

EDA on Used Cars Market in India

```
knitr::opts_chunk$set(echo = TRUE)
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.3      v dplyr   1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

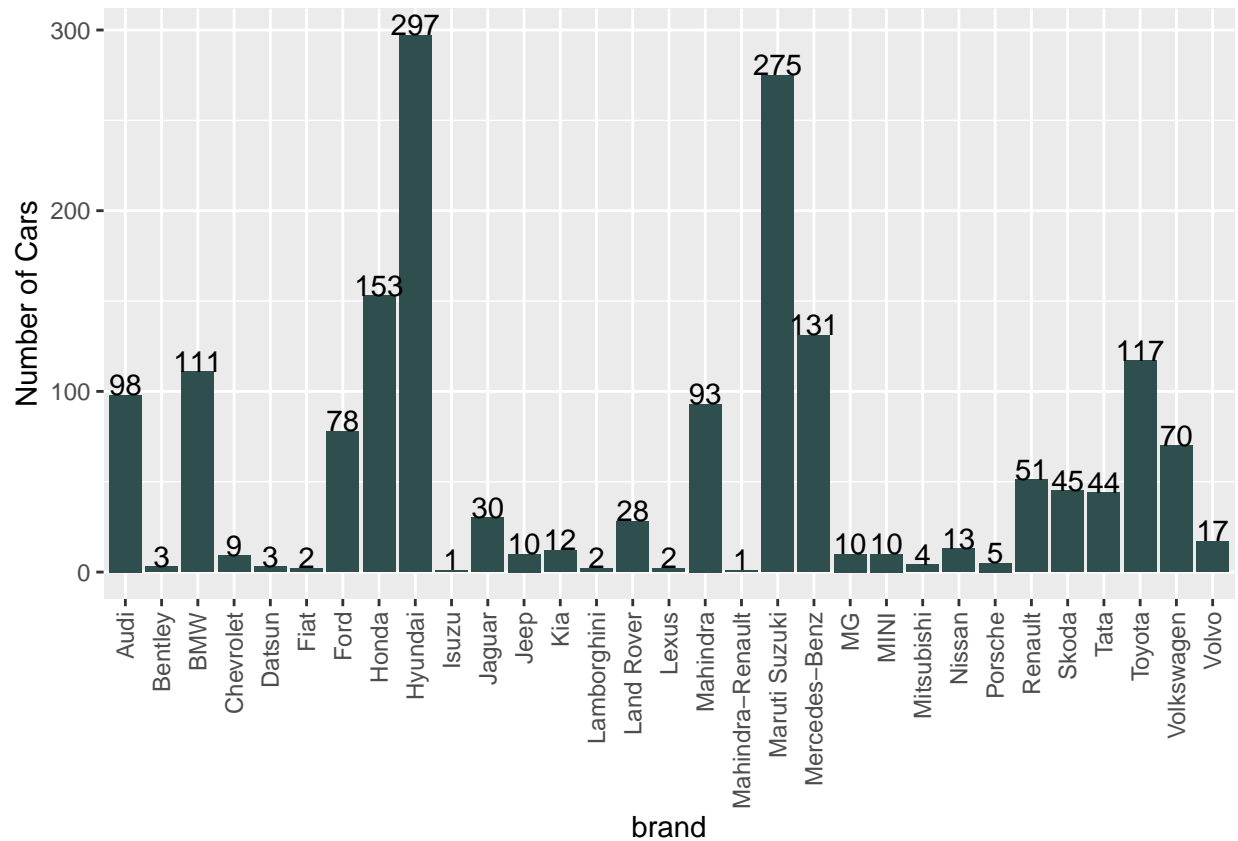
```
setwd("~/Desktop/Projects /Used_Cars")
```

```
#Importing Data
```

```
ds1 <- read.csv("UsedCars_Bangalore.csv")
ds2 <- read.csv("UsedCars_Chennai.csv")
ds3 <- read.csv("UsedCars_Delhi.csv")
ds4 <- read.csv("UsedCars_Hyderabad.csv")
ds5 <- read.csv("UsedCars_Kolkata.csv")
ds6 <- read.csv("UsedCars_Mumbai.csv")
ds7 <- read.csv("UsedCars_Pune.csv")
```

```
main_data <- rbind(ds1,ds2,ds3,ds4,ds5,ds6,ds7)
```

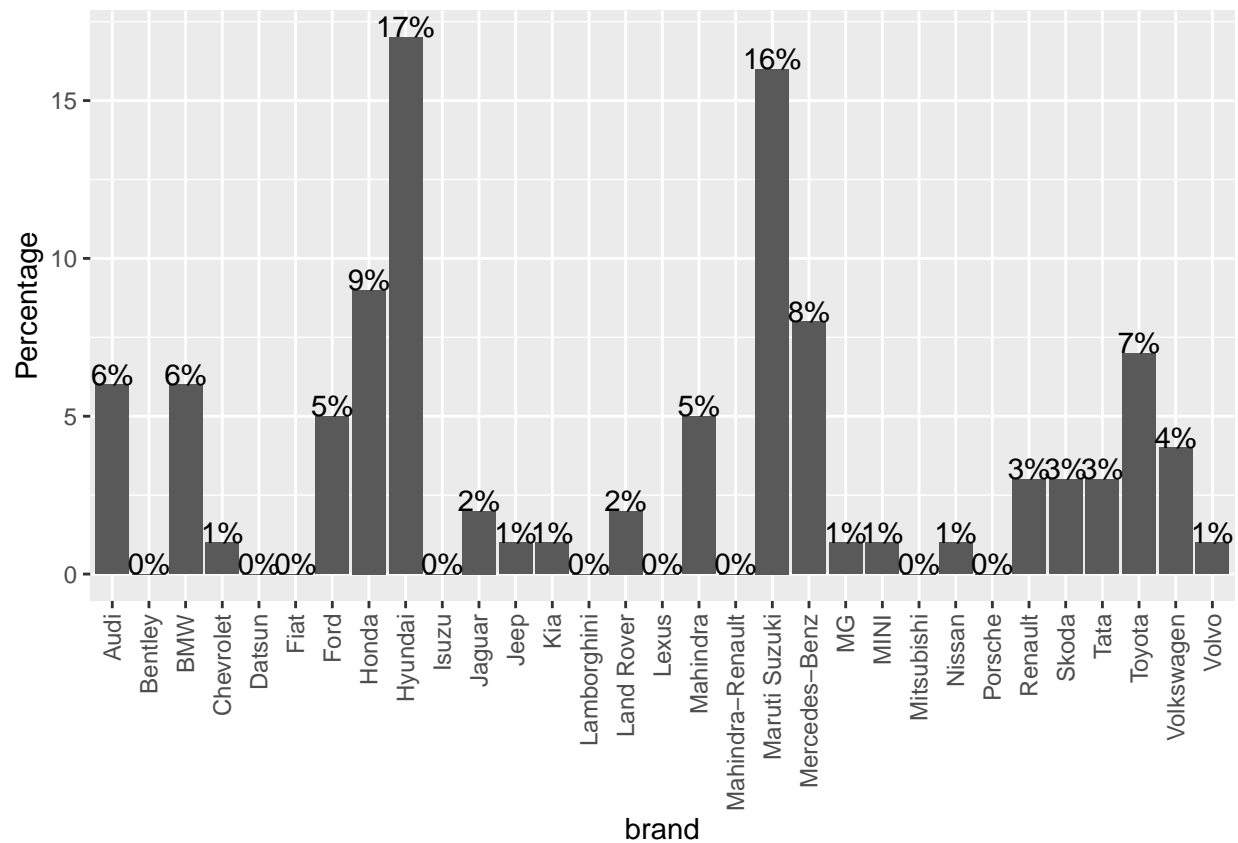
```
main_data %>%
  count(brand) %>%
  arrange(desc(n)) %>%
  ggplot(aes(brand,n))+geom_bar(stat = "identity", fill = "#2F4F4F")+geom_text(aes(label = n,vjust = 0),
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
  labs (y = "Number of Cars")
```



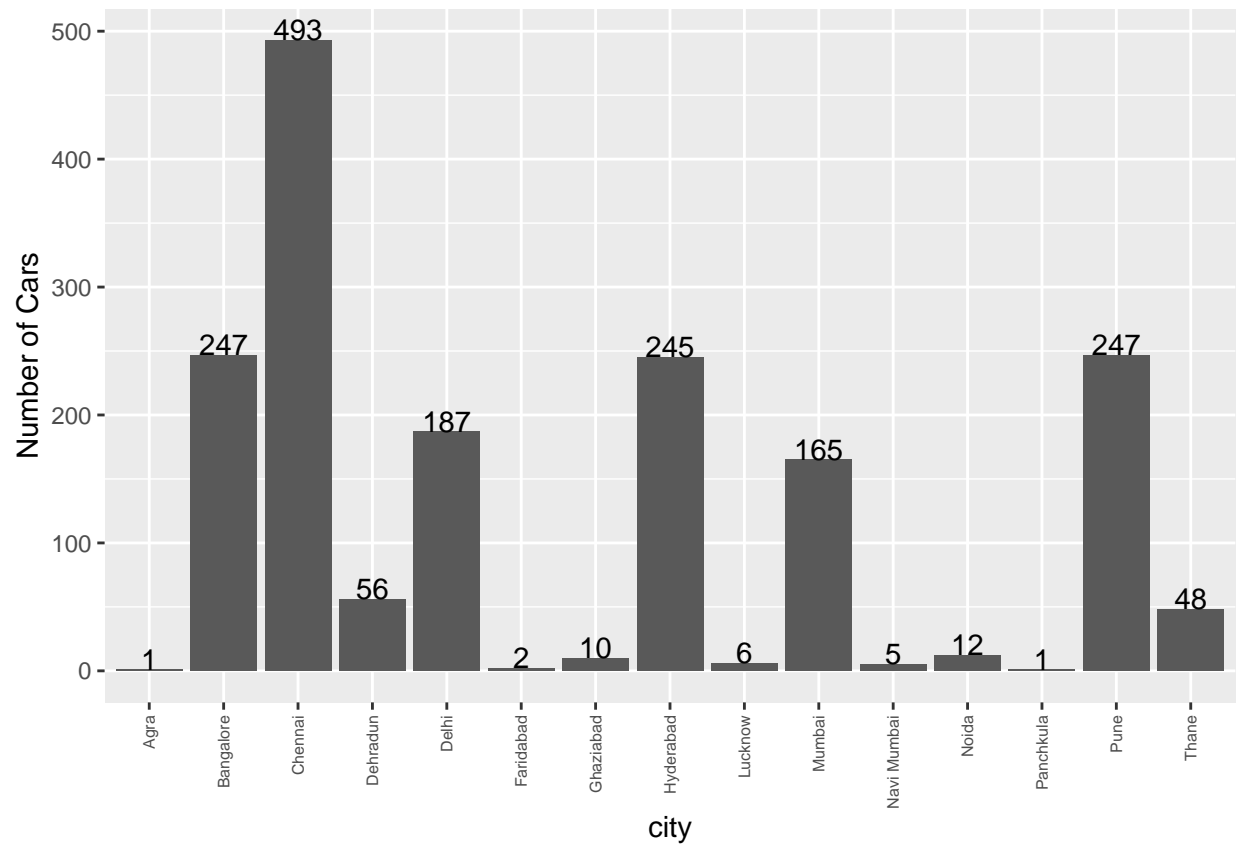
~~

```
main_data <- main_data %>%
  rename( ID = 1)
```

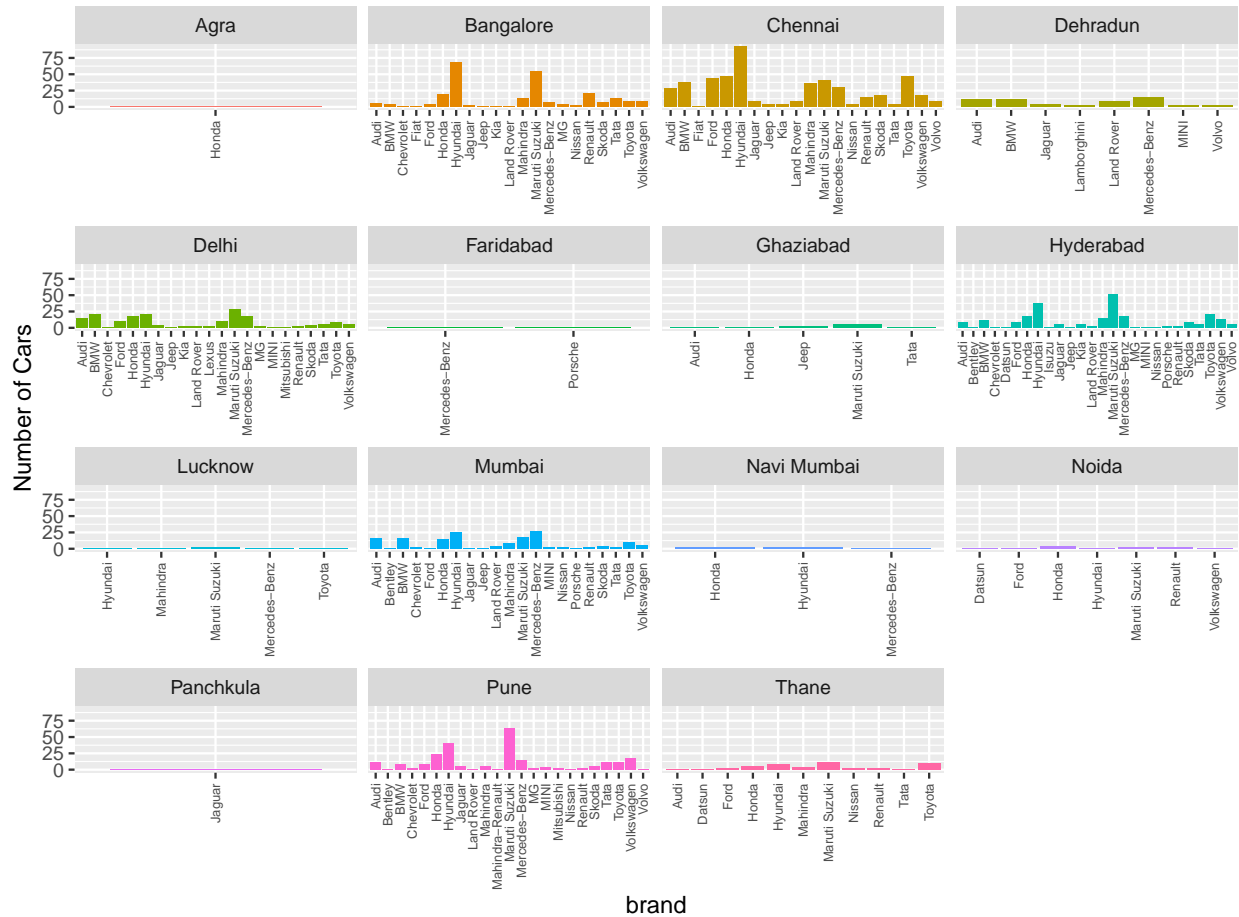
```
percentage <- main_data %>%
  group_by(brand) %>%
  summarise(Total = length(ID),
            Percentage = round(length(ID) / nrow(main_data) * 100)) %>%
  ggplot(aes(brand,Percentage))+geom_bar(stat = "identity")+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
  geom_text(aes(label = paste0(Percentage, "%"), vjust = 0))
percentage
```



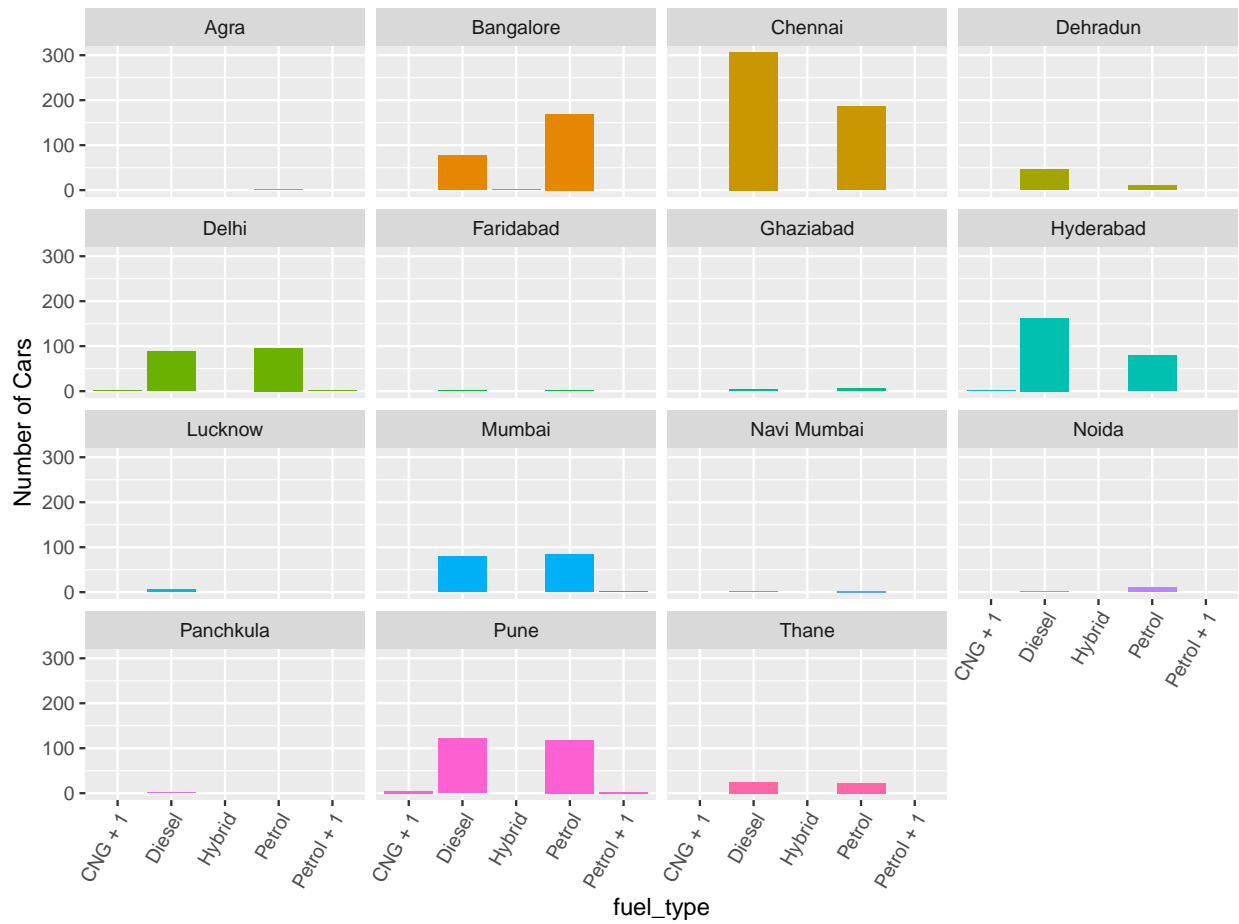
```
main_data %>%
  count(city) %>%
  ggplot(aes(city,n))+geom_bar(stat = "identity")+
  theme(axis.text.x = element_text(angle = 90,size = 6, vjust = 0.5, hjust=1))+
  geom_text(aes(label = n, vjust = 0))+labs(y = "Number of Cars")
```



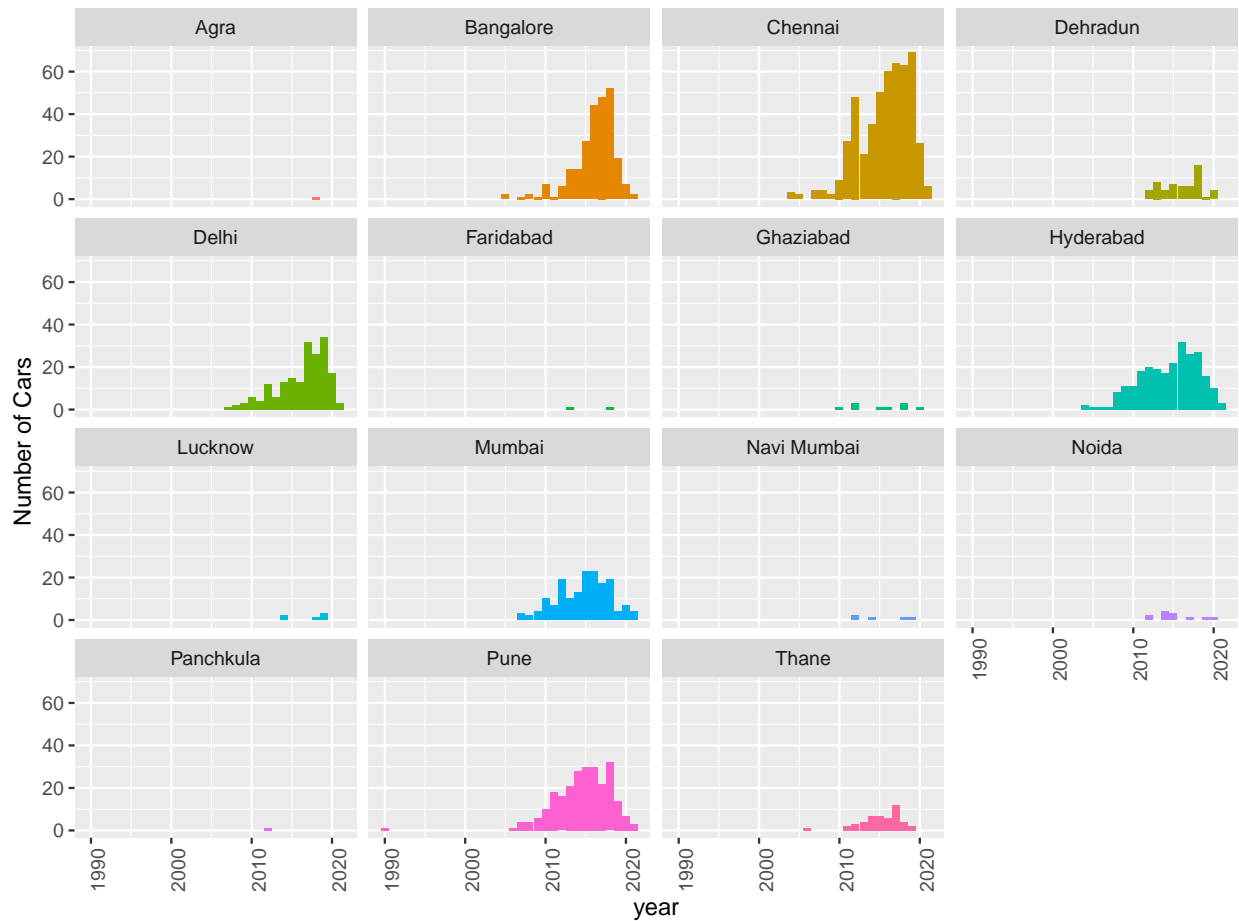
```
main_data %>%
  ggplot(aes(brand, fill = city))+geom_bar(stat = "count",show.legend = FALSE)+
  facet_wrap(~city , scales="free_x") +
  theme( axis.text.x = element_text(angle = 90,size = 6, vjust = 0.5, hjust=1)) +
  labs (y = "Number of Cars")
```



```
main_data %>%
  ggplot(aes(fuel_type, fill = city))+geom_bar(stat = "count",show.legend = FALSE)+facet_wrap(~city)+
  theme( axis.text.x = element_text(angle = 60, vjust = 1, hjust=1)) + labs (y = "Number of Cars")
```



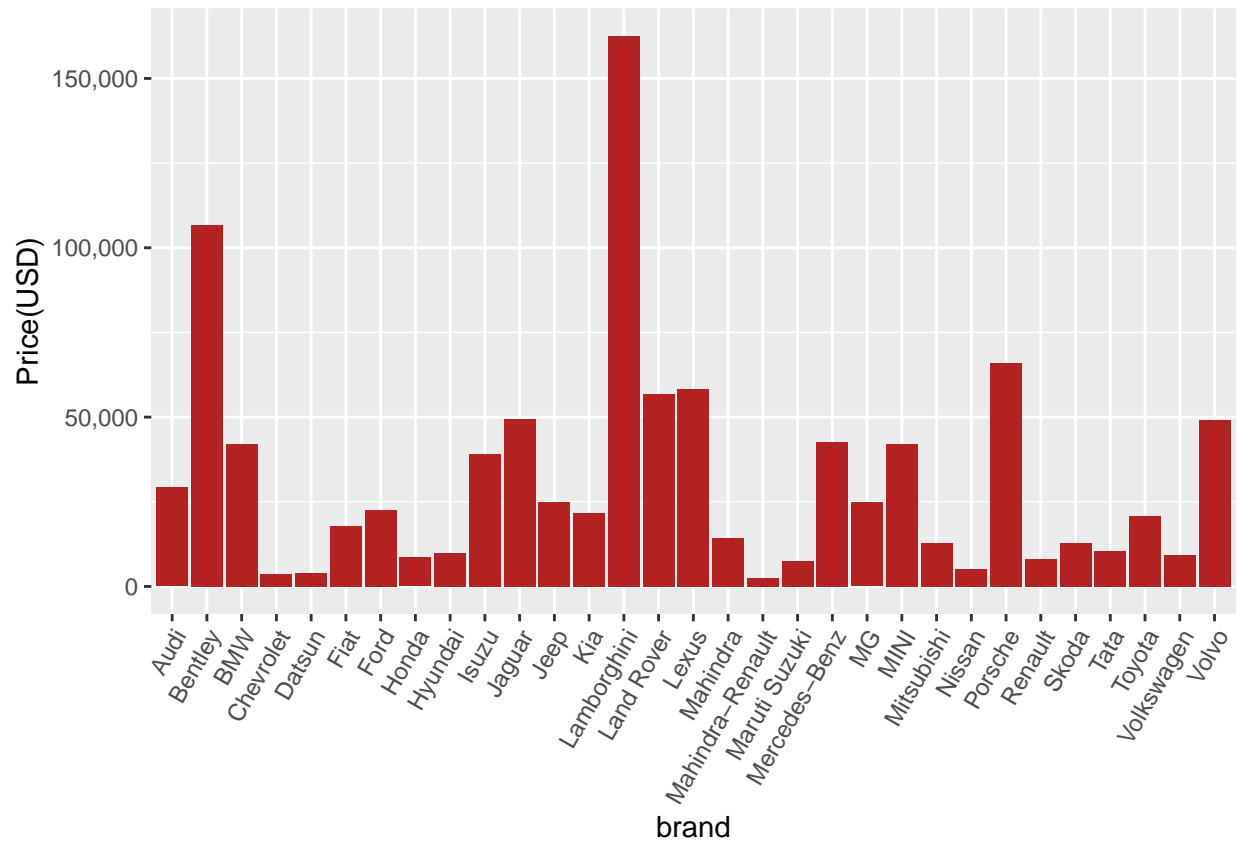
```
main_data %>%
  ggplot(aes(year, fill = city))+geom_bar(stat = "count",show.legend = FALSE)+
  facet_wrap(~city)+
  theme( axis.text.x = element_text(angle = 90, vjust = 1, hjust=1)) +
  labs (y = "Number of Cars")
```



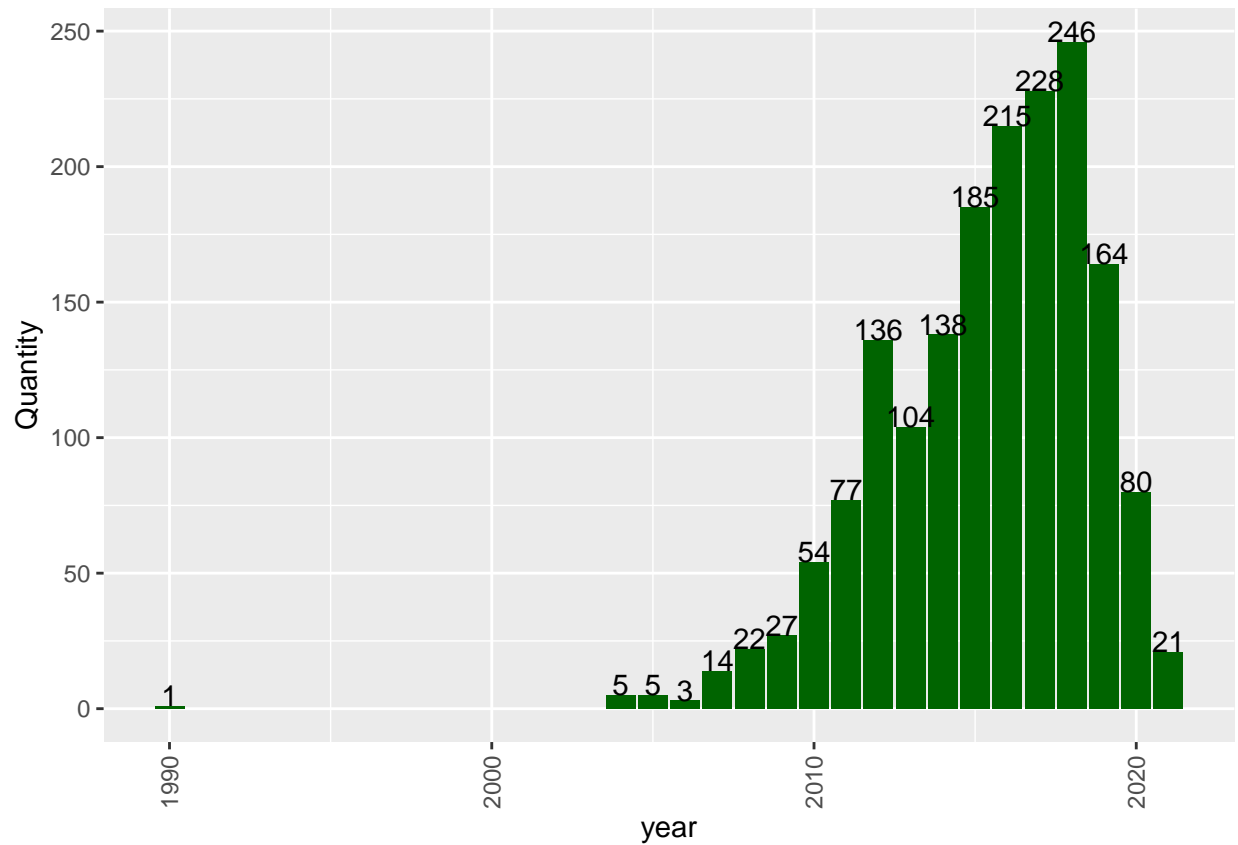
```
main_data <- main_data %>%
  mutate( Price_In_USD = price * 0.013)
```

```
aggregate(Price_In_USD~brand,main_data,mean) %>%
```

```
arrange(desc(Price_In_USD)) %>%
ggplot(aes(brand,Price_In_USD))+geom_bar(stat= "identity", fill = "#B22222")+
theme( axis.text.x = element_text(angle = 60, vjust = 1, hjust=1))+
scale_y_continuous(labels = scales::comma)+ labs(y = "Price(USD)")
```



```
main_data %>%
count(year) %>% ggplot(aes(year,n))+geom_bar(stat = "identity", fill = "#006400")+geom_text(aes(label =
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
labs(y = "Quantity")
```

```
main_data %>%
  group_by(brand) %>%
  summarise(Average = mean(distance_travelled.kms.)) %>%
  ggplot(aes(reorder(brand,Average),Average))+geom_bar(stat = "identity", fill = "#A52A2A")+
  theme(axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=1))+ scale_y_continuous(labels = scales:
  coord_flip()+labs(x = "Brand", y = "Average Distance (km)")
```

