

```

library(ggplot2)
library(tidyverse)
library(dplyr)

# Load data
rm(list=ls())
df<-read.csv(file.choose())

# Check data structure and summary
dim(df)
[1] 3164 6
> str(df)
'data.frame':      3164 obs. of  6 variables:
 $ Id      : Factor w/ 2967 levels "", "C4230142",...: 2317 2654 2536 355 2460 2311 2389 2535 2284 2829
 ...
 $ Address : Factor w/ 2946 levels "", "1 Aberfoyle Cres 1109, Toronto",...: 2163 1171 1348 1310 950 162
 7 2675 1347 925 2227 ...
 $ Bedrooms: int  1 1 1 3 1 1 5 1 1 1 ...
 $ Bathrooms: int  1 1 1 1 1 1 3 1 1 1 ...
 $ Type     : Factor w/ 99 levels "", "Att/Row/Twnhouse 2-Storey",...: 75 4 81 33 28 19 33 81 88 19 ...
 $ Price    : int  650 700 700 799 800 800 800 800 950 1000 ...
> summary(df)
      Id      Address      Bedrooms      Bathrooms      Type
C4311344: 2  101 Peter St 516, Toronto  : 3  Min. :1.0  Min. :1.000  Condo Apt Apartment  :
2066
C4320832: 2  18 Kenaston Gdns 1605, Toronto: 3  1st Qu.:1.0  1st Qu.:1.000  Detached 2-Storey
: 183
C4322238: 2  55 Stewart St 932, Toronto  : 3  Median :2.0  Median :1.000  Detached Bungalow
: 112
C4327202: 2  65 St Mary St 2503, Toronto  : 3  Mean   :2.1  Mean   :1.626  Comm Element Condo
Apartment: 56
C4327328: 2  1 Arundel Ave Main, Toronto  : 2  3rd Qu.:3.0  3rd Qu.:2.000  Semi-Detached 2-Store
y : 54
C4329247: 2  1 Bloor St E 1603, Toronto  : 2  Max.   :8.0  Max.   :8.000  Condo Townhouse 3-Stor
ey : 48
(Other) :3152 (Other)                  :3148 NA's :1  NA's :1  (Other)                  : 645
Price
Min.   : 650
1st Qu.: 2150
Median : 2500
Mean   : 3001
3rd Qu.: 3200
Max.   :22500
NA's   :1

# Change datatype
df$Price <- as.numeric(df$Price)
df$Bedrooms <- as.numeric(df$Bedrooms)
df$Bathrooms <- as.numeric(df$Bathrooms)
df$Type <- as.character(df$Type)
df$Address <- as.character(df$Address)
df$Id <- as.character(df$Id)

```

```

# Check duplicates and remove duplicates
duplicated(df$Id)
df <- df[!duplicated(df$Id), ]
dim(df)

# Checking missing values and remove them
colSums(is.na(df)|df=="")
df<-df[complete.cases(df),]

# Stats information about the Price,Bedrooms,Bathrooms after duplicates removed
dim(df)
summary(df$Price)
summary(df$Bedrooms)
summary(df$Bathrooms)

> dim(df)
[1] 2966      6
> summary(df$Price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   650    2100    2475    3022    3200    22500
> summary(df$Bedrooms)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  1.000  2.000  2.104  3.000  8.000
> summary(df$Bathrooms)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  1.000  1.000  1.634  2.000  8.000

# Change datatype of Bedrooms and Bathrooms for plotting
df$Bedrooms <- as.character(df$Bedrooms)
df$Bathrooms <- as.character(df$Bathrooms)

# Count the total number of properties by type
df %>% group_by(Type) %>% summarize(count=n())

Type                                count
  <chr>                                <int>
1 Att/Row/Twnhouse 2-Storey           21
2 Att/Row/Twnhouse 2 1/2 Storey         3
3 Att/Row/Twnhouse 3-Storey           32
4 Att/Row/Twnhouse Apartment           3
5 Att/Row/Twnhouse Other               2
6 Co-Op Apt Apartment                 6
7 Co-Ownership Apt 2-Storey           1
8 Co-Ownership Apt Apartment           1
9 Co-Ownership Apt Bachelor/Studio     1
10 Comm Element Condo 2-Storey         1
# ... with 88 more rows
# Total number of Type of Properties : 98

```

# List unique Type : total 98 types

unique(df\$Type)

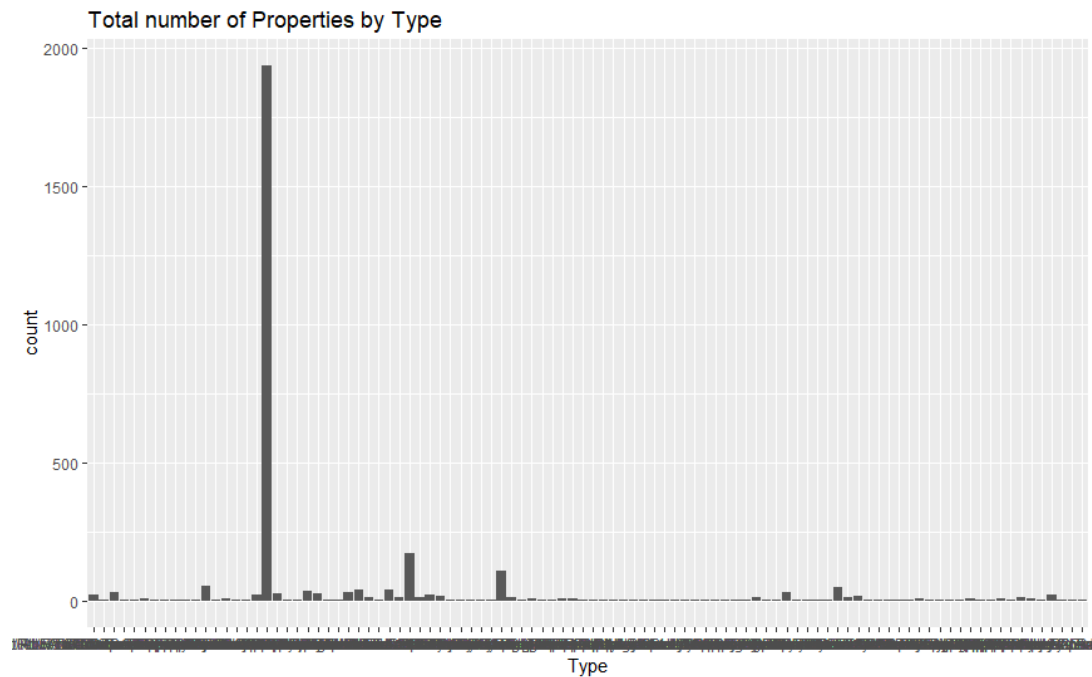
> unique(df\$Type)

[1] "Semi-Detached 2-Storey"	"Att/Row/Twnhouse 3-Storey"
"Semi-Detached Backsplit 5"	
[4] "Detached 2-Storey"	"Condo Townhouse 3-Storey"
"Condo Apt Apartment"	
[7] "Store w/Apt/Offc Apartment"	"Lower Level Bachelor/Studio"
"Multiplex Apartment"	
[10] "Detached Bungalow"	"Att/Row/Twnhouse Apartment"
"Att/Row/Twnhouse 2-Storey"	
[13] "Fourplex Apartment"	"Shared Room Apartment"
"Semi-Detached Other"	
[16] "Detached 1 1/2 Storey"	"Triplex Apartment"
"Upper Level Apartment"	
[19] "Detached Bungalow-Raised"	"Detached Apartment"
"Lower Level 2 1/2 Storey"	
[22] "Semi-Detached Bachelor/Studio"	"Detached Bungaloft"
"Multiplex Bachelor/Studio"	
[25] "Room 3-Storey"	"Lower Level 1 1/2 Storey"
"Detached Bachelor/Studio"	
[28] "Semi-Detached Apartment"	"Lower Level 2-Storey"
"Multiplex 3-Storey"	
[31] "Duplex 2-Storey"	"Semi-Detached Bungalow"
"Lower Level Bungalow-Raised"	
[34] "Upper Level Bachelor/Studio"	"Store w/Apt/Offc 2-Storey"
"Other Apartment"	
[37] "Condo Townhouse Stacked Townhse"	"Detached 2 1/2 Storey"
"Condo Apt Bungalow"	
[40] "Lower Level Apartment"	"Condo Apt Bachelor/Studio"
"Lower Level Bungalow"	
[43] "Semi-Detached Bungalow-Raised"	"Detached Sidesplit 4"
"Detached Backsplit 3"	
[46] "Comm Element Condo Apartment"	"Detached 3-Storey"
"Co-Ownership Apt Bachelor/Studio"	
[49] "Semi-Detached 3-Storey"	"Lower Level Backsplit 4"
"Detached Backsplit 4"	
[52] "Semi-Detached 2 1/2 Storey"	"Triplex 2-Storey"
"Duplex 2 1/2 Storey"	
[55] "Comm Element Condo Multi-Level"	"Condo Apt Loft"
"Condo Apt Multi-Level"	
[58] "Condo Townhouse 2-Storey"	"Other Multi-Level"
"Co-Op Apt Apartment"	
[61] "Detached Other"	"Duplex Bungalow"
"Semi-Detached 1 1/2 Storey"	
[64] "Upper Level 2-Storey"	"Upper Level Backsplit 4"
"Upper Level 3-Storey"	
[67] "Triplex 1 1/2 Storey"	"Condo Townhouse Apartment"
"Condo Apt Stacked Townhse"	
[70] "Condo Apt 2-Storey"	"Duplex Apartment"
"Att/Row/Twnhouse 2 1/2 Storey"	
[73] "Detached Sidesplit 3"	"Upper Level Other"
"Co-Ownership Apt Apartment"	
[76] "Multiplex 2-Storey"	"Triplex 3-Storey"
"Store w/Apt/Offc 3-Storey"	
[79] "Co-Ownership Apt 2-Storey"	"Detached Backsplit 5"
"Condo Apt Other"	

[82] "Condo Townhouse Multi-Level"	"Duplex 3-Storey"
"Other 2-Storey"	
[85] "Comm Element Condo Stacked Townhse"	"Fourplex 3-Storey"
"Comm Element Condo Loft"	
[88] "Fourplex 1 1/2 Storey"	"Other Other"
"Att/Row/Twnhouse Other"	
[91] "Fourplex 2-Storey"	"Store w/Apt/Offc Other"
"Comm Element Condo Other"	
[94] "Semi-Detached Backsplit 3"	"Detached Sidesplit 5"
"Condo Apt Industrial Loft"	
[97] "Comm Element Condo 2-Storey"	"Comm Element Condo 3-Storey"

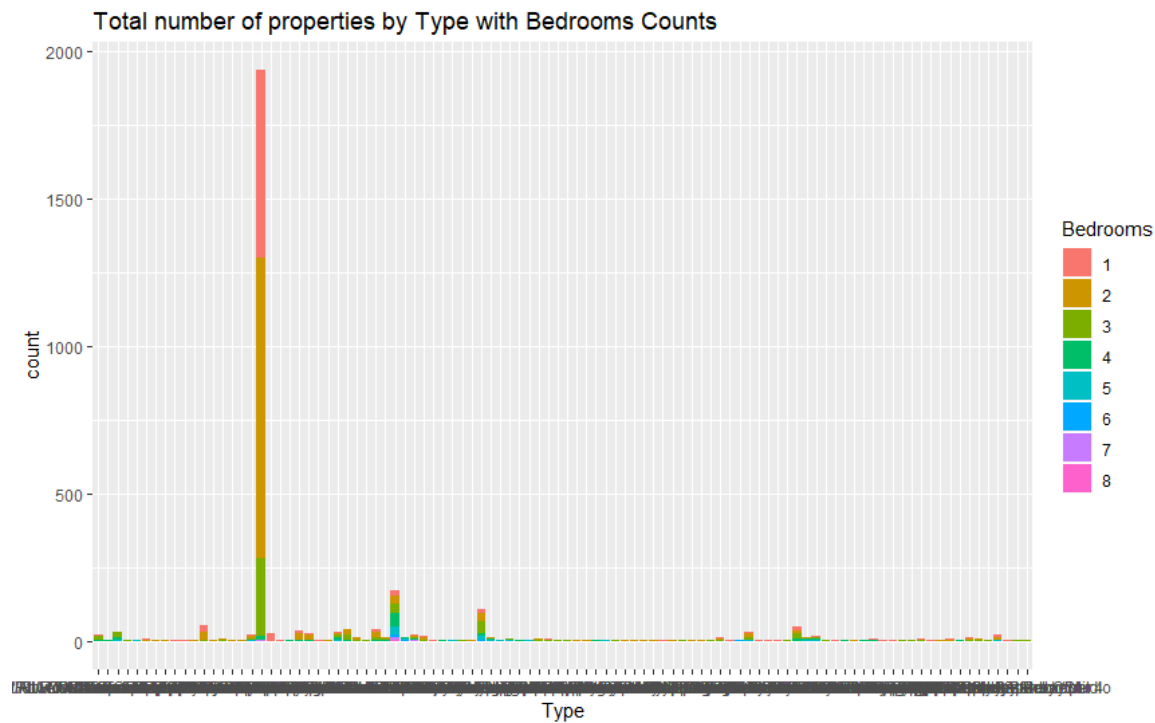
# Visualize the total number of properties by Type

```
ggplot(data=df)+geom_bar(aes(x=Type))+ggtitle("Total number of Properties by Type")
```



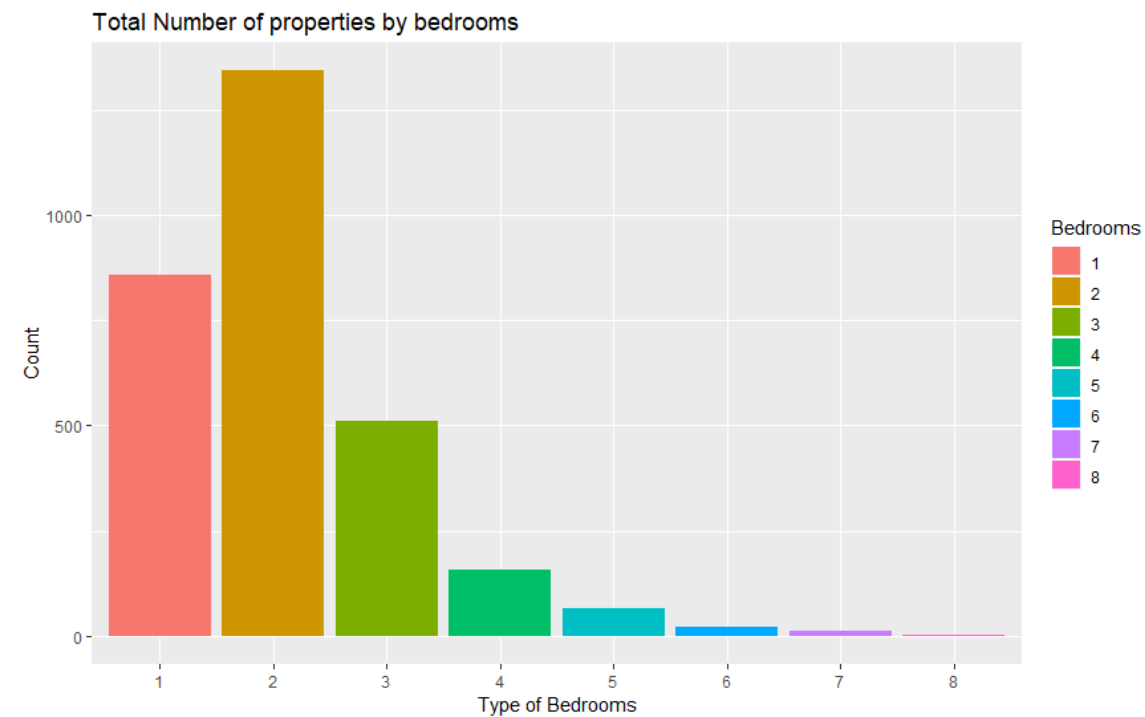
# Visualize the total number of properties by Type with fill "Bedrooms"

```
ggplot(data=df)+geom_bar(aes(x=Type,fill=Bedrooms))+ggtitle("Total number of properties by Type with Bedrooms Counts")
```

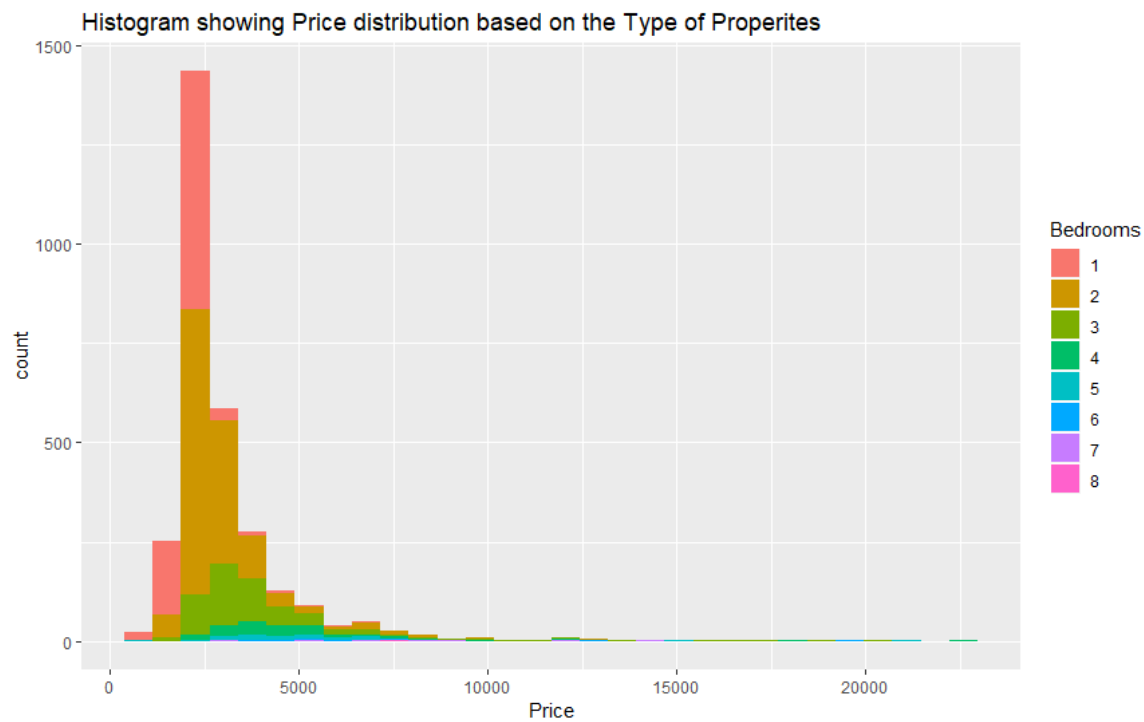


# Visualize the number of properties by Bedrooms

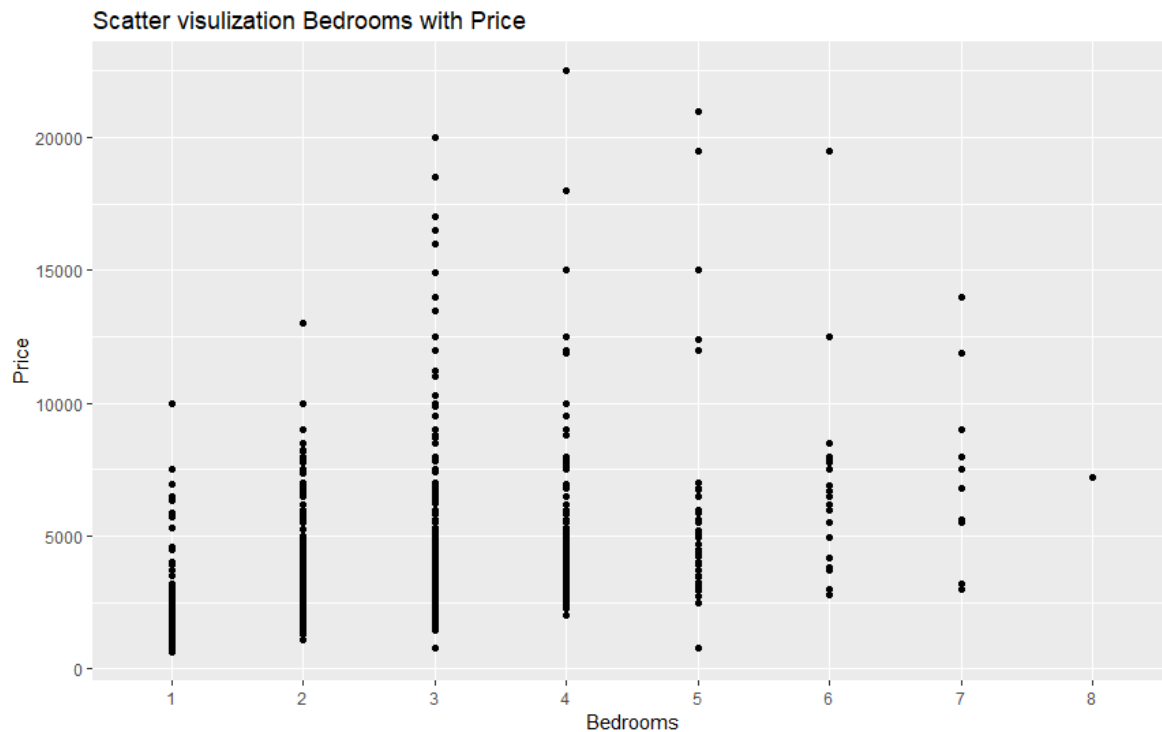
```
ggplot(data = df, aes(x= Bedrooms, fill = Bedrooms))+  
geom_bar()+ggtitle("Total Number of properties by bedrooms")+  
xlab("Type of Bedrooms")+ ylab("Count")
```



```
# Histogram visualize Price distribution based on type of properties
ggplot(data = df, aes(x= Price, bins=10, fill= Bedrooms))+
  geom_histogram()+
  ggtitle("Histogram showing Price distribution based on the Type of Properites")
```

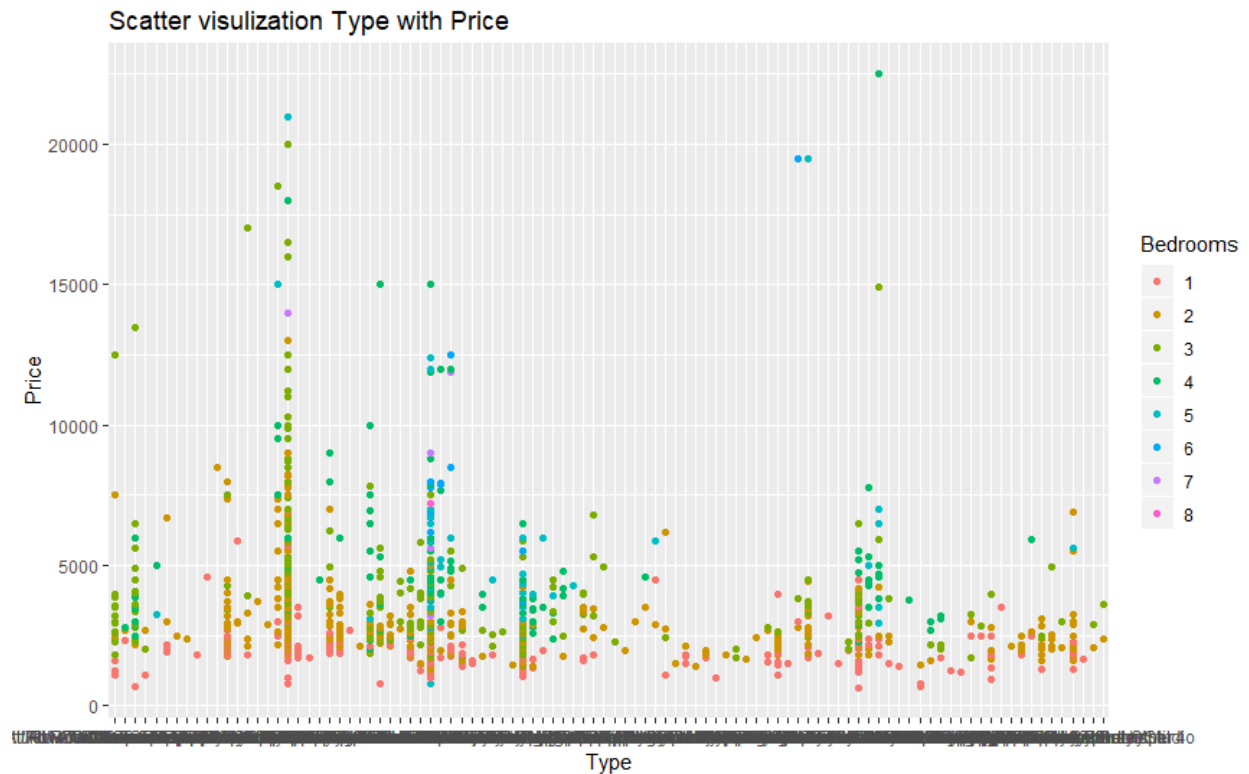


```
# Visualize Price with Bedrooms
ggplot(data = df, aes(x=Bedrooms, y=Price))+geom_point()+ggtitle("Scatter vizulization Bedrooms with Price")
```



```
# Visualize Price with Type
```

```
ggplot(data=df)+geom_point(aes(x=Type,y=Price,color=Bedrooms))+ggtitle("Scatter visulization Type  
with Price")
```



```
# Settings
```

```
mycolors <- c("#771C19", "#AA3929", "#8E9CA3", "#556670", "#000000",  
              "#E25033", "#F27314", "#F8A31B", "#E2C59F", "#B6C5CC",  
              "#99CCCC", "#FFCC99")
```

```
mytheme <- theme(axis.text.x = element_text(angle = 90, size = 10, vjust = .4),  
                 plot.title = element_text(size = 15, vjust = 2),  
                 axis.title.x = element_text(size = 12, vjust = -.35))
```

```
mytheme2 <- theme(axis.text.x = element_text(size = 10, vjust = .4),  
                  plot.title = element_text(size = 15, vjust = 2),  
                  axis.title.x = element_text(size = 12, vjust = -.35))
```

```
# Top 10 Type by the Number of Properties
```

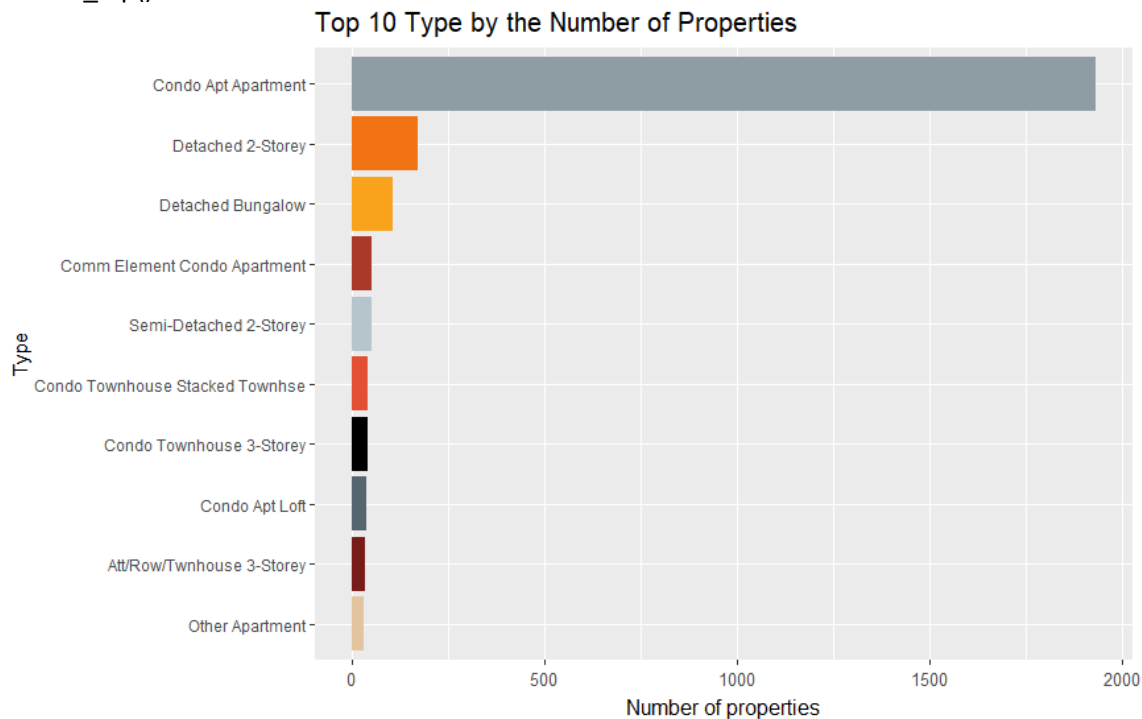
```
top10_type <- df %>% group_by(Type) %>%  
  summarise(Number = n()) %>%  
  arrange(desc(Number)) %>%  
  head(10)
```

```
ggplot(top10_type, aes(reorder(Type, Number), Number, fill = Type))+  
  geom_bar(stat = "identity")+mytheme2+  
  theme(legend.position = "none")+  
  labs(x = "Type", y = "Number of properties",
```

```

title = "Top 10 Type by the Number of Properties")+
scale_fill_manual(values = mycolors)+
coord_flip()

```

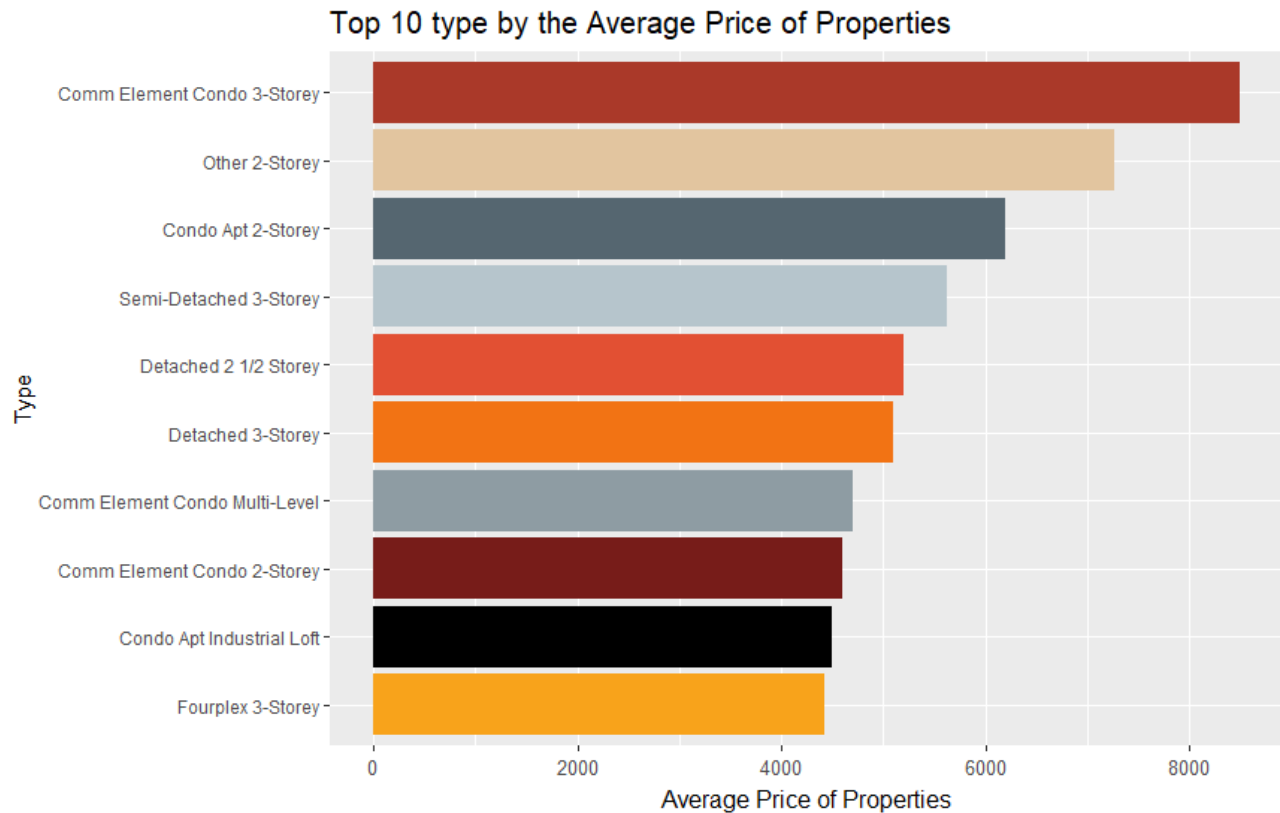


```

#Top 10 Type by the Average Price of Properties
type_vs_price <- df[c("Type","Price")] %>%na.omit()
top10type_by_averprice <- type_vs_price %>%
  group_by(Type) %>%
  summarise(Average = sum(Price)/n()) %>%
  arrange(desc(Average)) %>%
  head(10)
ggplot(top10type_by_averprice, aes(reorder(Type, Average), Average, fill = Type))+
  geom_bar(stat = "identity")+mytheme2+theme(legend.position = "none")+
  labs(x = "Type", y = "Average Price of Properties",
       title = "Top 10 type by the Average Price of Properties")+
  scale_fill_manual(values = mycolors)+
  coord_flip()

```





# Summarize the Price with Type, Bedrooms, Bathrooms and Look at Price Trend

```
df1<-df%>%
```

```
group_by(Type,Bedrooms,Bathrooms)%>%
```

```
summarize(mean_price=mean(Price,na.rm=TRUE))
```

```
write.csv(df1, file = "Summary_Type_Beds_Baths.csv",row.names=TRUE)
```

```
#see output "Summary_type_Beds_Baths.csv" file
```

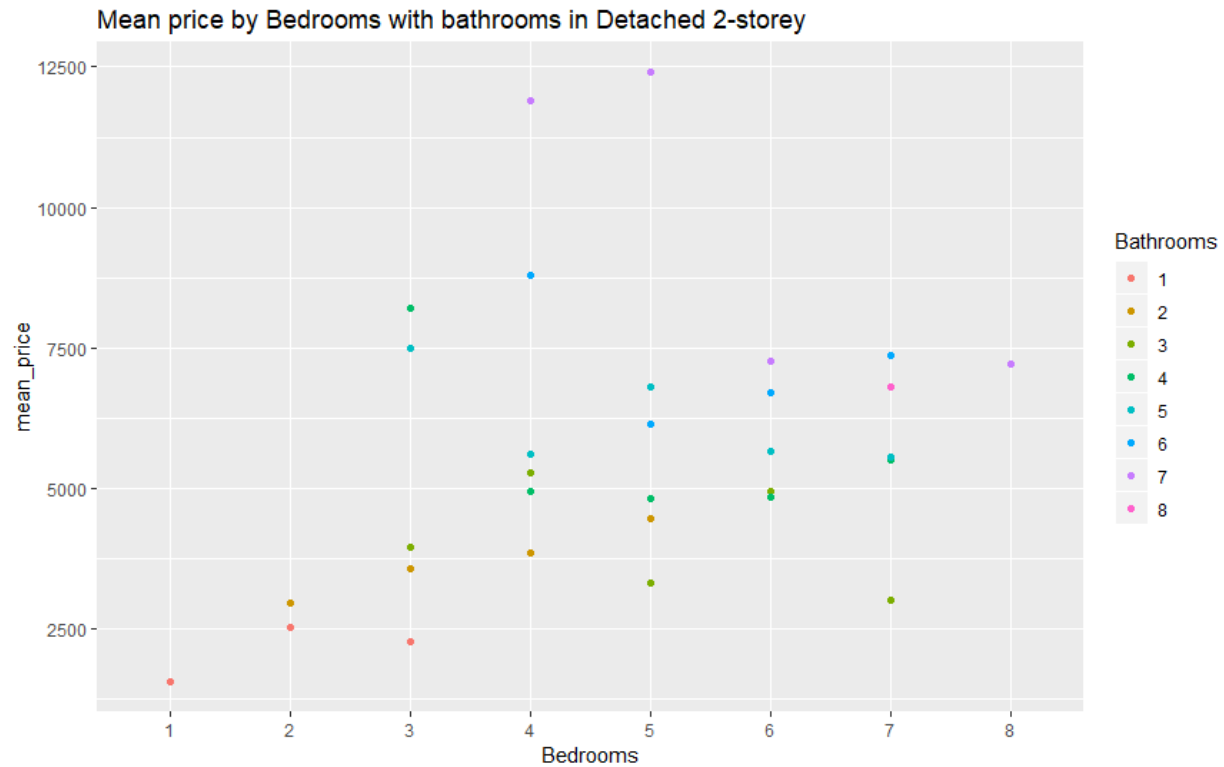
# Look the price trend in Type in "Detached 2-storey"

```
df1 %>%
```

```
filter(Type=="Detached 2-Storey")%>%
```

```
ggplot(aes(x=Bedrooms,y=mean_price))+geom_point(aes(color=Bathrooms))+geom_smooth(se=FALSE)
```

```
+ ggtitle("Mean price by Bedrooms with bathrooms in Detached 2-storey")
```



# Summarize the Price with Bedrooms, Bathrooms and Look at Price Trend

```
df2<-df1%>%
```

```
  group_by(Bedrooms,Bathrooms)%>%
```

```
  summarize(mean_price1=mean(mean_price,na.rm=TRUE))
```

```
write.csv(df, file = "Summary_Beds_Baths.csv",row.names=TRUE)
```

# see output "Summary\_Beds\_Baths.csv" file

# Summarize the Price with Bedrooms and Look at Price Trend

```
df3<-df1%>%
```

```
  group_by(Bedrooms)%>%
```

```
  summarize(mean_price2=mean(mean_price,na.rm=TRUE))
```

```
write.csv(df, file = "Summary_Beds.csv",row.names=TRUE)
```

#See output"Summary\_Beds.csv"file