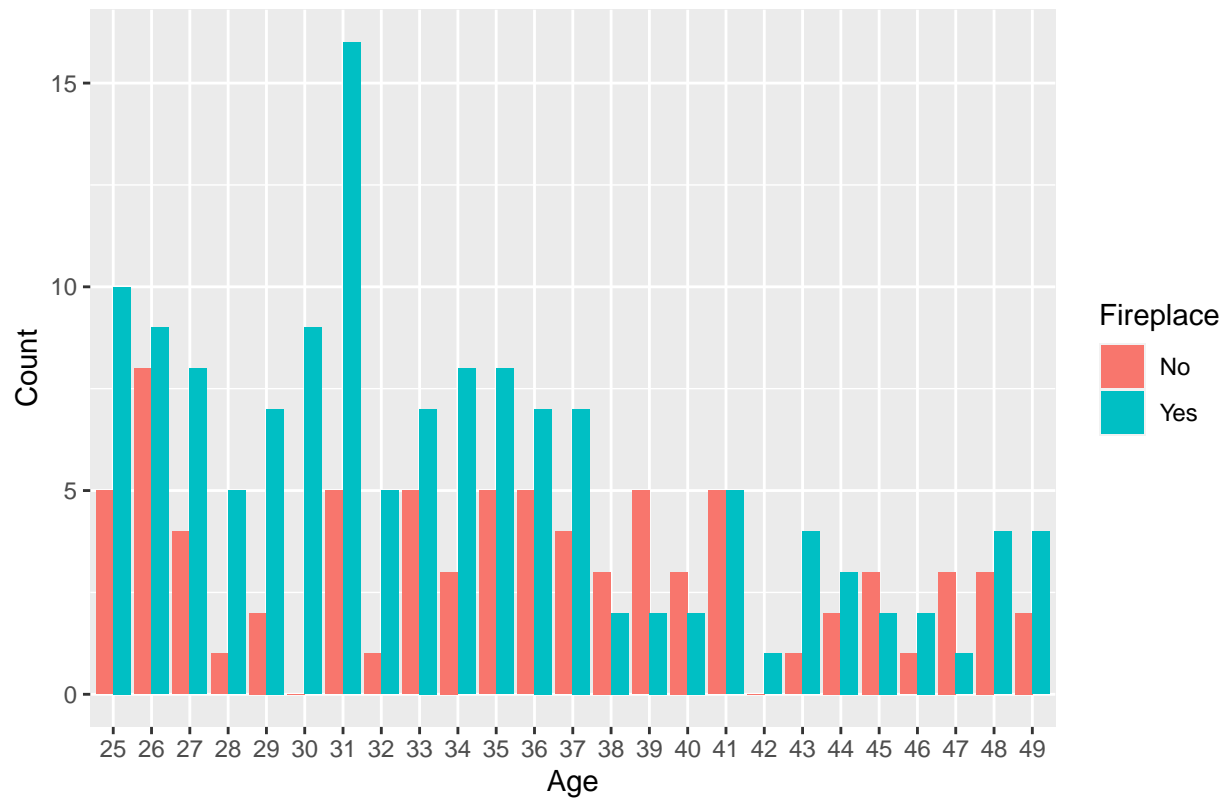
```
# As 6 groups of age grouped bars
# Age 0 to 24
ggplot(shp4[0:50,], aes(fill=Var1, y=Freq, x=Var2)) +
  geom_bar(position="dodge", stat="identity") +
  labs( title = "Grouped bar plot for fireplace/age combinations",
        x = "Age",
        y = "Count") +
  scale_fill_discrete(name = "Fireplace", labels = c("No", "Yes"))
```

Grouped bar plot for fireplace/age combinations



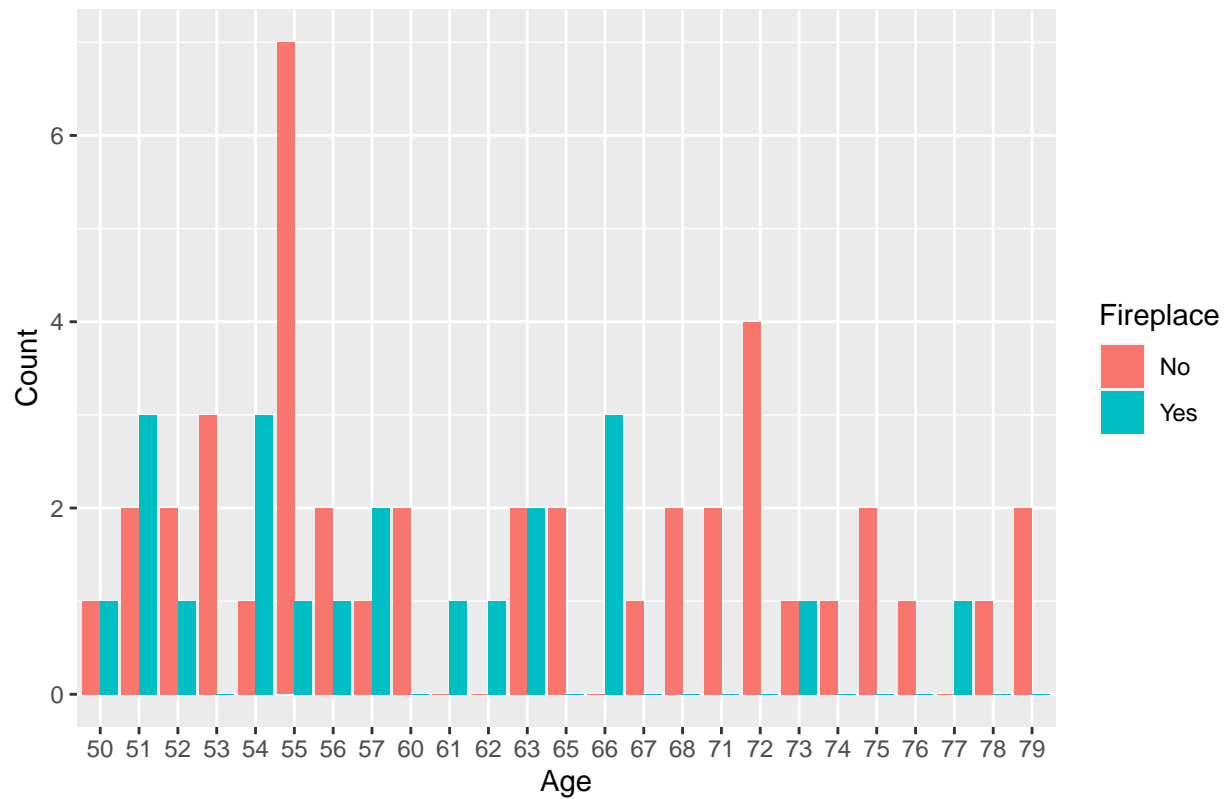
```
# Age 25 to 49
ggplot(shp4[51:100,], aes(fill=Var1, y=Freq, x=Var2)) +
  geom_bar(position="dodge", stat="identity") +
  labs( title = "Grouped bar plot for fireplace/age combinations",
        x = "Age",
        y = "Count") +
  scale_fill_discrete(name = "Fireplace", labels = c("No", "Yes"))
```

Grouped bar plot for fireplace/age combinations

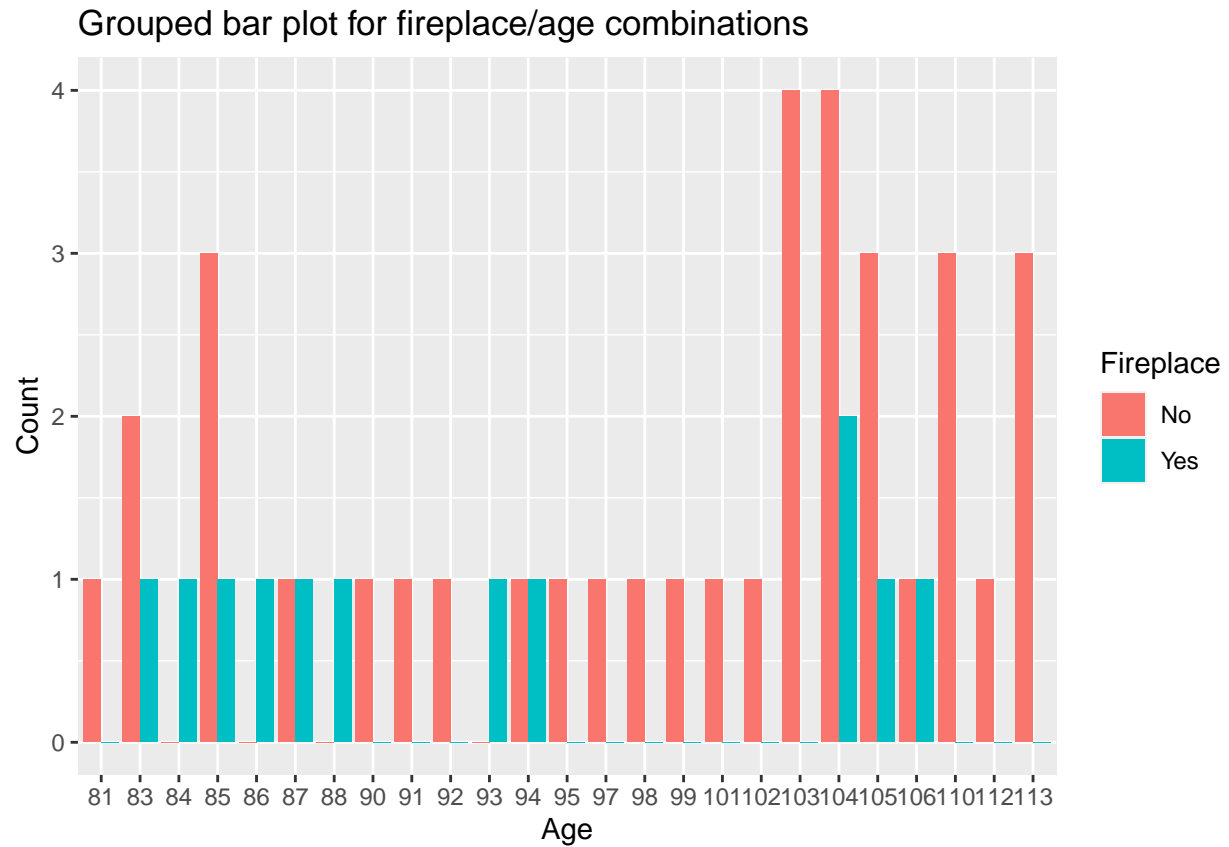


```
# Age 50 to 79
ggplot(shp4[101:150,], aes(fill=Var1, y=Freq, x=Var2)) +
  geom_bar(position="dodge", stat="identity") +
  labs( title = "Grouped bar plot for fireplace/ages combinations",
        x = "Age",
        y = "Count") +
  scale_fill_discrete(name = "Fireplace", labels = c("No", "Yes"))
```

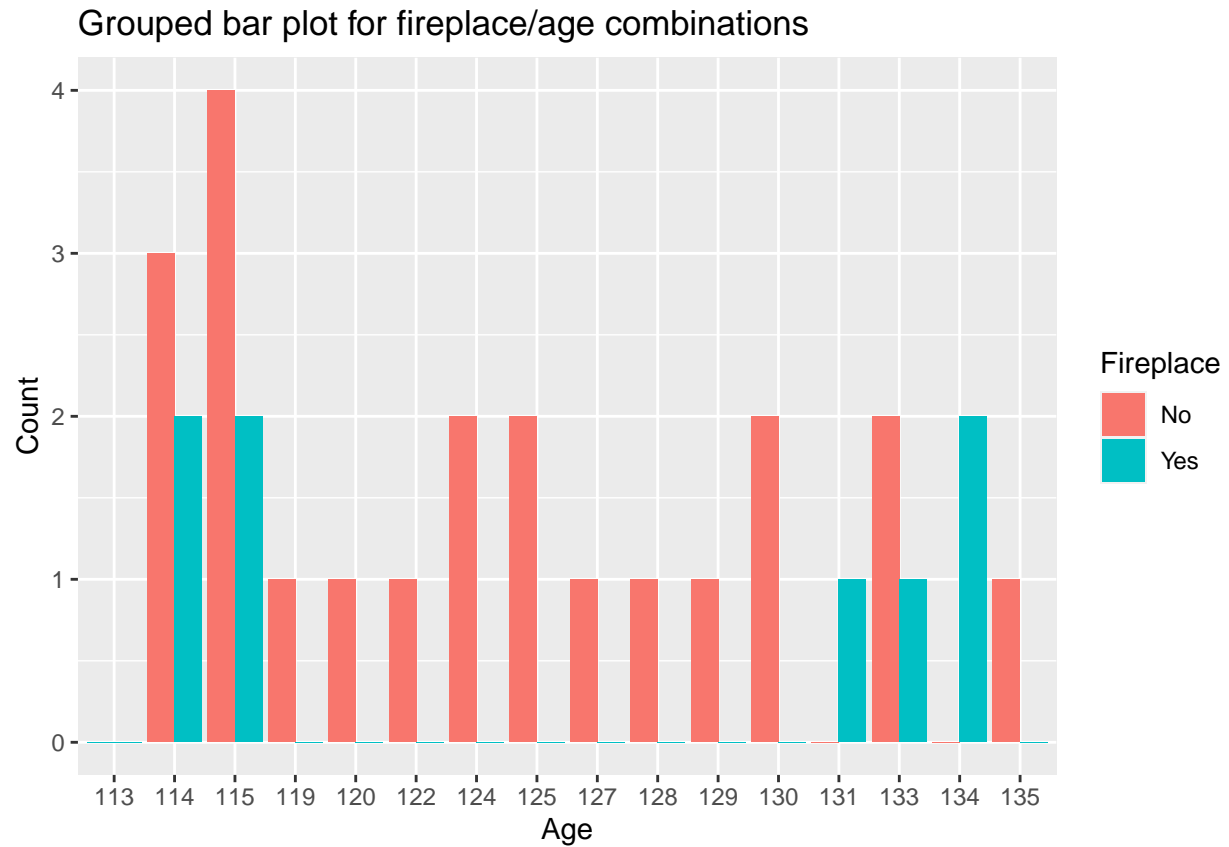
Grouped bar plot for fireplace/ages combinations



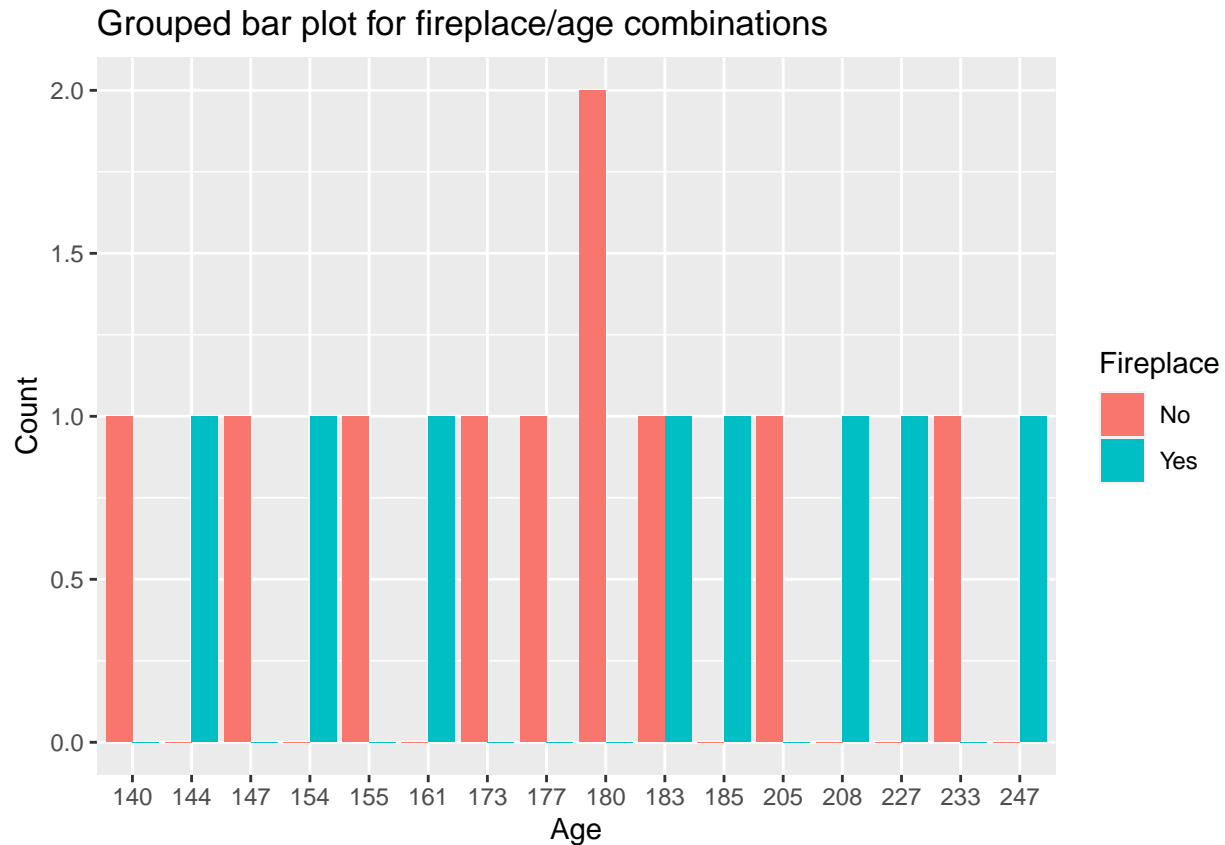
```
# Age 80 to 113
ggplot(shp4[151:200,], aes(fill=Var1, y=Freq, x=Var2)) +
  geom_bar(position="dodge", stat="identity") +
  labs( title = "Grouped bar plot for fireplace/age combinations",
        x = "Age",
        y = "Count") +
  scale_fill_discrete(name = "Fireplace", labels = c("No", "Yes"))
```



```
# Age 114 to 135
ggplot(shp4[200:230,], aes(fill=Var1, y=Freq, x=Var2)) +
  geom_bar(position="dodge", stat="identity") +
  labs( title = "Grouped bar plot for fireplace/age combinations",
        x = "Age",
        y = "Count") +
  scale_fill_discrete(name = "Fireplace", labels = c("No", "Yes"))
```

```
# Age 136 to 247
ggplot(shp4[231:262,], aes(fill=Var1, y=Freq, x=Var2)) +
  geom_bar(position="dodge", stat="identity") +
  labs( title = "Grouped bar plot for fireplace/age combinations",
        x = "Age",
        y = "Count") +
  scale_fill_discrete(name = "Fireplace", labels = c("No", "Yes"))
```



I used six different age groups because the age range varies from 0 to 247, and representing in one plot would make it unclear. I used grouped bars, since it represents correct proportions and count for all fireplace/age combinations, and two colors to show fireplace availability

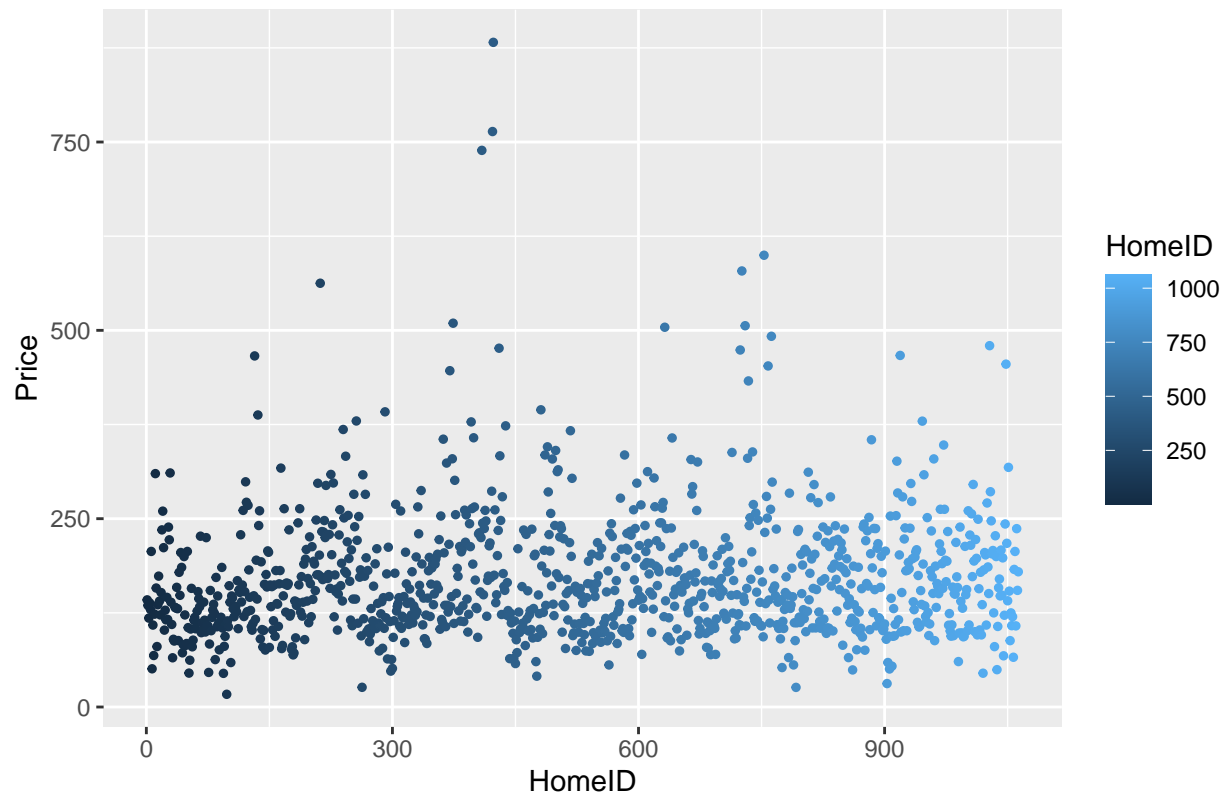
4. Realtor Regina is also interested in showing her customers the distribution of home prices in Saratoga.

- Create a visual (you choose the type!) to help Regina show the distribution to her clients.
- Describe your design choices.

```
# Creating a ID for each home as variable called "HomeID"
shp$HomeID <- 1:nrow(shp)
shp %>% relocate(HomeID, .before= Price)

# Creating scatter plot to represent the distribution of home prices for each "HomeID"
ggplot(shp, aes(x= HomeID, y= Price, color = HomeID)) +
  geom_point(size=1) +
  labs( title = "Scatter plot for distribution of home prices for each HomeID")
```

Scatter plot for distribution of home prices for each HomeID



I created a specific HomeID for each home with different prices in order to better represent the distribution of home prices for each HomeID. We used a scatter plot, in order to clearly see how the prices have fluctuated for different homes.