airbnb_eda

Ningze Zu

11/16/2018

Abstract

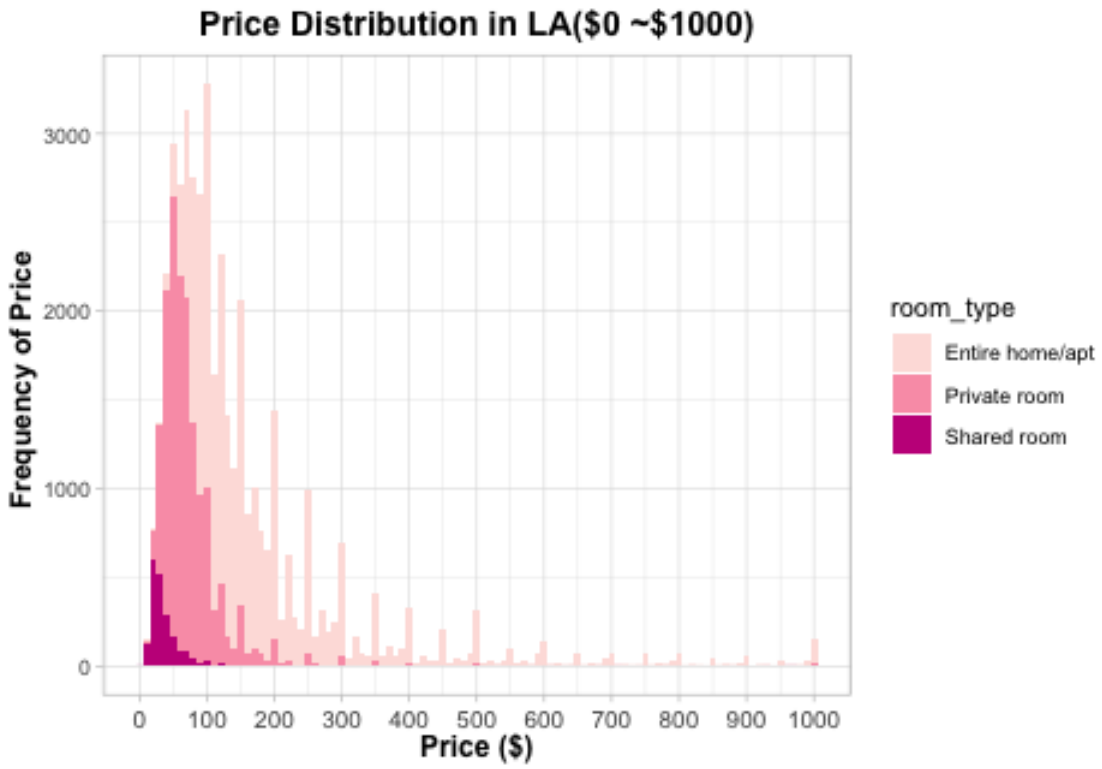
Data Cleaning

```
##Data Cleaning
##Separte four cities and Add city name with new variable "Area"
airbnb_la <- airbnb_pre2 %>% filter(str_detect(street, "CA"))
airbnb_la$Area <- rep("Los Angelas",nrow(airbnb_la))
airbnb_ny <- airbnb_pre2 %>% filter(str_detect(street, "NY"))
airbnb_ny$Area <- rep("New York",nrow(airbnb_ny))
airbnb_se <- airbnb_pre2 %>% filter(str_detect(street, "WA"))
airbnb_se$Area <- rep("Seattle",nrow(airbnb_se))
airbnb_va <- airbnb_pre2 %>% filter(str_detect(street, "Canada"))
airbnb_va$Area <- rep("Vancouver",nrow(airbnb_va))
```

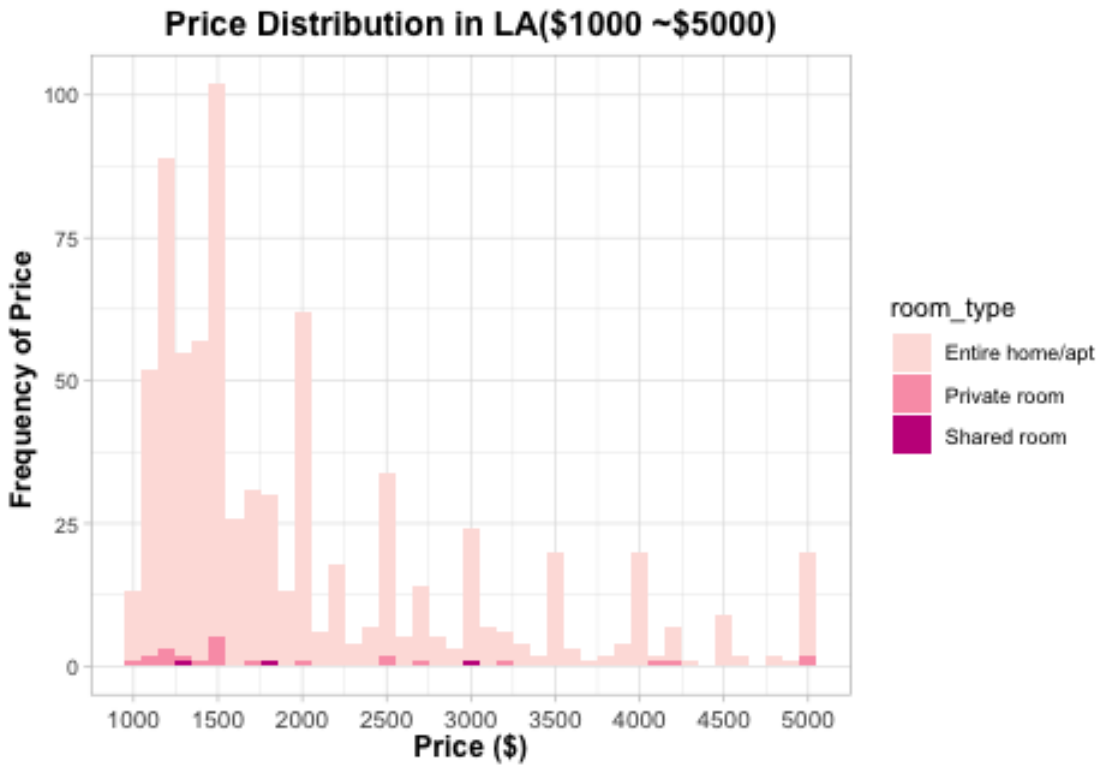
EDA

Price Distribution

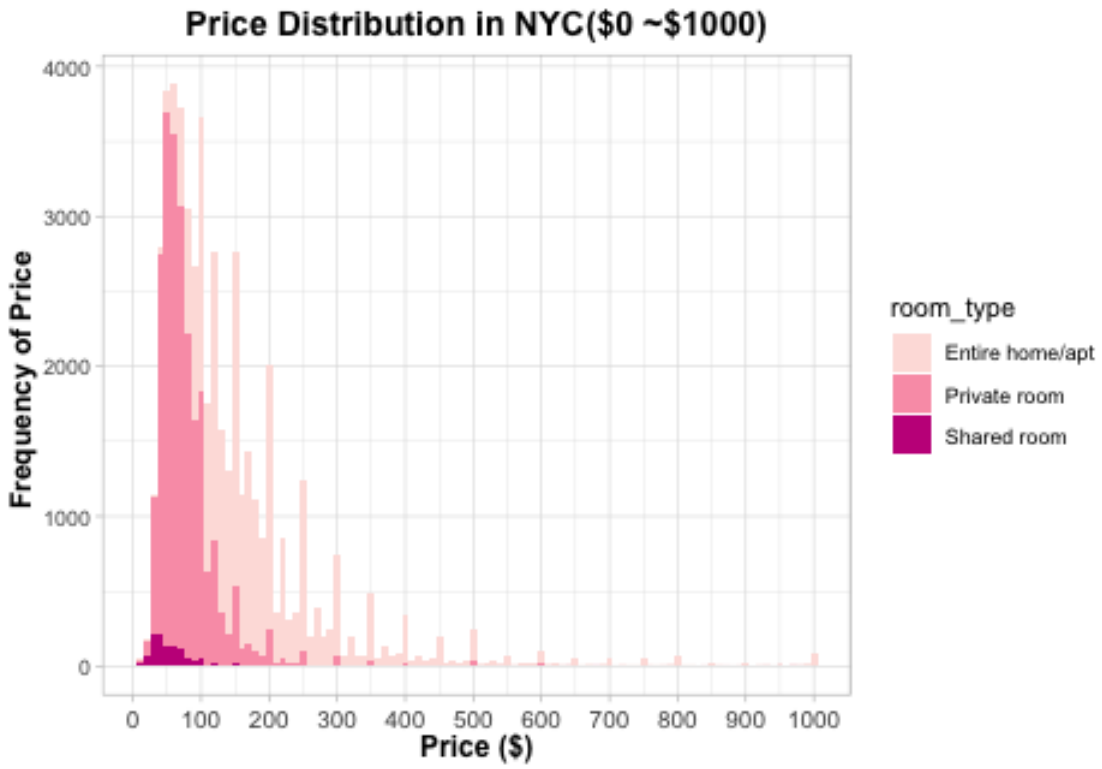
```
#Price Distribution - LA
#0-1000
airbnb_la_1 <- airbnb_la %>% filter(price <= 1000)
ggplot(airbnb_la_1, aes(price, fill = room_type)) +
  geom_histogram(binwidth = 10) + theme_light() +
  scale_fill_brewer(palette = "RdPu") +
  scale_x_continuous(name ="Price ($)", breaks = seq(0, 1000, by = 100)) +
  scale_y_continuous(name="Frequency of Price ")+
  ggtitle("Price Distribution in LA($0 ~$1000)") +
  theme(axis.title.x = element_text(face="bold",  size=12),
        axis.title.y = element_text(face="bold",  size=12),
        plot.title = element_text(size=14, face="bold"),
        axis.text.x  = element_text(vjust=0.5, size=10)) +
  theme(plot.title = element_text(hjust = 0.5))
```



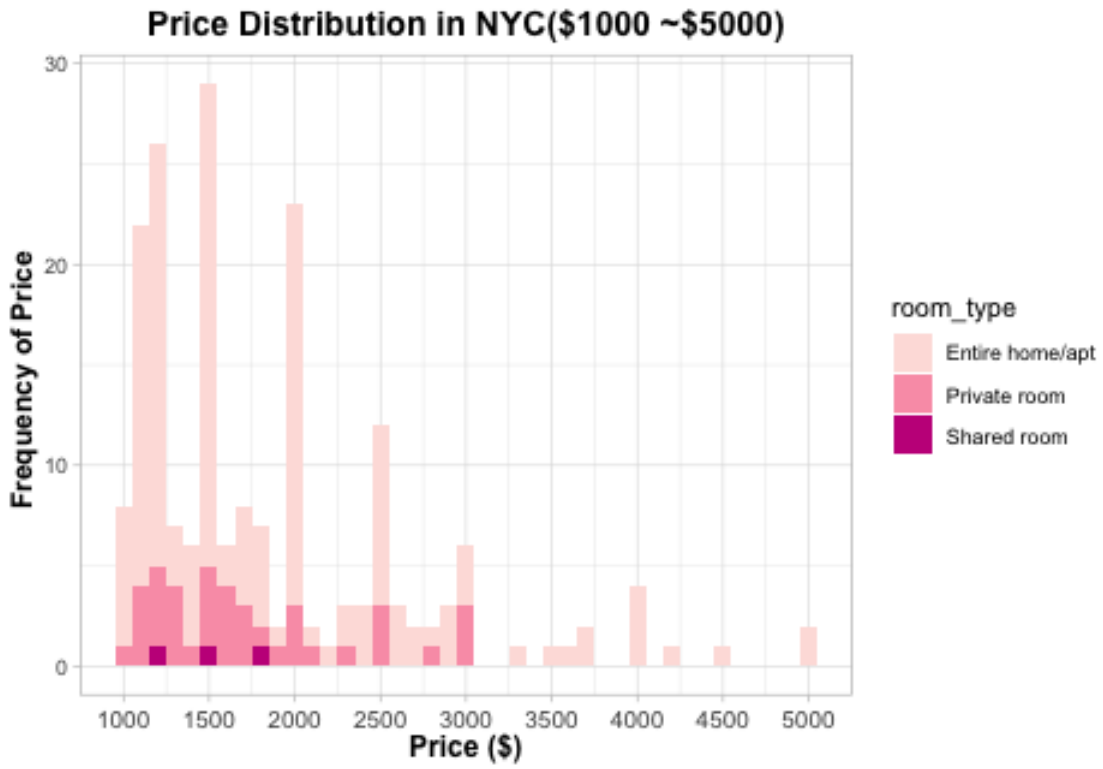

```
#1000-5000
airbnb_la_2 <- airbnb_la %>% filter(price > 1000)
ggplot(airbnb_la_2, aes(price, fill = room_type)) +
  geom_histogram(binwidth = 100) + theme_light() +
  scale_fill_brewer(palette = "RdPu") +
  scale_x_continuous(name ="Price ($)", breaks = seq(1000, 5000, by = 500)) +
  scale_y_continuous(name="Frequency of Price ")+
  ggtitle("Price Distribution in LA($1000 ~$5000)") +
  theme(axis.title.x = element_text(face="bold",  size=12),
        axis.title.y = element_text(face="bold",  size=12),
        plot.title = element_text(size=14, face="bold"),
        axis.text.x  = element_text(vjust=0.5, size=10)) +
  theme(plot.title = element_text(hjust = 0.5))
```



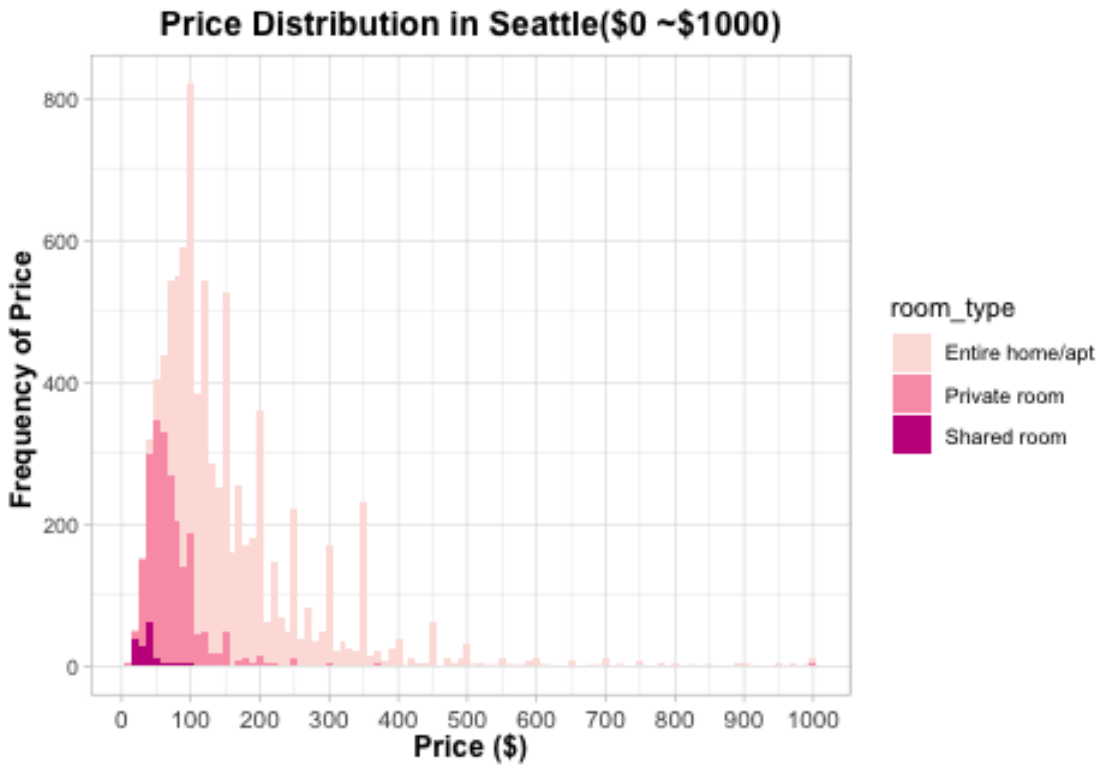

```
#Price Distribution - NY
#0-1000
airbnb_ny_1 <- airbnb_ny %>% filter(price <= 1000)
ggplot(airbnb_ny_1, aes(price, fill = room_type)) +
  geom_histogram(binwidth = 10) + theme_light() +
  scale_fill_brewer(palette = "RdPu") +
  scale_x_continuous(name ="Price ($)", breaks = seq(0, 1000, by = 100)) +
  scale_y_continuous(name="Frequency of Price ")+
  ggtitle("Price Distribution in NYC($0 ~$1000)") +
  theme(axis.title.x = element_text(face="bold",  size=12),
        axis.title.y = element_text(face="bold",  size=12),
        plot.title = element_text(size=14, face="bold"),
        axis.text.x  = element_text(vjust=0.5, size=10)) +
  theme(plot.title = element_text(hjust = 0.5))
```



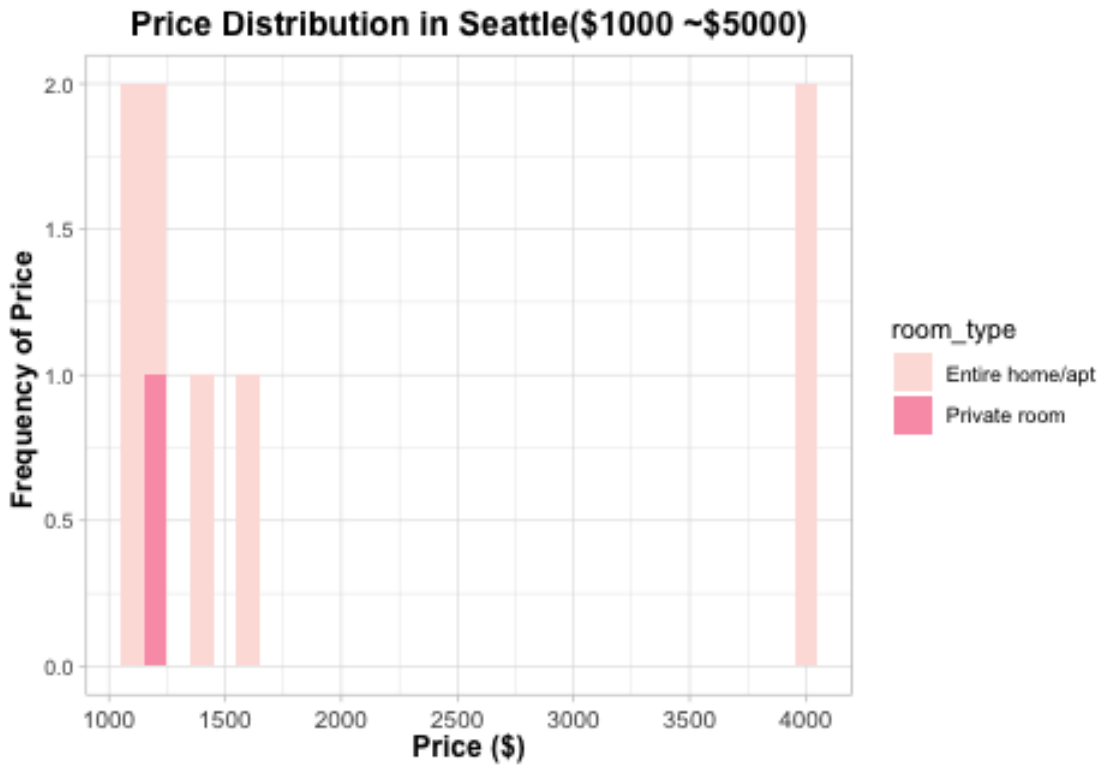

```
#1000-5000
airbnb_ny_2 <- airbnb_ny %>% filter(price > 1000)
ggplot(airbnb_ny_2, aes(price, fill = room_type)) +
  geom_histogram(binwidth = 100) + theme_light() +
  scale_fill_brewer(palette = "RdPu") +
  scale_x_continuous(name ="Price ($)", breaks = seq(1000, 5000, by = 500)) +
  scale_y_continuous(name="Frequency of Price ")+
  ggtitle("Price Distribution in NYC($1000 ~$5000)") +
  theme(axis.title.x = element_text(face="bold",  size=12),
        axis.title.y = element_text(face="bold",  size=12),
        plot.title = element_text(size=14, face="bold"),
        axis.text.x  = element_text(vjust=0.5, size=10)) +
  theme(plot.title = element_text(hjust = 0.5))
```



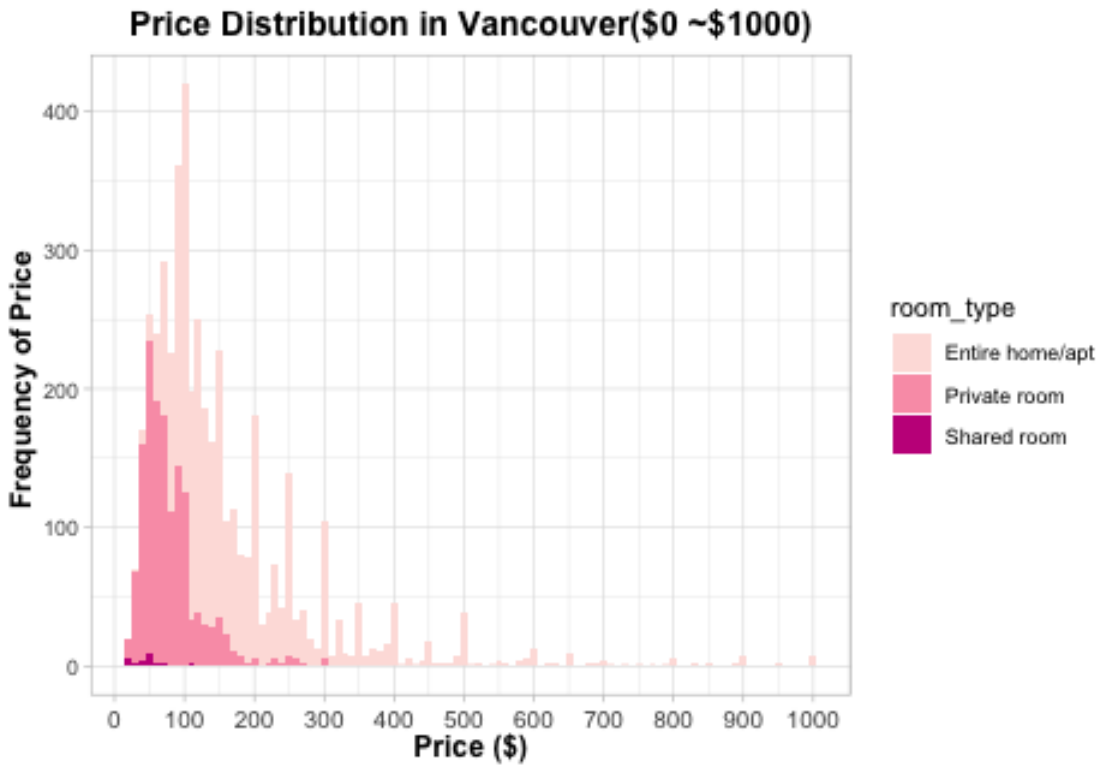

```
#Price Distribution - SEATTLE
#0-1000
airbnb_se_1 <- airbnb_se %>% filter(price <= 1000)
ggplot(airbnb_se_1, aes(price, fill = room_type)) +
  geom_histogram(binwidth = 10) + theme_light() +
  scale_fill_brewer(palette = "RdPu") +
  scale_x_continuous(name ="Price ($)", breaks = seq(0, 1000, by = 100)) +
  scale_y_continuous(name="Frequency of Price ")+
  ggtitle("Price Distribution in Seattle($0 ~$1000)") +
  theme(axis.title.x = element_text(face="bold",  size=12),
        axis.title.y = element_text(face="bold",  size=12),
        plot.title = element_text(size=14, face="bold"),
        axis.text.x  = element_text(vjust=0.5, size=10)) +
  theme(plot.title = element_text(hjust = 0.5))
```



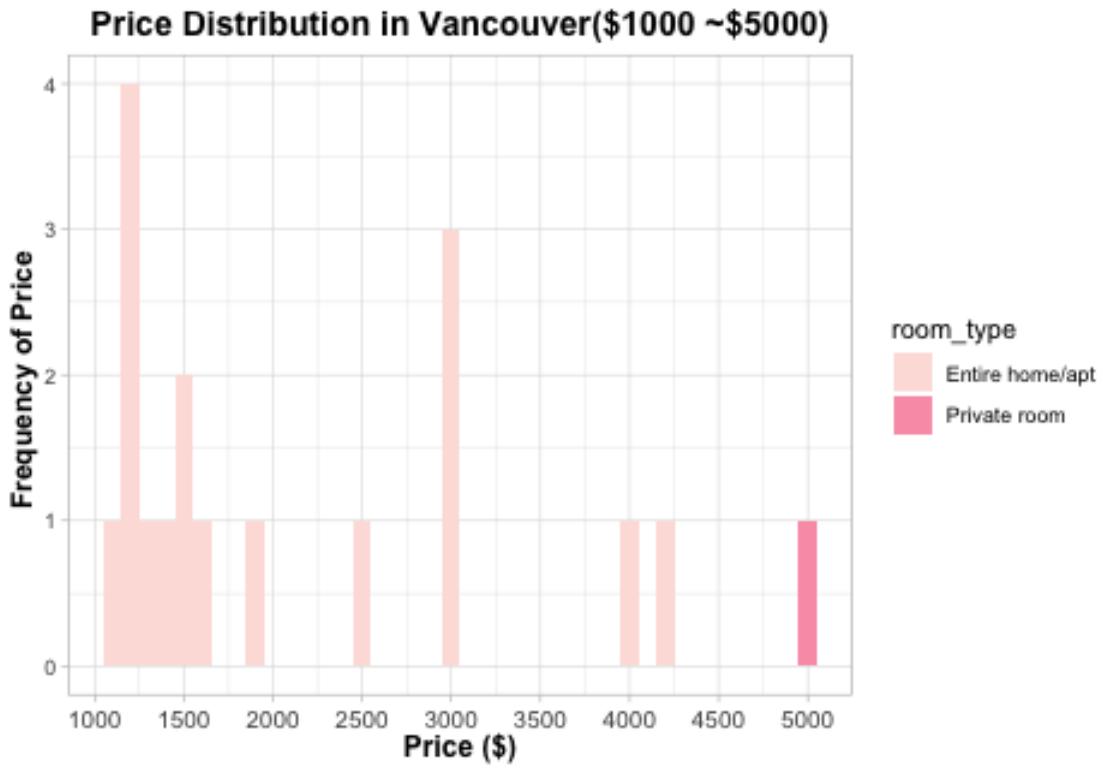

```
#1000-5000
airbnb_se_2 <- airbnb_se %>% filter(price > 1000)
ggplot(airbnb_se_2, aes(price, fill = room_type)) +
  geom_histogram(binwidth = 100) + theme_light() +
  scale_fill_brewer(palette = "RdPu") +
  scale_x_continuous(name ="Price ($)", breaks = seq(1000, 5000, by = 500)) +
  scale_y_continuous(name="Frequency of Price ")+
  ggtitle("Price Distribution in Seattle($1000 ~$5000)") +
  theme(axis.title.x = element_text(face="bold",  size=12),
        axis.title.y = element_text(face="bold",  size=12),
        plot.title = element_text(size=14, face="bold"),
        axis.text.x  = element_text(vjust=0.5, size=10)) +
  theme(plot.title = element_text(hjust = 0.5))
```




```
#Price Distribution - Vancouver
#0-1000
airbnb_va_1 <- airbnb_va %>% filter(price <= 1000)
ggplot(airbnb_va_1, aes(price, fill = room_type)) +
  geom_histogram(binwidth = 10) + theme_light() +
  scale_fill_brewer(palette = "RdPu") +
  scale_x_continuous(name ="Price ($)", breaks = seq(0, 1000, by = 100)) +
  scale_y_continuous(name="Frequency of Price ")+
  ggtitle("Price Distribution in Vancouver($0 ~$1000)") +
  theme(axis.title.x = element_text(face="bold",  size=12),
        axis.title.y = element_text(face="bold",  size=12),
        plot.title = element_text(size=14, face="bold"),
        axis.text.x  = element_text(vjust=0.5, size=10)) +
  theme(plot.title = element_text(hjust = 0.5))
```

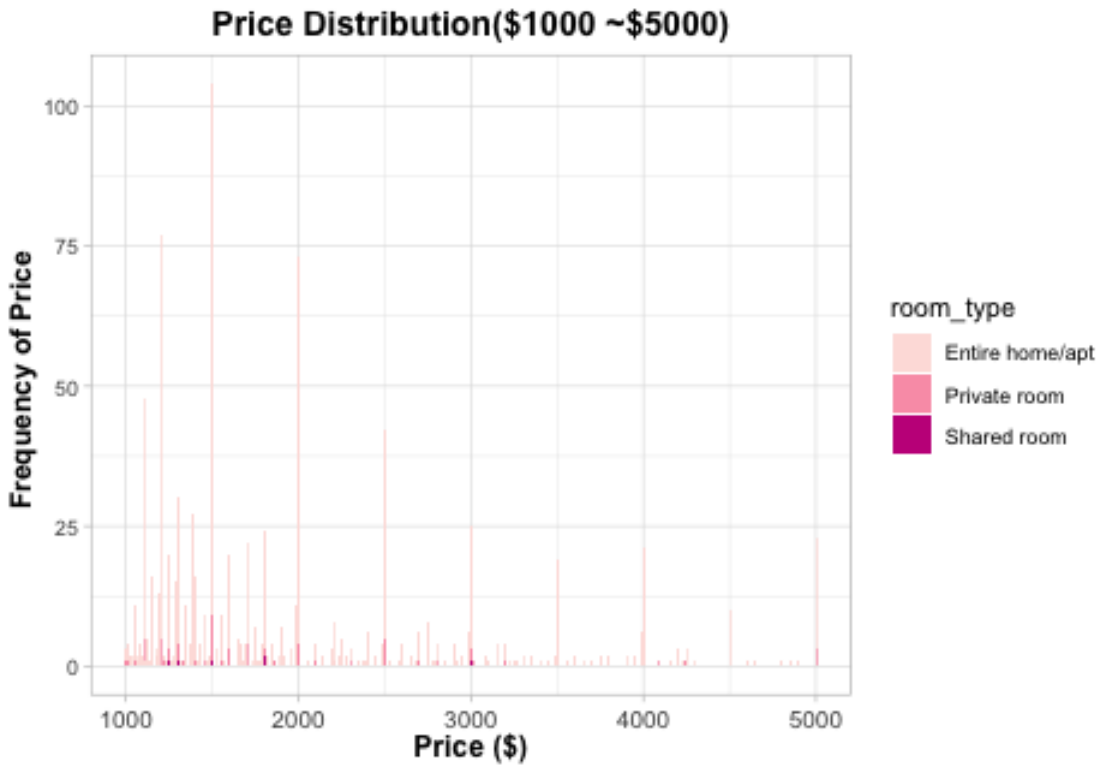



```
#1000-5000
airbnb_va_2 <- airbnb_va %>% filter(price > 1000)
ggplot(airbnb_va_2, aes(price, fill = room_type)) +
  geom_histogram(binwidth = 100) + theme_light() +
  scale_fill_brewer(palette = "RdPu") +
  scale_x_continuous(name ="Price ($)", breaks = seq(1000, 5000, by = 500)) +
  scale_y_continuous(name="Frequency of Price ")+
  ggtitle("Price Distribution in Vancouver($1000 ~$5000)") +
  theme(axis.title.x = element_text(face="bold",  size=12),
        axis.title.y = element_text(face="bold",  size=12),
        plot.title = element_text(size=14, face="bold"),
        axis.text.x  = element_text(vjust=0.5, size=10)) +
  theme(plot.title = element_text(hjust = 0.5))
```




```
##price distribution in all four cities
##0~1000
airbnb_pre2_1 <- airbnb_pre2 %>% filter(price > 0 & price <= 1000)
ggplot(airbnb_pre2_1, aes(price, fill = room_type)) +
  geom_histogram(binwidth = 10) + theme_light() +
  scale_fill_brewer(palette = "RdPu") +
  scale_x_continuous(name ="Price ($)", breaks = seq(0, 1000, by = 100)) +
  scale_y_continuous(name="Frequency of Price ")+
  ggtitle("Price Distribution($0 ~$1000)") +
  theme(axis.title.x = element_text(face="bold",  size=12),
        axis.title.y = element_text(face="bold",  size=12),
        plot.title = element_text(size=14, face="bold"),
        axis.text.x  = element_text(vjust=0.5, size=10)) +
  theme(plot.title = element_text(hjust = 0.5))
```


```
#500~1000
airbnb_pre2_2 <- airbnb_pre2 %>% filter(price > 1000 & price <= 5000)
ggplot(airbnb_pre2_2, aes(price, fill = room_type)) +
  geom_histogram(binwidth = 10) + theme_light() +
  scale_fill_brewer(palette = "RdPu") +
  scale_x_continuous(name ="Price ($)", breaks = seq(1000, 5000, by = 1000)) +
  scale_y_continuous(name="Frequency of Price ")+
  ggtitle("Price Distribution($1000 ~$5000)") +
  theme(axis.title.x = element_text(face="bold",  size=12),
        axis.title.y = element_text(face="bold",  size=12),
        plot.title = element_text(size=14, face="bold"),
        axis.text.x  = element_text(vjust=0.5, size=10)) +
  theme(plot.title = element_text(hjust = 0.5))
```

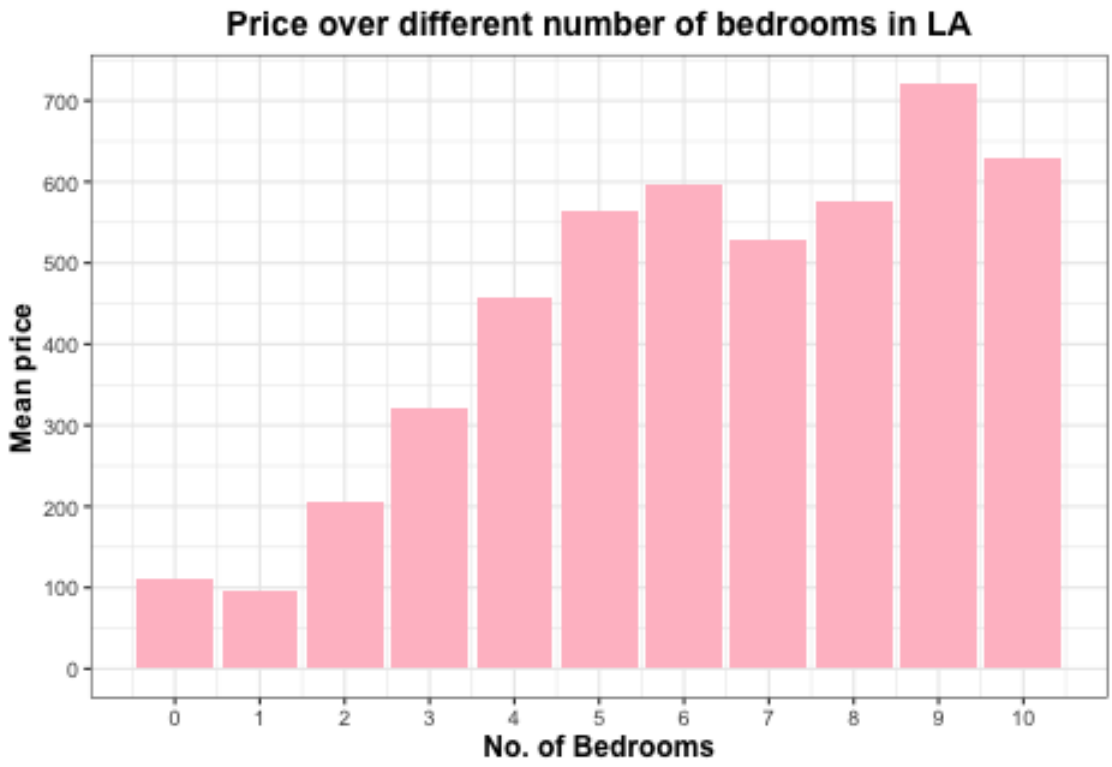


Price ($0 ~ $1000)

Price over Bedrooms

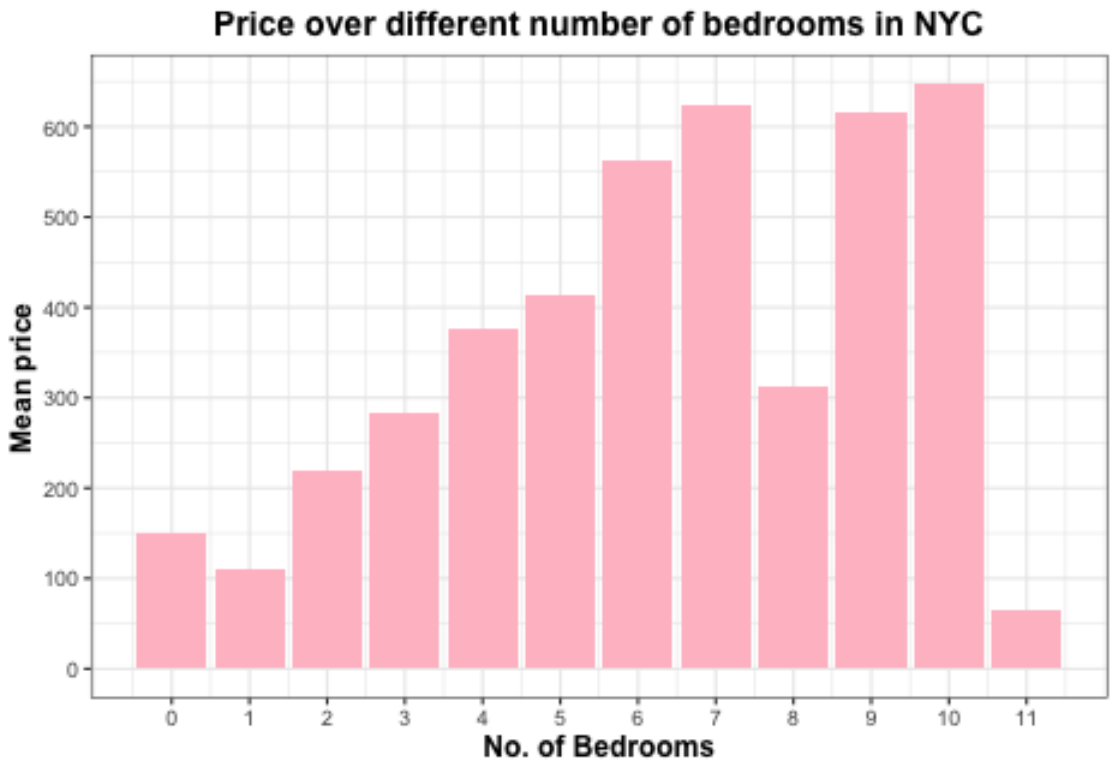
```
##bedroom ~ price in LA
airbnb_la_bed <- airbnb_la %>% filter(price > 0 & price <= 1000)
bedroom_la <- airbnb_la_bed %>% dplyr::select(price, bedrooms)
bedroom_la <- bedroom_la  %>% group_by(bedrooms) %>% summarise(mean_price = mean(price))
#barplot
ggplot(bedroom_la, aes(y=mean_price, x=bedrooms,fill=as.factor(bedrooms))) +
      geom_bar(stat="identity", fill="pink") +
      scale_x_continuous(name ="No. of Bedrooms", breaks = seq(0, 10, by = 1)) +
      scale_y_continuous(name = "Mean price", breaks = seq(0, 1000, by = 100)) + theme_bw() +
      ggtitle("Price over different number of bedrooms in LA") +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold")) +
      theme(plot.title = element_text(hjust = 0.5))
```

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




```
##bedroom ~ price in NY
airbnb_ny_bed <- airbnb_ny %>% filter(price > 0 & price <= 1000)
bedroom_ny <- airbnb_ny_bed %>% dplyr::select(price, bedrooms)
bedroom_ny <- bedroom_ny  %>% group_by(bedrooms) %>% summarise(mean_price = mean(price))
#barplot
ggplot(bedroom_ny, aes(y=mean_price, x=bedrooms,fill=as.factor(bedrooms))) +
      geom_bar(stat="identity", fill="pink") +
      scale_x_continuous(name ="No. of Bedrooms", breaks = seq(0, 11, by = 1)) +
      scale_y_continuous(name = "Mean price", breaks = seq(0, 1000, by = 100))+ theme_bw() +
      ggtitle("Price over different number of bedrooms in NYC") +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold")) +
      theme(plot.title = element_text(hjust = 0.5))
```

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

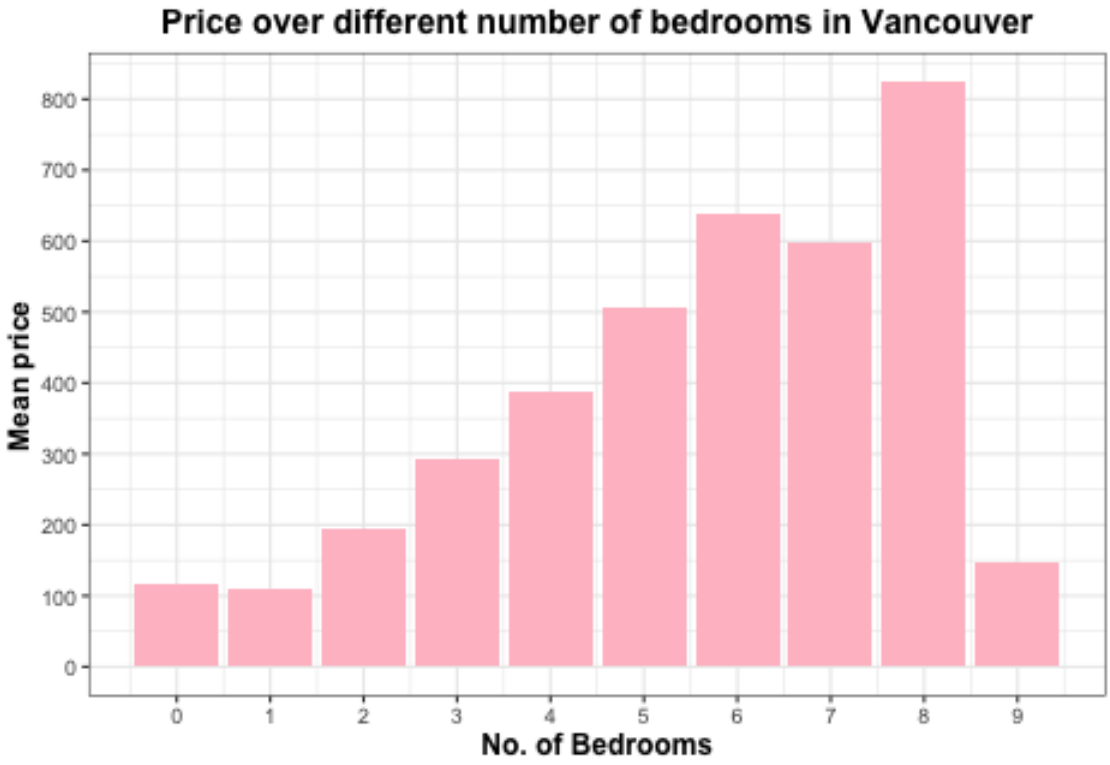



```
##bedroom ~ price in SE
airbnb_se_bed <- airbnb_se %>% filter(price > 0 & price <= 1000)
bedroom_se <- airbnb_se_bed %>% dplyr::select(price, bedrooms)
bedroom_se <- bedroom_se  %>% group_by(bedrooms) %>% summarise(mean_price = mean(price))
#barplot
ggplot(bedroom_se, aes(y=mean_price, x=bedrooms,fill=as.factor(bedrooms))) +
      geom_bar(stat="identity", fill="pink") +
      scale_x_continuous(name = "No. of Bedrooms", breaks = seq(0, 10, by = 1)) +
      scale_y_continuous(name = "Mean price", breaks = seq(0, 1000, by = 100))+ theme_bw()+
      ggtitle("Price over different number of bedrooms in Seattle") +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold")) +
      theme(plot.title = element_text(hjust = 0.5))
```

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


```
##bedroom ~ price in VA
airbnb_va_bed <- airbnb_va %>% filter(price > 0 & price <= 1000)
bedroom_va <- airbnb_va_bed %>% dplyr::select(price, bedrooms)
bedroom_va <- bedroom_va  %>% group_by(bedrooms) %>% summarise(mean_price = mean(price))
#barplot
ggplot(bedroom_va, aes(y=mean_price, x=bedrooms,fill=as.factor(bedrooms))) +
      geom_bar(stat="identity", fill="pink") +
      scale_x_continuous(name ="No. of Bedrooms", breaks = seq(0, 9, by = 1)) +
      scale_y_continuous(name = "Mean price", breaks = seq(0, 1000, by = 100))+ theme_bw() +
      ggtitle("Price over different number of bedrooms in Vancouver") +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold")) +
      theme(plot.title = element_text(hjust = 0.5))
```

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

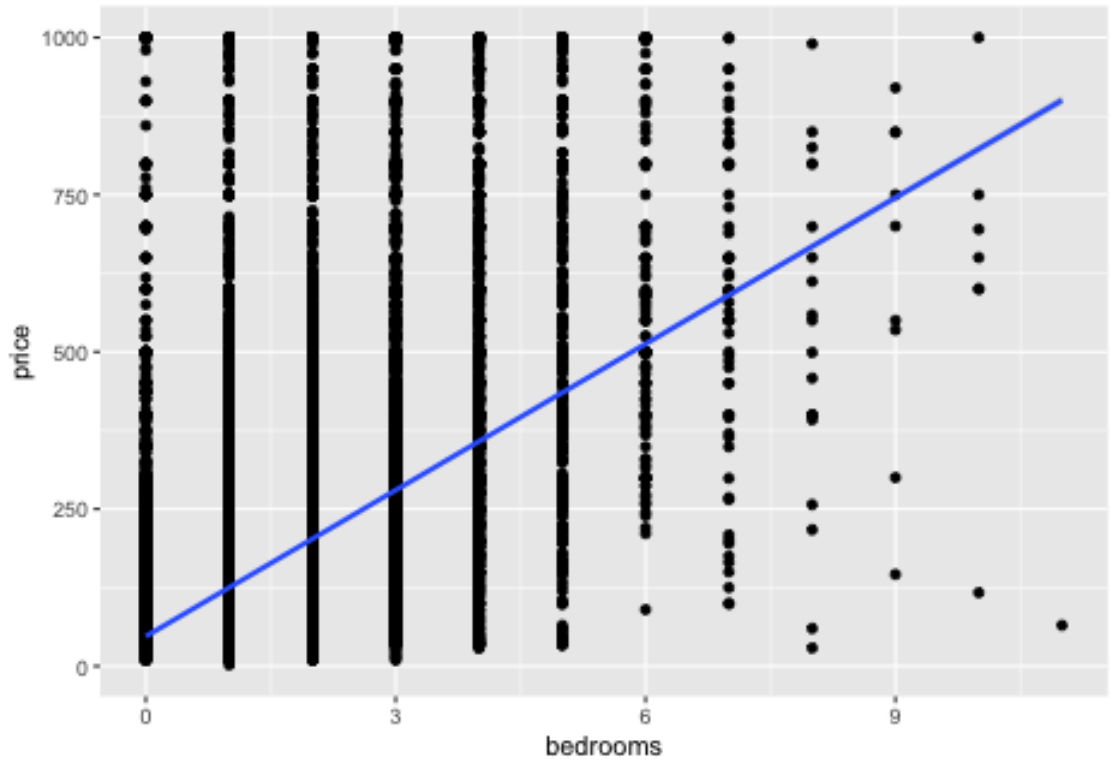


Geom smooth line ~ bedrooms/bathrooms

```
airbnb_pre2_1 <- airbnb_pre2 %>% filter(price > 0 & price <= 1000)
ggplot(airbnb_pre2_1, aes(x=bedrooms, y=price)) + geom_point() + geom_smooth(method = "lm")
```

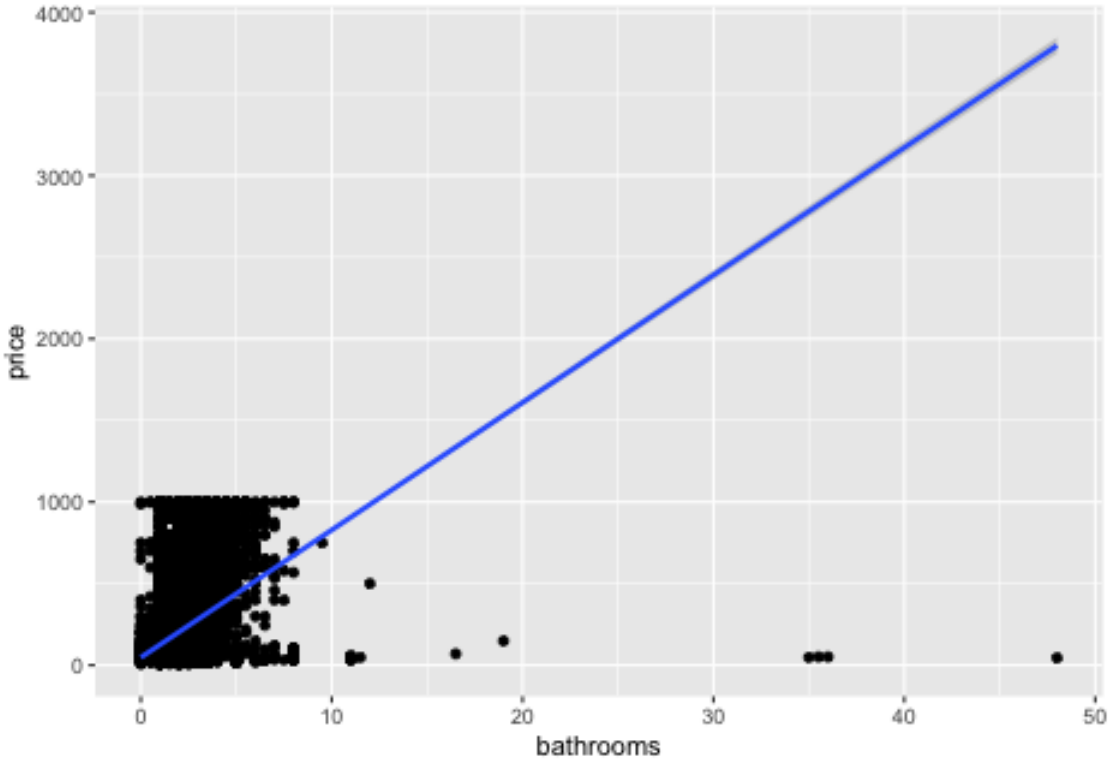
```
## Warning: Removed 82 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 82 rows containing missing values (geom_point).
```



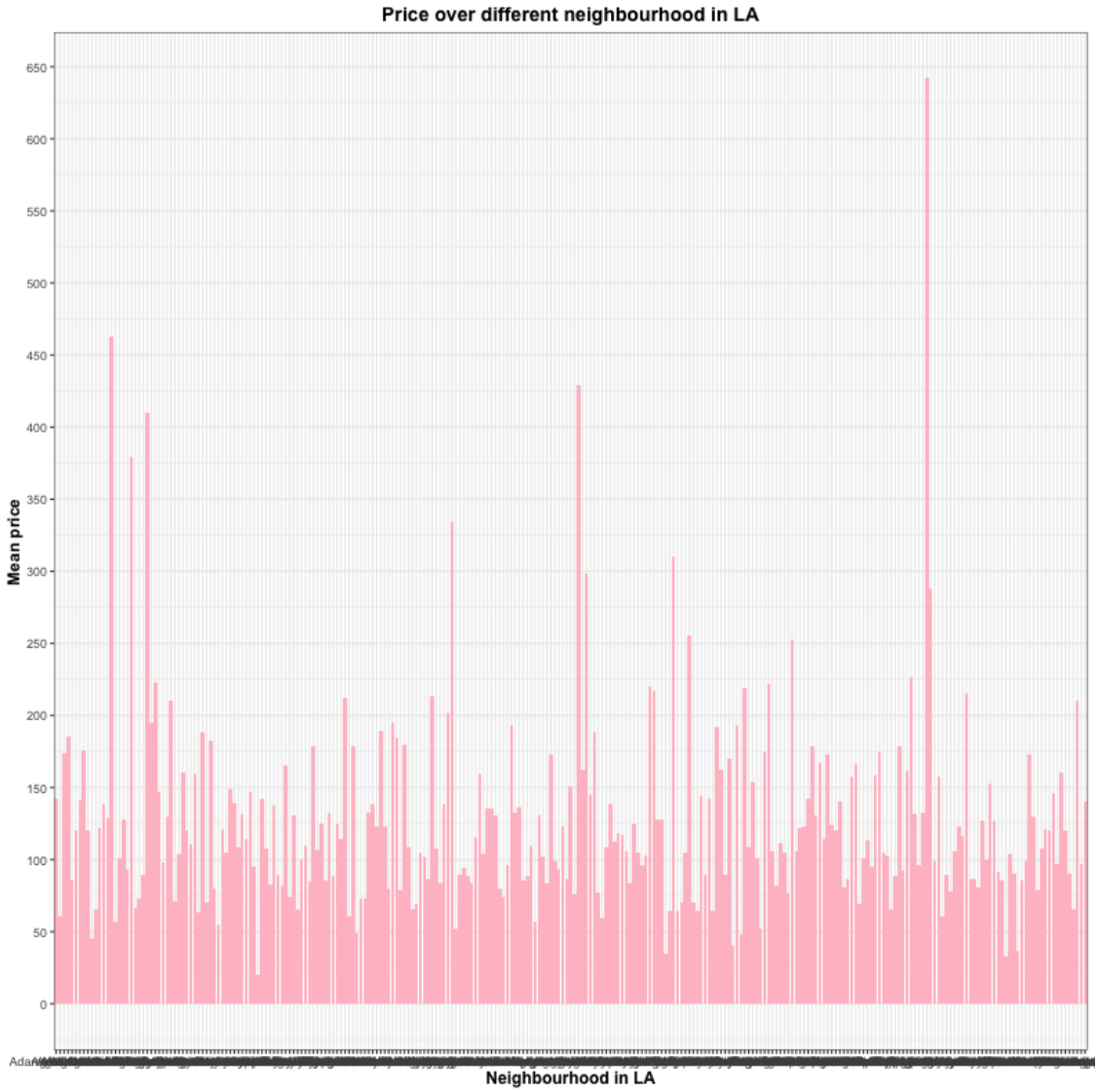

```
ggplot(airbnb_pre2_1, aes(x=bathrooms, y=price)) + geom_point() + geom_smooth(method = "lm")
```

```
## Warning: Removed 131 rows containing non-finite values (stat_smooth).
## Warning: Removed 131 rows containing missing values (geom_point).
```



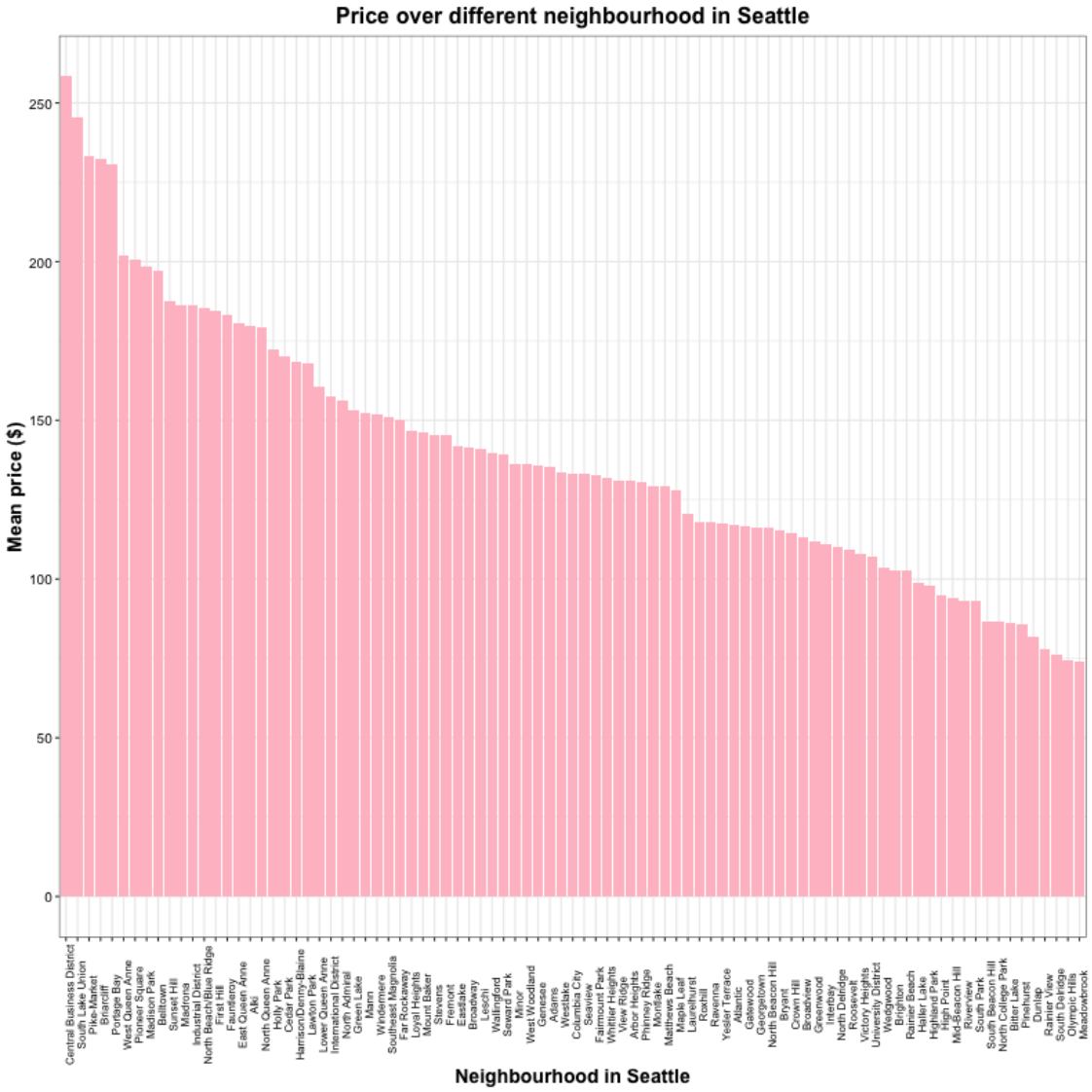
Price over Different Neighbourhoods

```
###LA neighbourhood ~ Price
###since LA has 587 neighbourhood, I decide to use city as alternative choice.
airbnb_la_nei <- airbnb_la %>% filter(price > 0 & price <= 1000)
neigh_la <- airbnb_la_nei %>% dplyr::select(price, neighbourhood_cleansed)
neigh_la <- neigh_la  %>% group_by(neighbourhood_cleansed) %>% summarise(mean_price = mean(price))
ggplot(neigh_la, aes(y=mean_price, x=neighbourhood_cleansed,fill=as.factor(neighbourhood_cleansed))) +
      geom_bar(stat="identity", fill="pink") +
      scale_x_discrete(name ="Neighbourhood in LA") +
      scale_y_continuous(name = "Mean price", breaks = seq(0, 1000, by = 50))+ theme_bw() +
      ggtitle("Price over different neighbourhood in LA") +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold")) +
      theme(plot.title = element_text(hjust = 0.5))
```



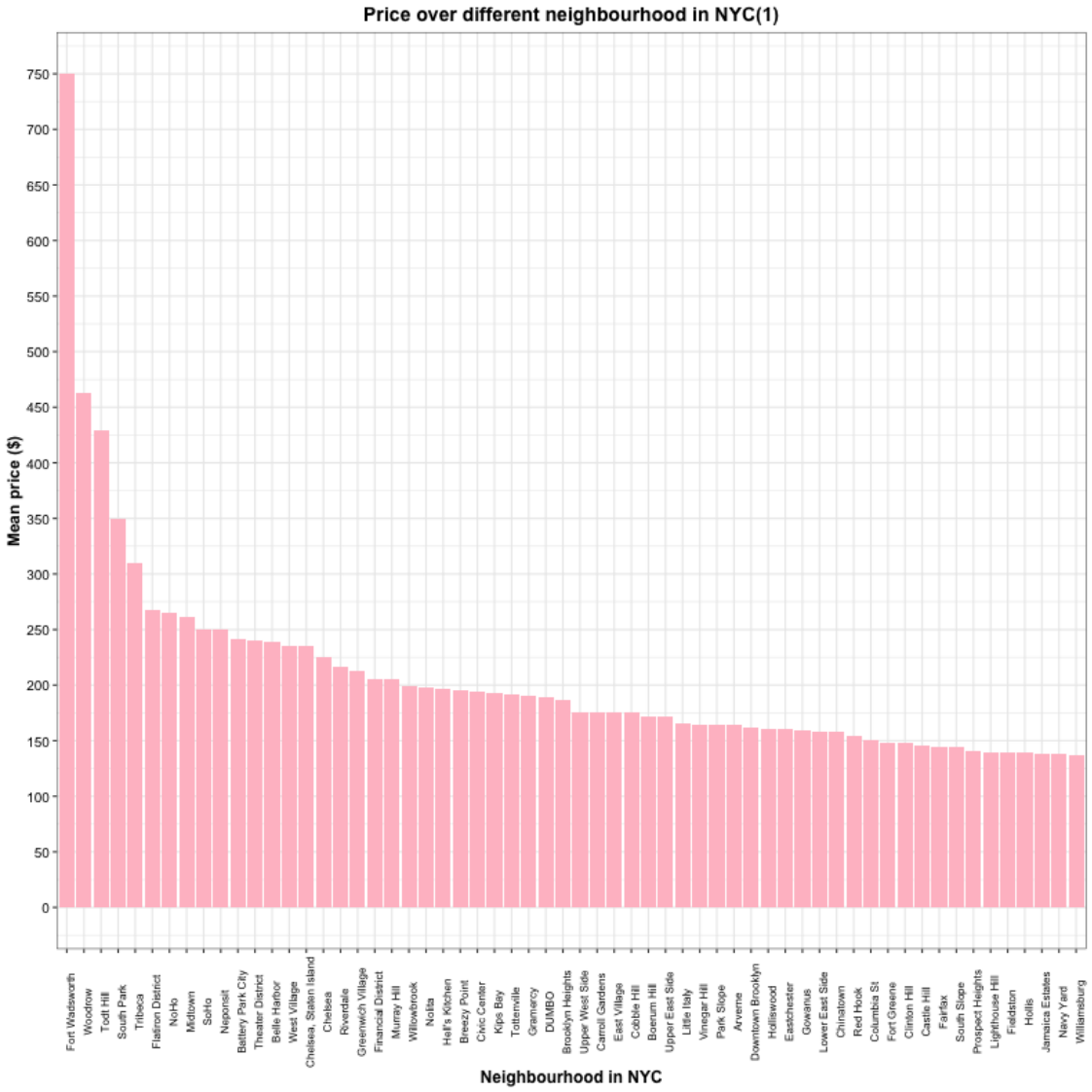

```
#unique(airbnb_la$city)



###Seattle Neightbourhood ~ Price
airbnb_se_nei <- airbnb_se %>% filter(price > 0 & price <= 1000)
neigh_se <- airbnb_se_nei %>% dplyr::select(price, neighbourhood_cleansed)
neigh_se <- neigh_se  %>% group_by(neighbourhood_cleansed) %>% summarise(mean_price = mean(price))
ggplot(neigh_se, aes(y=mean_price, x =
                     reorder(x=neighbourhood_cleansed,-mean_price),
                     fill=as.factor(neighbourhood_cleansed))) +
      geom_bar(stat="identity", fill="pink", position = "dodge") +
      scale_x_discrete(name ="Neighbourhood in Seattle") +
      scale_y_continuous(name = "Mean price ($)", breaks = seq(0, 1000, by = 50))+ theme_bw() +
      ggtitle("Price over different neighbourhood in Seattle") +
      theme(axis.title.x = element_text(face="bold",  size=14),
            axis.title.y = element_text(face="bold",  size=14),
            plot.title = element_text(size=16, face="bold"),
```

```
        axis.text.x=element_text(angle=90, colour="black", size = 8),
        axis.text.y=element_text(colour="black", size = 10)) +
  theme(plot.title = element_text(hjust = 0.5))
```

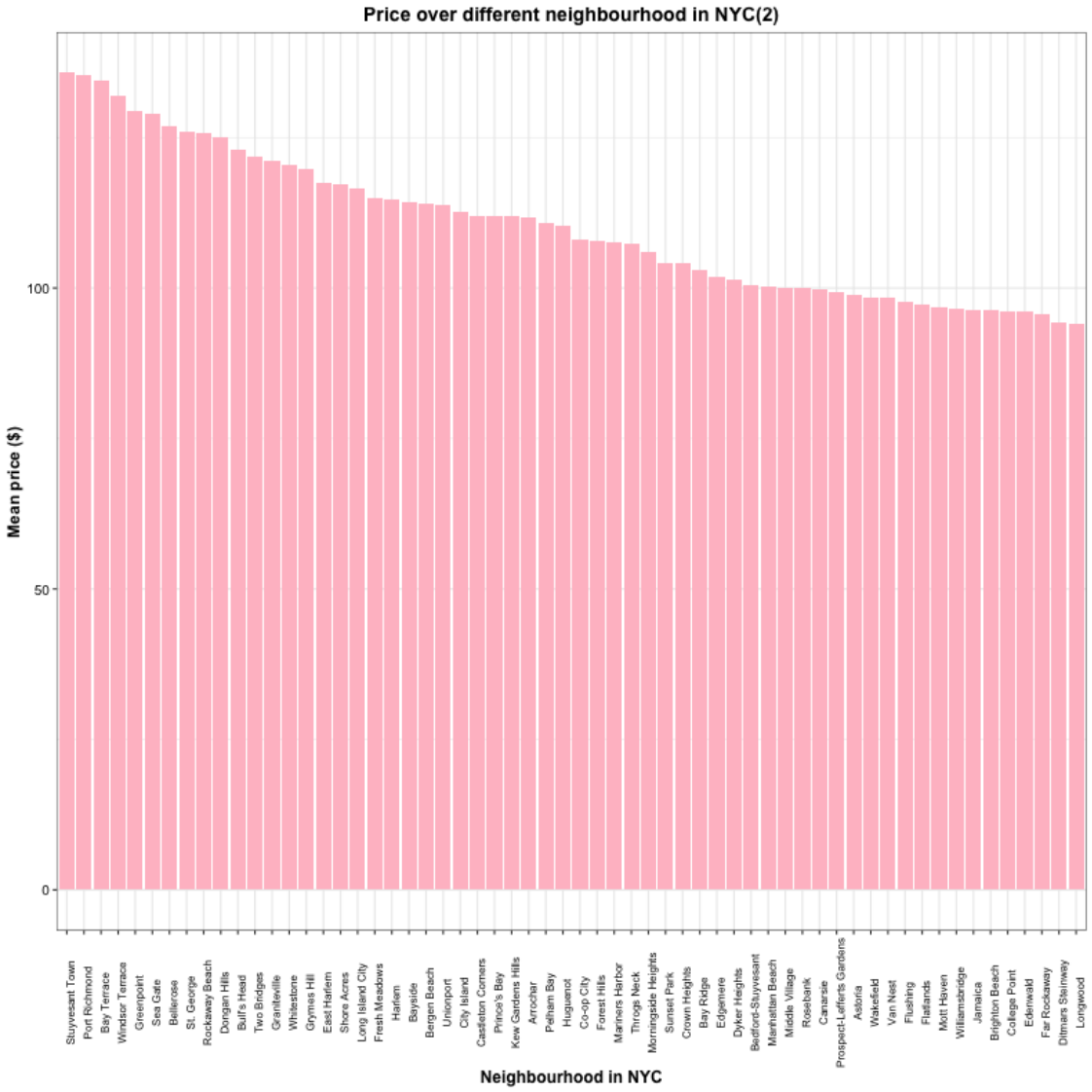



```
###NYC Neightbourhood ~ Price
airbnb_ny_nei <- airbnb_ny %>% filter(price > 0 & price <= 1000)
neigh_ny <- airbnb_ny_nei %>% dplyr::select(price, neighbourhood_cleansed)
neigh_ny <- neigh_ny  %>% group_by(neighbourhood_cleansed) %>% summarise(mean_price = mean(price)) %>% a

g1 <- ggplot(subset(neigh_ny, mean_price %in% mean_price[1:60]), aes(y=mean_price,
                      reorder(x=neighbourhood_cleansed,-mean_price),
                      fill=as.factor(neighbourhood_cleansed))) +
      geom_bar(stat="identity", fill="pink", position = "dodge") +
      scale_x_discrete(name ="Neighbourhood in NYC") +
      scale_y_continuous(name = "Mean price ($)", breaks = seq(0, 1000, by = 50))+ theme_bw() +
      ggtitle("Price over different neighbourhood in NYC(1)") +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
```

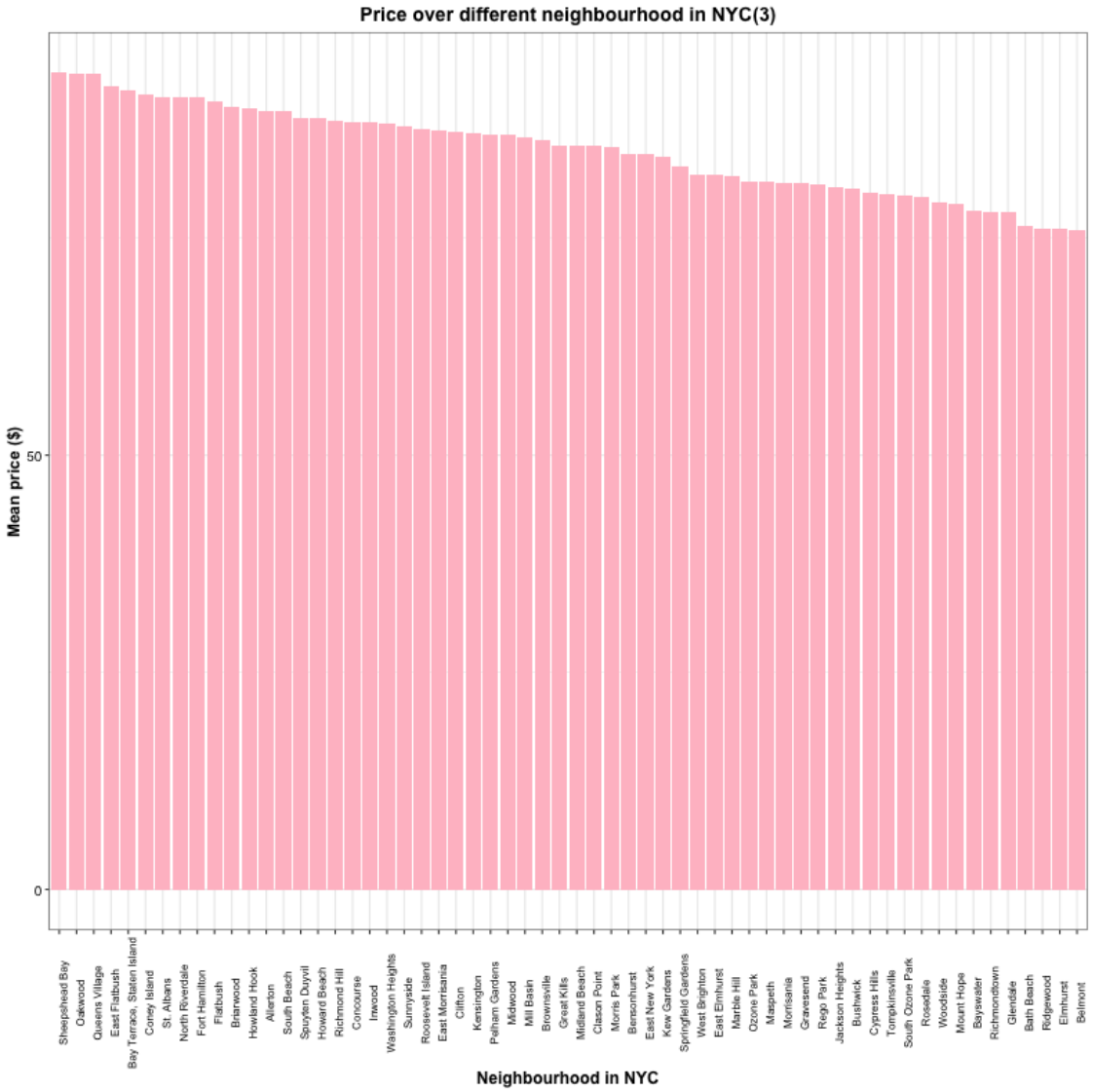
```
            plot.title = element_text(size=14, face="bold"),
            axis.text.x=element_text(angle=90, colour="black", size = 8),
            axis.text.y=element_text(colour="black", size = 10)) +
      theme(plot.title = element_text(hjust = 0.5))


g2 <- ggplot(subset(neigh_ny, mean_price %in% mean_price[61:120]), aes(y=mean_price,
                        reorder(x=neighbourhood_cleansed,-mean_price),
                        fill=as.factor(neighbourhood_cleansed))) +
      geom_bar(stat="identity", fill="pink", position = "dodge") +
      scale_x_discrete(name ="Neighbourhood in NYC") +
      scale_y_continuous(name = "Mean price ($)", breaks = seq(0, 1000, by = 50))+ theme_bw() +
      ggtitle("Price over different neighbourhood in NYC(2)") +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold"),
            axis.text.x=element_text(angle=90, colour="black", size = 8),
            axis.text.y=element_text(colour="black", size = 10)) +
      theme(plot.title = element_text(hjust = 0.5))

g3 <- ggplot(subset(neigh_ny, mean_price %in% mean_price[121:180]), aes(y=mean_price,
                        reorder(x=neighbourhood_cleansed,-mean_price),
                        fill=as.factor(neighbourhood_cleansed))) +
      geom_bar(stat="identity", fill="pink", position = "dodge") +
      scale_x_discrete(name ="Neighbourhood in NYC") +
      scale_y_continuous(name = "Mean price ($)", breaks = seq(0, 1000, by = 50))+ theme_bw() +
      ggtitle("Price over different neighbourhood in NYC(3)") +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold"),
            axis.text.x=element_text(angle=90, colour="black", size = 8),
            axis.text.y=element_text(colour="black", size = 10)) +
      theme(plot.title = element_text(hjust = 0.5))

g4 <- ggplot(subset(neigh_ny, mean_price %in% mean_price[181:226]), aes(y=mean_price,
                        reorder(x=neighbourhood_cleansed,-mean_price),
                        fill=as.factor(neighbourhood_cleansed))) +
      geom_bar(stat="identity", fill="pink", position = "dodge") +
      scale_x_discrete(name ="Neighbourhood in NYC") +
      scale_y_continuous(name = "Mean price ($)", breaks = seq(0, 1000, by = 50))+ theme_bw() +
      ggtitle("Price over different neighbourhood in NYC(4)") +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold"),
            axis.text.x=element_text(angle=90, colour="black", size = 8),
            axis.text.y=element_text(colour="black", size = 10)) +
      theme(plot.title = element_text(hjust = 0.5))
g1
```


Price over different neighbourhood in NYC(1)
Mean price ($)
750
700
650
600
550
500
450
400
350
300
250
200
150
100
50
0
Fort Wadsworth
Woodrow
Todt Hill
South Park
Tribeca
Flatiron District
NoHo
Midtown
SoHo
Neponsit
Battery Park City
Theater District
Belle Harbor
West Village
Chelsea, Staten Island
Chelsea
Riverdale
Greenwich Village
Financial District
Murray Hill
Willowbrook
Nolita
Hell's Kitchen
Breezy Point
Civic Center
Kips Bay
Tottenville
Gramercy
DUMBO
Brooklyn Heights
Upper West Side
Carroll Gardens
East Village
Cobble Hill
Boerum Hill
Upper East Side
Little Italy
Vinegar Hill
Park Slope
Arverne
Downtown Brooklyn
Hollswood
Eastchester
Gowanus
Lower East Side
Chinatown
Red Hook
Columbia St
Fort Greene
Clinton Hill
Castle Hill
Fairfax
South Slope
Prospect Heights
Lighthouse Hill
Fieldston
Hollis
Jamaica Estates
Navy Yard
Williamsburg
Neighbourhood in NYC

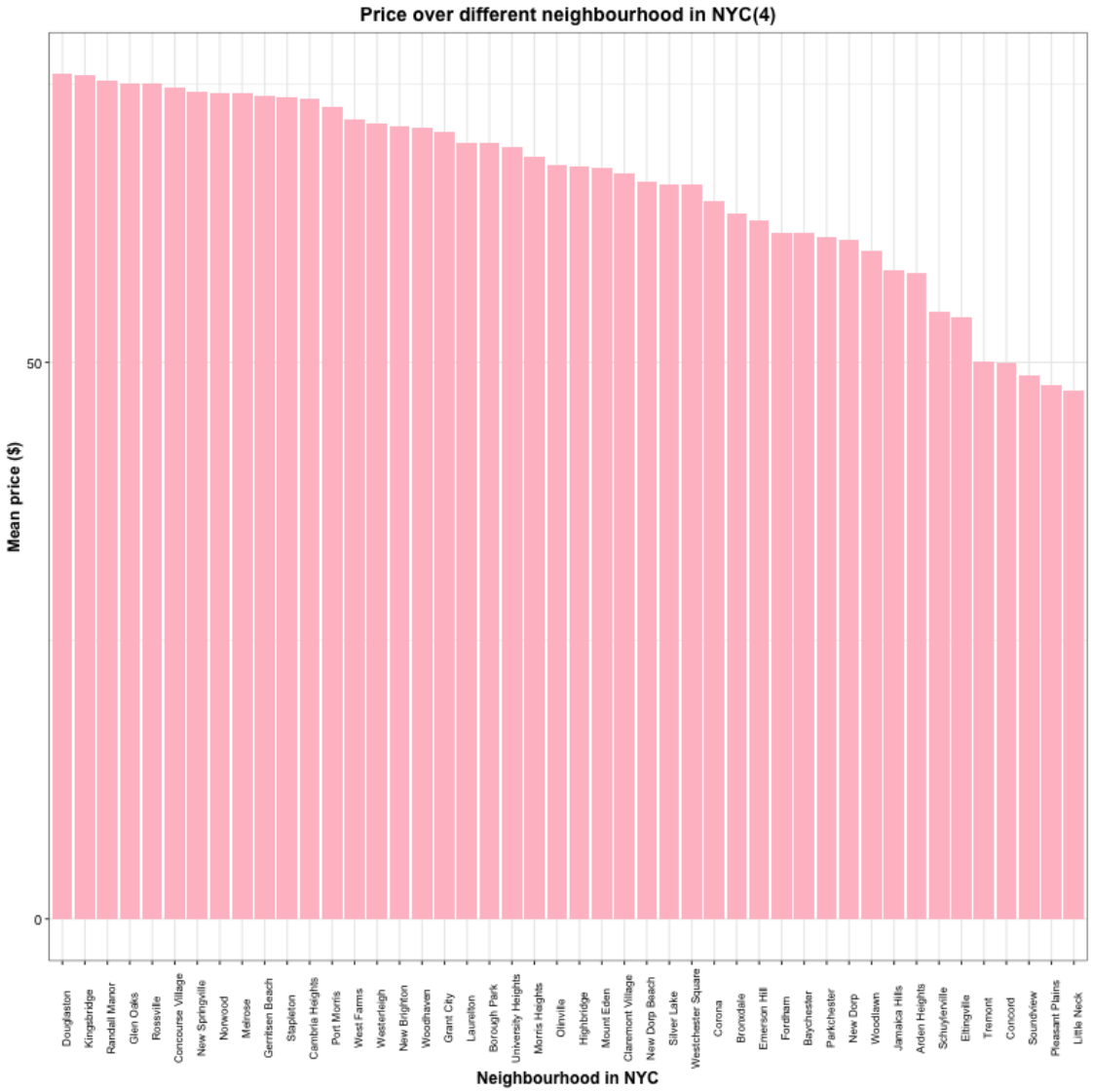
g2


Price over different neighbourhood in NYC(2)
Mean price ($)
100
50
0
Stuyvesant Town
Port Richmond
Bay Terrace
Windsor Terrace
Greenpoint
Sea Gate
Bellerose
St. George
Rockaway Beach
Dongan Hills
Bull's Head
Two Bridges
Graniteville
Whitestone
Grymes Hill
East Harlem
Shore Acres
Long Island City
Fresh Meadows
Harlem
Bayside
Bergen Beach
Unionport
City Island
Castleton Corners
Prince's Bay
Kew Gardens Hills
Arrochar
Pelham Bay
Huguenot
Co-op City
Forest Hills
Mariners Harbor
Throgs Neck
Morningside Heights
Sunset Park
Crown Heights
Bay Ridge
Edgemere
Dyker Heights
Bedford-Stuyvesant
Manhattan Beach
Middle Village
Rosebank
Canarsie
Prospect-Lefferts Gardens
Astoria
Wakefield
Van Nest
Flushing
Flatlands
Mott Haven
Williamsbridge
Jamaica
Brighton Beach
College Point
Edenwald
Far Rockaway
Ditmars Steinway
Longwood
Neighbourhood in NYC

g3

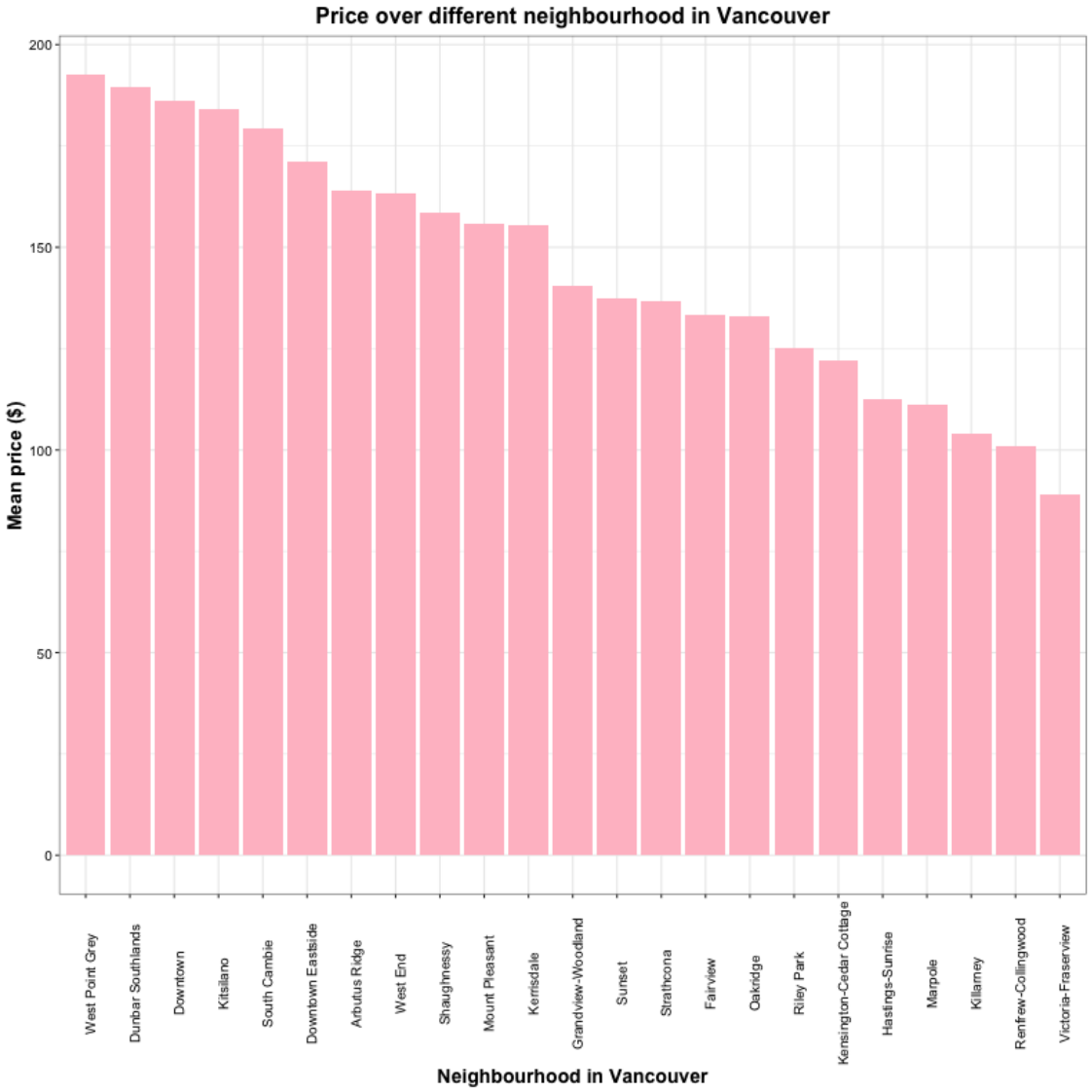

Price over different neighbourhood in NYC(3)
Mean price ($)
50
0
Sheepshead Bay
Oakwood
Queens Village
East Flatbush
Bay Terrace, Staten Island
Coney Island
St. Albans
North Riverdale
Fort Hamilton
Flatbush
Briarwood
Howland Hook
Allerton
South Beach
Spuyten Duyvil
Howard Beach
Richmond Hill
Concourse
Inwood
Washington Heights
Sunnyside
Roosevelt Island
East Morrisania
Clifton
Kensington
Pelham Gardens
Midwood
Mill Basin
Brownsville
Great Kills
Midland Beach
Claxon Point
Morris Park
Bensonhurst
East New York
Kew Gardens
Springfield Gardens
West Brighton
East Elmhurst
Marble Hill
Ozone Park
Maspeth
Morrisania
Gravesend
Rego Park
Jackson Heights
Bushwick
Cypress Hills
Tompkinsville
South Ozone Park
Rosedale
Woodside
Mount Hope
Bayswater
Richmondtown
Glendale
Bath Beach
Ridgewood
Elmhurst
Belmont
Neighbourhood in NYC

g4




```
#grid.arrange(g1,g2,g3,g4)


###Vancouver Neightbourhood ~ Price
airbnb_va_nei <- airbnb_va %>% filter(price > 0 & price <= 1000)
neigh_va <- airbnb_va_nei %>% dplyr::select(price, neighbourhood_cleansed)
neigh_va <- neigh_va  %>% group_by(neighbourhood_cleansed) %>% summarise(mean_price = mean(price))
ggplot(neigh_va, aes(y=mean_price, x =
                     reorder(x=neighbourhood_cleansed,-mean_price),
                     fill=as.factor(neighbourhood_cleansed))) +
      geom_bar(stat="identity", fill="pink", position = "dodge") +
      scale_x_discrete(name ="Neighbourhood in Vancouver") +
      scale_y_continuous(name = "Mean price ($)", breaks = seq(0, 1000, by = 50))+ theme_bw() +
      ggtitle("Price over different neighbourhood in Vancouver") +
      theme(axis.title.x = element_text(face="bold",  size=14),
            axis.title.y = element_text(face="bold",  size=14),
            plot.title = element_text(size=16, face="bold"),
            axis.text.x=element_text(angle=90, colour="black", size = 10),
```

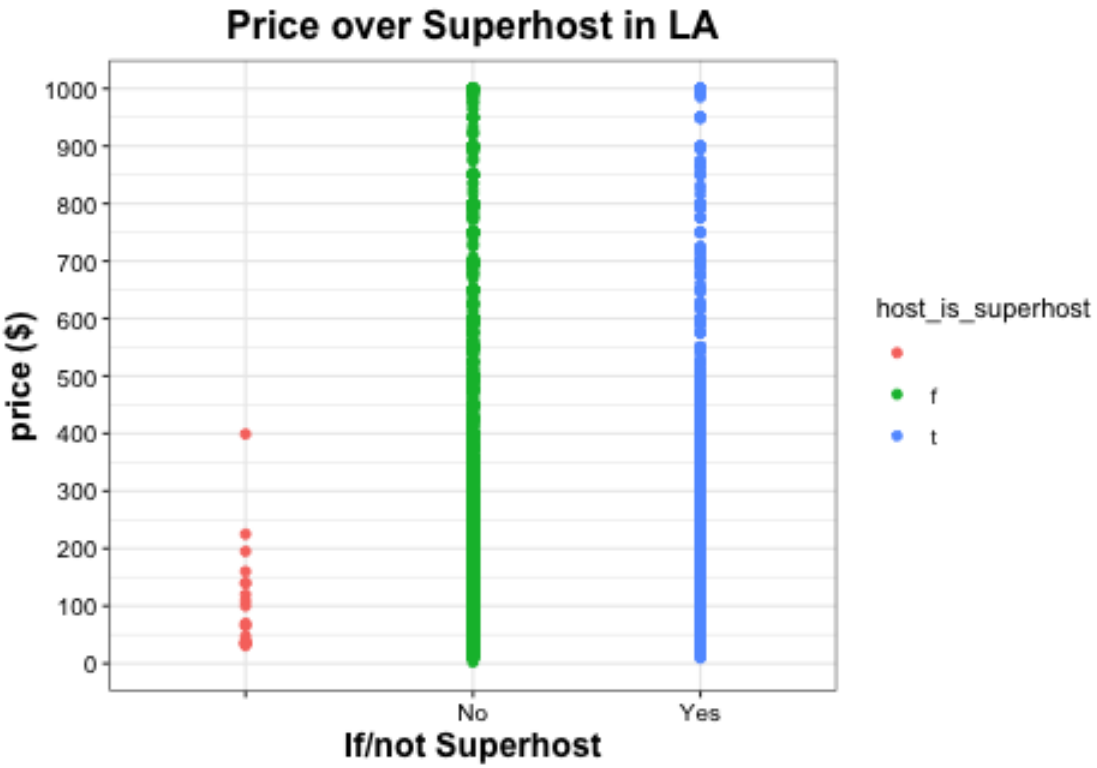
```
        axis.text.y=element_text(colour="black", size = 10)) +
  theme(plot.title = element_text(hjust = 0.5))
```



Price ~ SuperHost

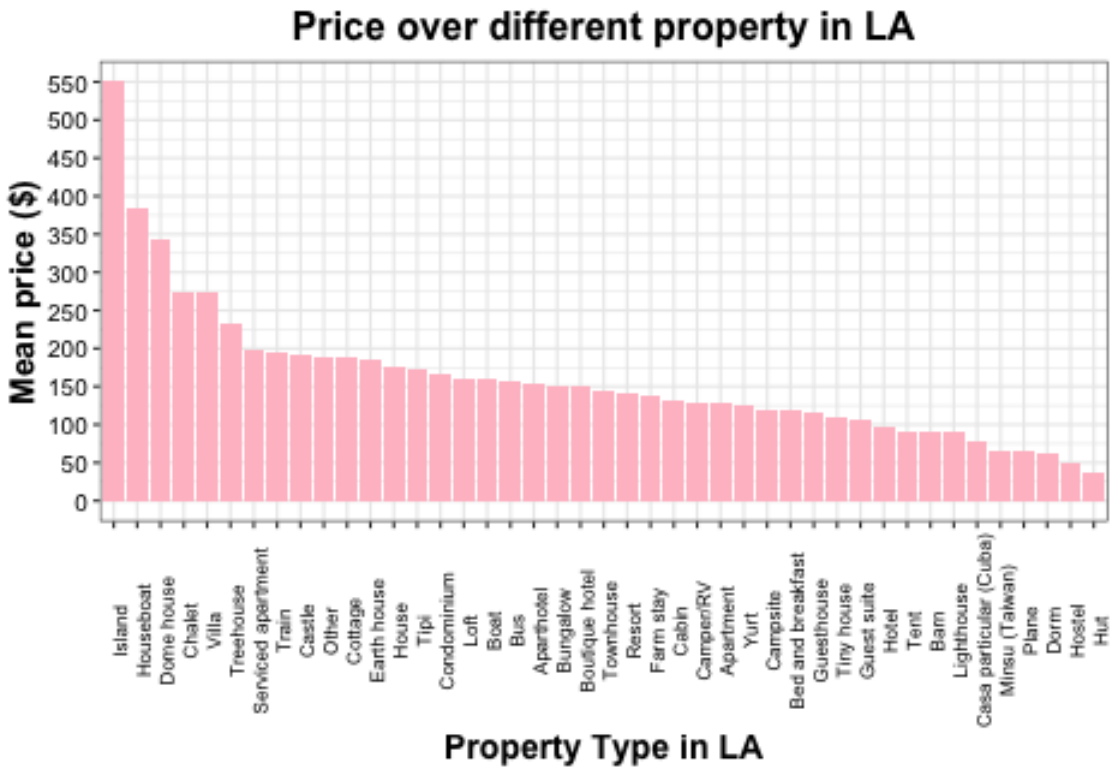
```
airbnb_la_bed <- airbnb_la %>% filter(price > 0 & price <= 1000)
airbnb_la_bed <- airbnb_la_bed %>% na.omit(host_is_superhost)
ggplot(airbnb_la_bed, aes(x=host_is_superhost, y=price)) +
      geom_point( aes(color=host_is_superhost)) + geom_smooth(method = "glm") +
      scale_x_discrete(name ="If/not Superhost",labels=c("f"="No" ,"t"="Yes")) +
      scale_y_continuous(name = "price ($)", breaks = seq(0, 1000, by = 100))+ theme_bw() +
      ggtitle("Price over Superhost in LA") +
      theme(axis.title.x = element_text(face="bold",  size=14),
            axis.title.y = element_text(face="bold",  size=14),
            plot.title = element_text(size=16, face="bold"),
            axis.text.x=element_text(colour="black", size = 10),
```

```
            axis.text.y=element_text(colour="black", size = 10)) +
      theme(plot.title = element_text(hjust = 0.5))
```

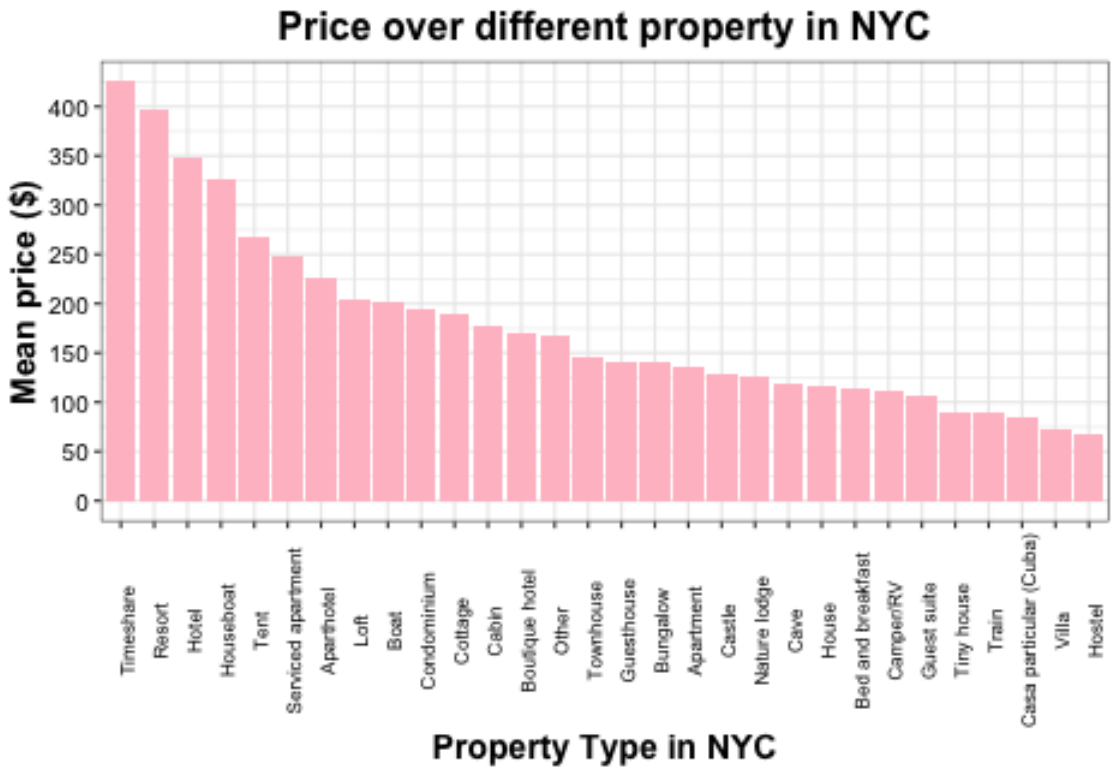


Price ~ Property type

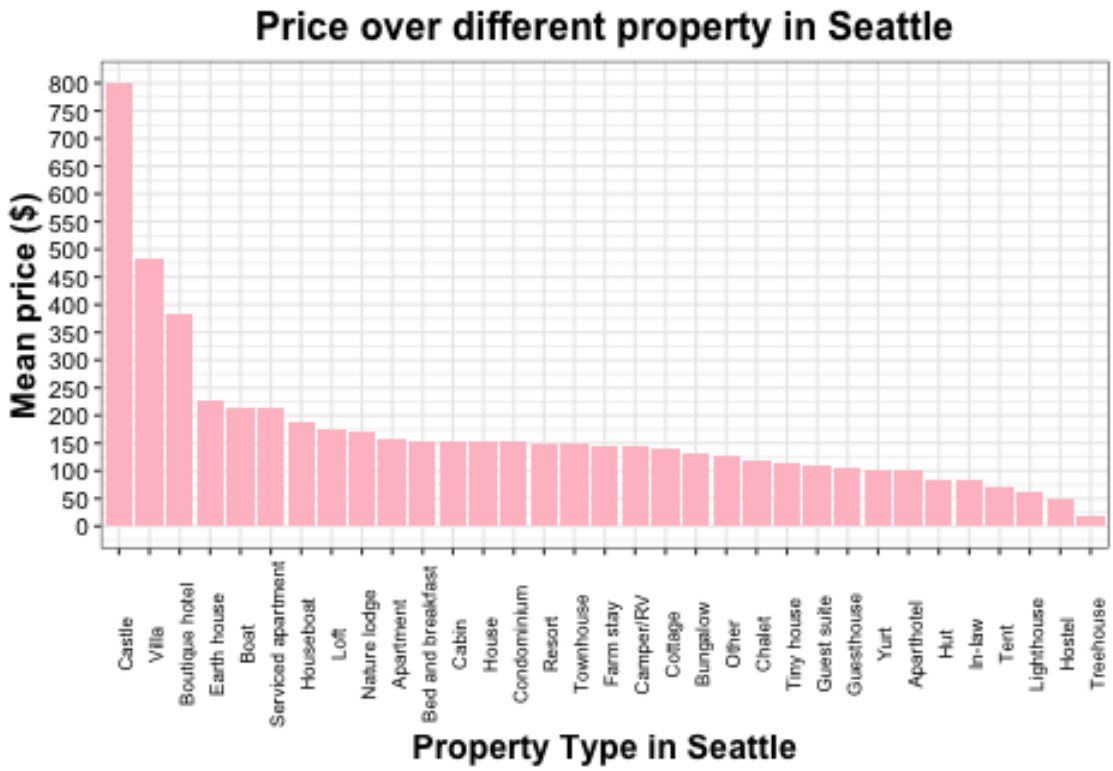
```
###property type in LA
airbnb_la_room <- airbnb_la %>% filter(price > 0 & price <= 1000) %>% dplyr::select(property_type, bedr
room_la <- airbnb_la_room  %>% group_by(property_type) %>% summarise(mean_price = mean(price)) %>% arra
ggplot(room_la, aes(y=mean_price, x =
                    reorder(x=property_type,-mean_price),
                    fill=as.factor(property_type))) +
      geom_bar(stat="identity", fill="pink", position = "dodge") +
      scale_x_discrete(name ="Property Type in LA") +
      scale_y_continuous(name = "Mean price ($)", breaks = seq(0, 1000, by = 50))+ theme_bw() +
      ggtitle("Price over different property in LA") +
      theme(axis.title.x = element_text(face="bold",  size=14),
            axis.title.y = element_text(face="bold",  size=14),
            plot.title = element_text(size=16, face="bold"),
            axis.text.x=element_text(angle=90, colour="black", size = 8),
            axis.text.y=element_text(colour="black", size = 10)) +
      theme(plot.title = element_text(hjust = 0.5))
```



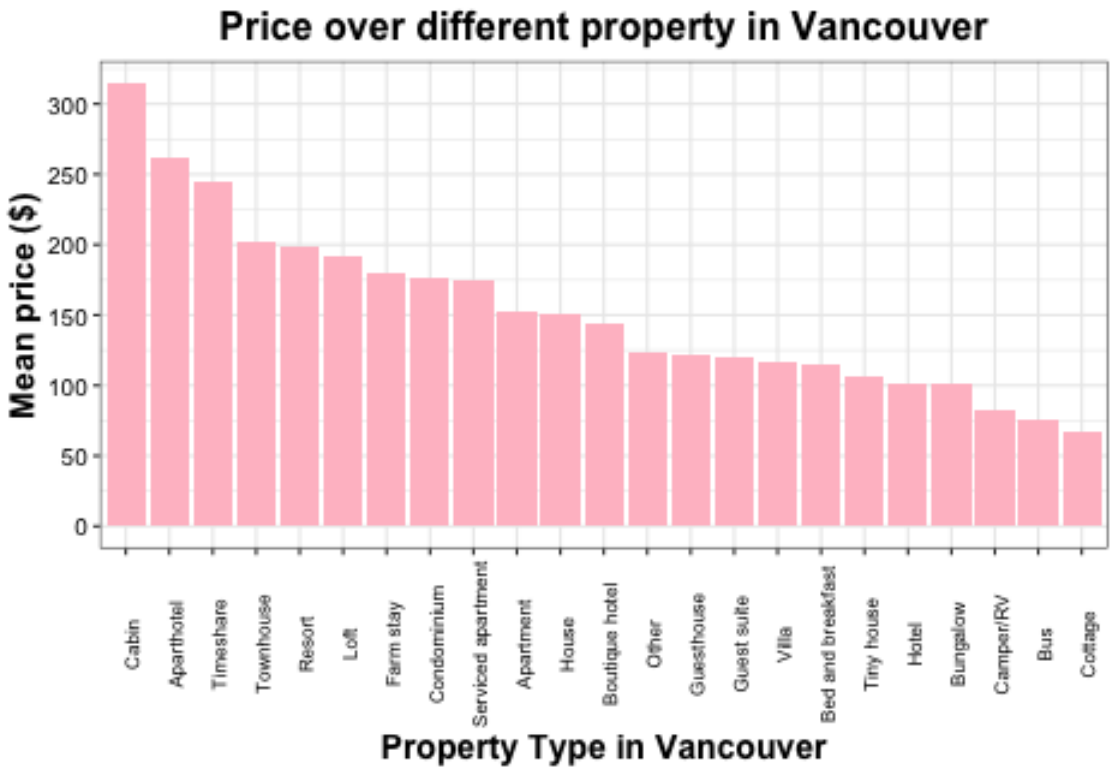

```
###property type in NY
airbnb_ny_room <- airbnb_ny %>% filter(price > 0 & price <= 1000) %>% dplyr::select(property_type, bedr
room_ny <- airbnb_ny_room  %>% group_by(property_type) %>% summarise(mean_price = mean(price)) %>% arra
ggplot(room_ny, aes(y=mean_price, x =
                    reorder(x=property_type,-mean_price),
                    fill=as.factor(property_type))) +
      geom_bar(stat="identity", fill="pink", position = "dodge") +
      scale_x_discrete(name ="Property Type in NYC") +
      scale_y_continuous(name = "Mean price ($)", breaks = seq(0, 1000, by = 50))+ theme_bw() +
      ggtitle("Price over different property in NYC") +
      theme(axis.title.x = element_text(face="bold",  size=14),
            axis.title.y = element_text(face="bold",  size=14),
            plot.title = element_text(size=16, face="bold"),
            axis.text.x=element_text(angle=90, colour="black", size = 8),
            axis.text.y=element_text(colour="black", size = 10)) +
      theme(plot.title = element_text(hjust = 0.5))
```




```
###property type in seattle
airbnb_se_room <- airbnb_se %>% filter(price > 0 & price <= 1000) %>% dplyr::select(property_type, bedr
room_se <- airbnb_se_room  %>% group_by(property_type) %>% summarise(mean_price = mean(price)) %>% arra
ggplot(room_se, aes(y=mean_price, x =
                    reorder(x=property_type,-mean_price),
                    fill=as.factor(property_type))) +
      geom_bar(stat="identity", fill="pink", position = "dodge") +
      scale_x_discrete(name ="Property Type in Seattle") +
      scale_y_continuous(name = "Mean price ($)", breaks = seq(0, 1000, by = 50))+ theme_bw() +
      ggtitle("Price over different property in Seattle") +
      theme(axis.title.x = element_text(face="bold",  size=14),
            axis.title.y = element_text(face="bold",  size=14),
            plot.title = element_text(size=16, face="bold"),
            axis.text.x=element_text(angle=90, colour="black", size = 8),
            axis.text.y=element_text(colour="black", size = 10)) +
      theme(plot.title = element_text(hjust = 0.5))
```

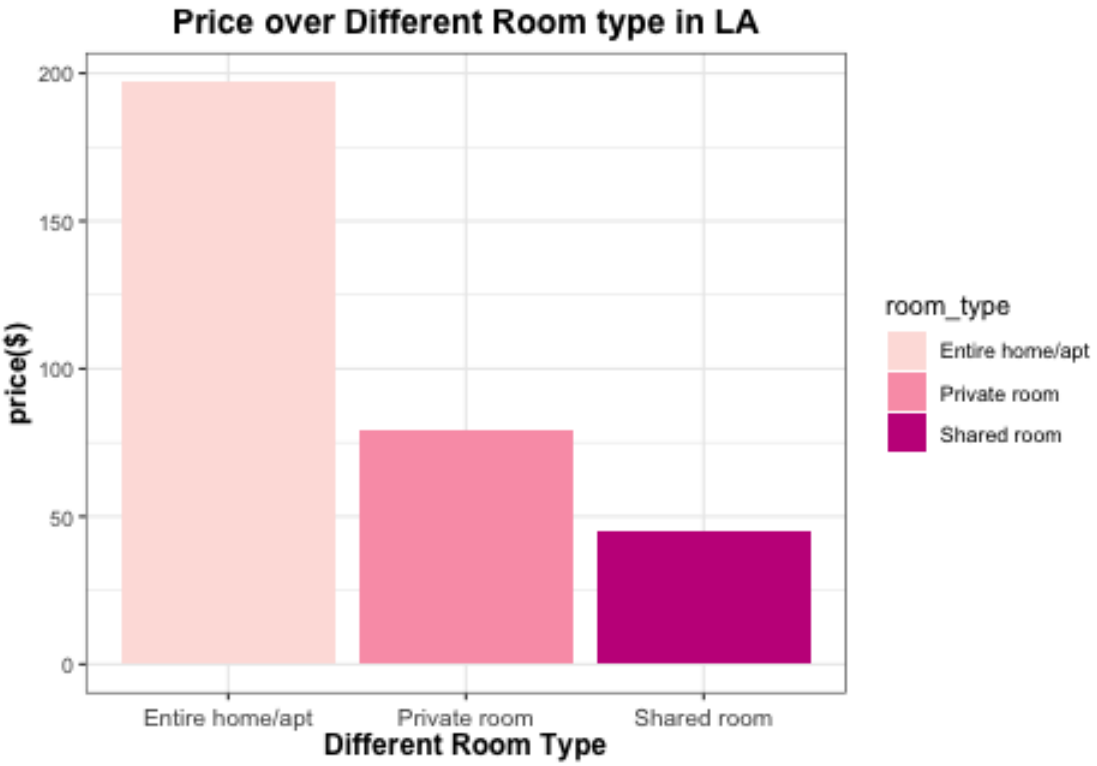



```
###property type in Vancouver
airbnb_va_room <- airbnb_va %>% filter(price > 0 & price <= 1000) %>% dplyr::select(property_type, bedr
room_va <- airbnb_va_room  %>% group_by(property_type) %>% summarise(mean_price = mean(price)) %>% arra
ggplot(room_va, aes(y=mean_price, x =
                    reorder(x=property_type,-mean_price),
                    fill=as.factor(property_type))) +
      geom_bar(stat="identity", fill="pink", position = "dodge") +
      scale_x_discrete(name ="Property Type in Vancouver") +
      scale_y_continuous(name = "Mean price ($)", breaks = seq(0, 1000, by = 50))+ theme_bw() +
      ggtitle("Price over different property in Vancouver") +
      theme(axis.title.x = element_text(face="bold",  size=14),
            axis.title.y = element_text(face="bold",  size=14),
            plot.title = element_text(size=16, face="bold"),
            axis.text.x=element_text(angle=90, colour="black", size = 8),
            axis.text.y=element_text(colour="black", size = 10)) +
      theme(plot.title = element_text(hjust = 0.5))
```

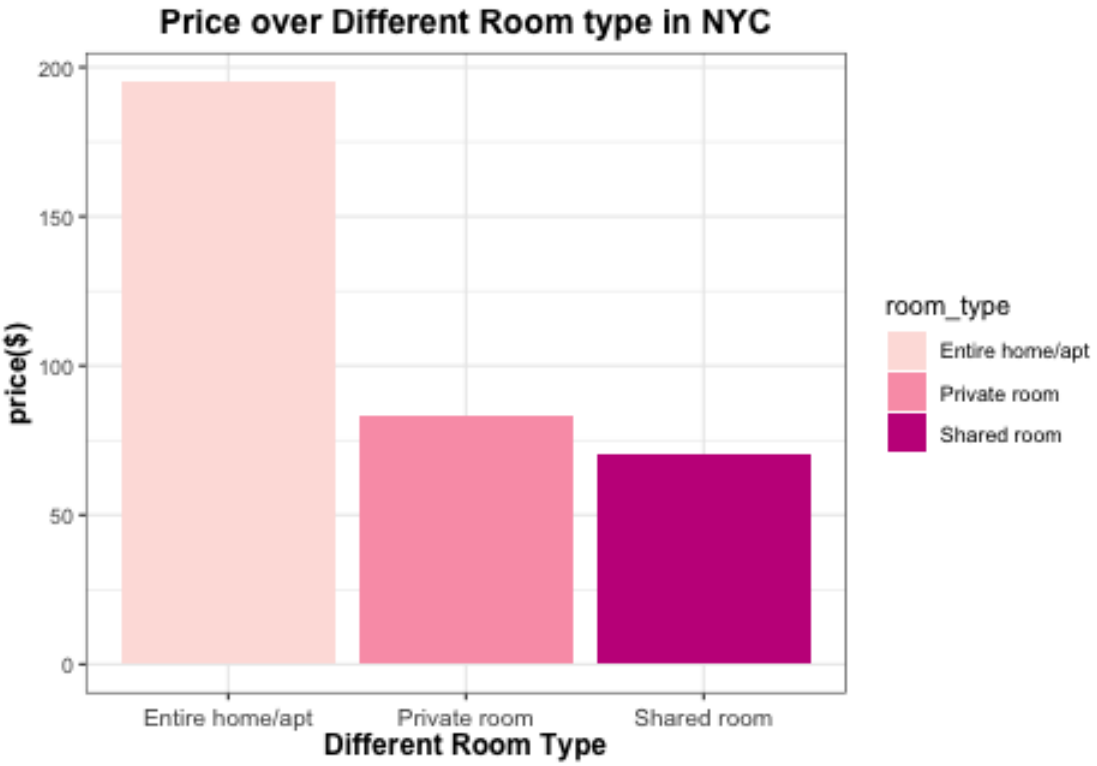


Price ~ room type

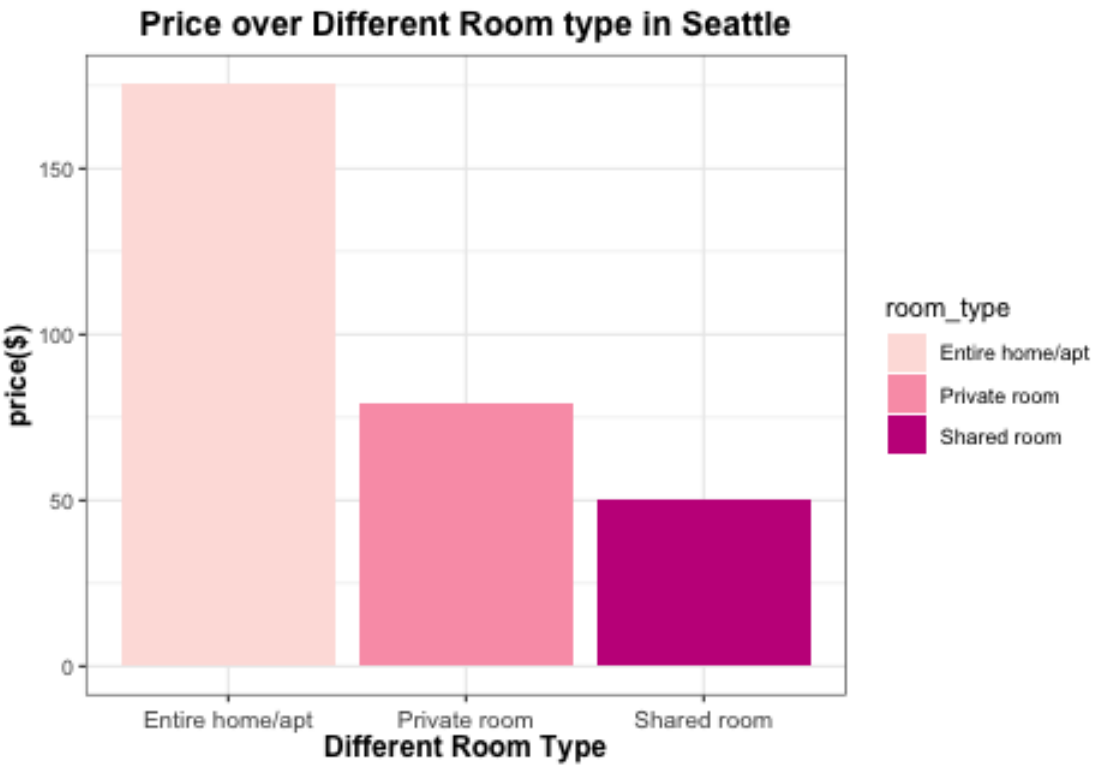
```
## Room type in LA
airbnb_la_type <- airbnb_la %>% filter(price > 0 & price <= 1000) %>% dplyr::select(room_type, bedrooms
type_la <- airbnb_la_type  %>% group_by(room_type) %>% summarise(mean_price = mean(price)) %>% arrange(
ggplot(type_la, aes(y=mean_price, x=room_type, fill=room_type)) +
      geom_bar(stat="identity") + scale_fill_brewer(palette = "RdPu") +
      scale_x_discrete(name ="Different Room Type") +
      scale_y_continuous(name="price($)", breaks = seq(0, 1000, by = 50)) +
      ggtitle("Price over Different Room type in LA") +
      theme_bw() +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold"),
            axis.text.x  = element_text( vjust=0.5, size=10)) +
      theme(plot.title = element_text(hjust = 0.5))
```



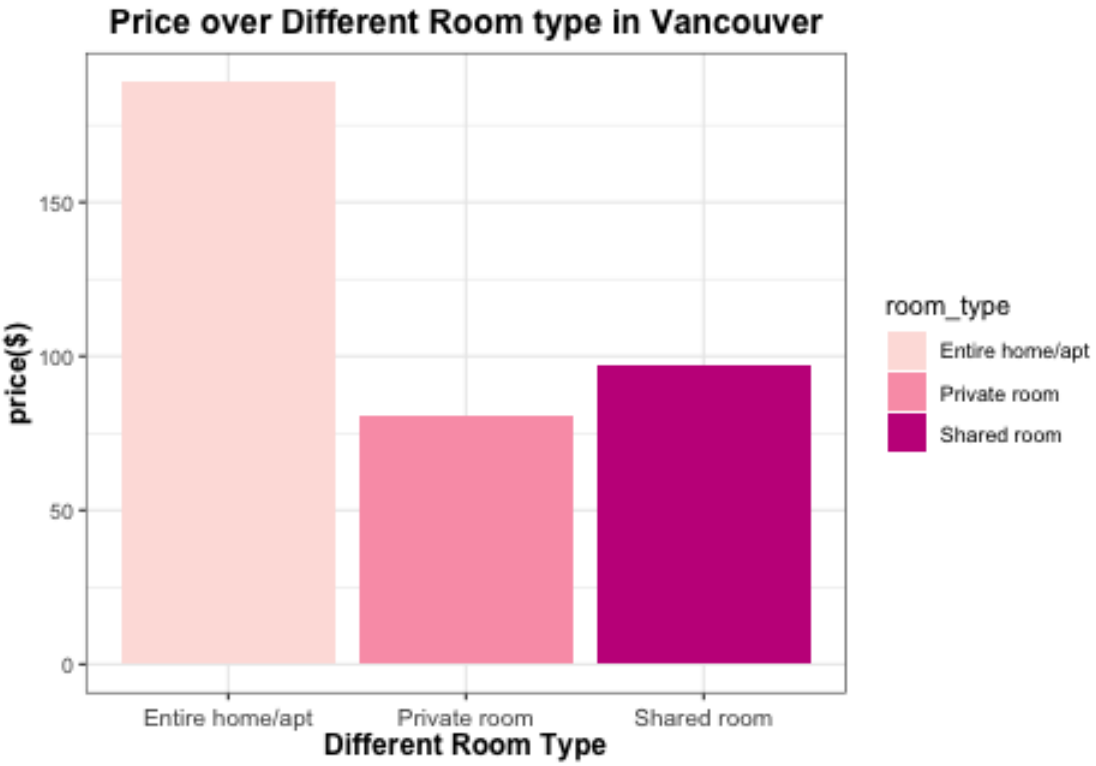

```
## Room type in NY
airbnb_ny_type <- airbnb_ny %>% filter(price > 0 & price <= 1000) %>% dplyr::select(room_type, bedrooms
type_ny <- airbnb_ny_type  %>% group_by(room_type) %>% summarise(mean_price = mean(price)) %>% arrange(
ggplot(type_ny, aes(y=mean_price, x=room_type, fill=room_type)) +
      geom_bar(stat="identity") + scale_fill_brewer(palette = "RdPu") +
      scale_x_discrete(name ="Different Room Type") +
      scale_y_continuous(name="price($)", breaks = seq(0, 1000, by = 50)) +
      ggtitle("Price over Different Room type in NYC") +
      theme_bw() +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold"),
            axis.text.x  = element_text( vjust=0.5, size=10)) +
      theme(plot.title = element_text(hjust = 0.5))
```




```
## Room type in Seattle
airbnb_se_type <- airbnb_se %>% filter(price > 0 & price <= 1000) %>% dplyr::select(room_type, bedrooms
type_se <- airbnb_se_type  %>% group_by(room_type) %>% summarise(mean_price = mean(price)) %>% arrange(
ggplot(type_se, aes(y=mean_price, x=room_type, fill=room_type)) +
      geom_bar(stat="identity") + scale_fill_brewer(palette = "RdPu") +
      scale_x_discrete(name ="Different Room Type") +
      scale_y_continuous(name="price($)", breaks = seq(0, 1000, by = 50)) +
      ggtitle("Price over Different Room type in Seattle") +
      theme_bw() +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold"),
            axis.text.x  = element_text( vjust=0.5, size=10)) +
      theme(plot.title = element_text(hjust = 0.5))
```




```
## Room type in Vancouver
airbnb_va_type <- airbnb_va %>% filter(price > 0 & price <= 1000) %>% dplyr::select(room_type, bedrooms
type_va <- airbnb_va_type  %>% group_by(room_type) %>% summarise(mean_price = mean(price)) %>% arrange(
ggplot(type_va, aes(y=mean_price, x=room_type, fill=room_type)) +
      geom_bar(stat="identity") + scale_fill_brewer(palette = "RdPu") +
      scale_x_discrete(name ="Different Room Type") +
      scale_y_continuous(name="price($)", breaks = seq(0, 1000, by = 50)) +
      ggtitle("Price over Different Room type in Vancouver") +
      theme_bw() +
      theme(axis.title.x = element_text(face="bold",  size=12),
            axis.title.y = element_text(face="bold",  size=12),
            plot.title = element_text(size=14, face="bold"),
            axis.text.x  = element_text( vjust=0.5, size=10)) +
      theme(plot.title = element_text(hjust = 0.5))
```


Price over Different Room type in Vancouver
price($)
150
100
50
0
Entire home/apt
Private room
Shared room
Different Room Type
room_type
Entire home/apt
Private room
Shared room