amanchauhan22bce0476.R

Batch1

2024-09-05

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
```

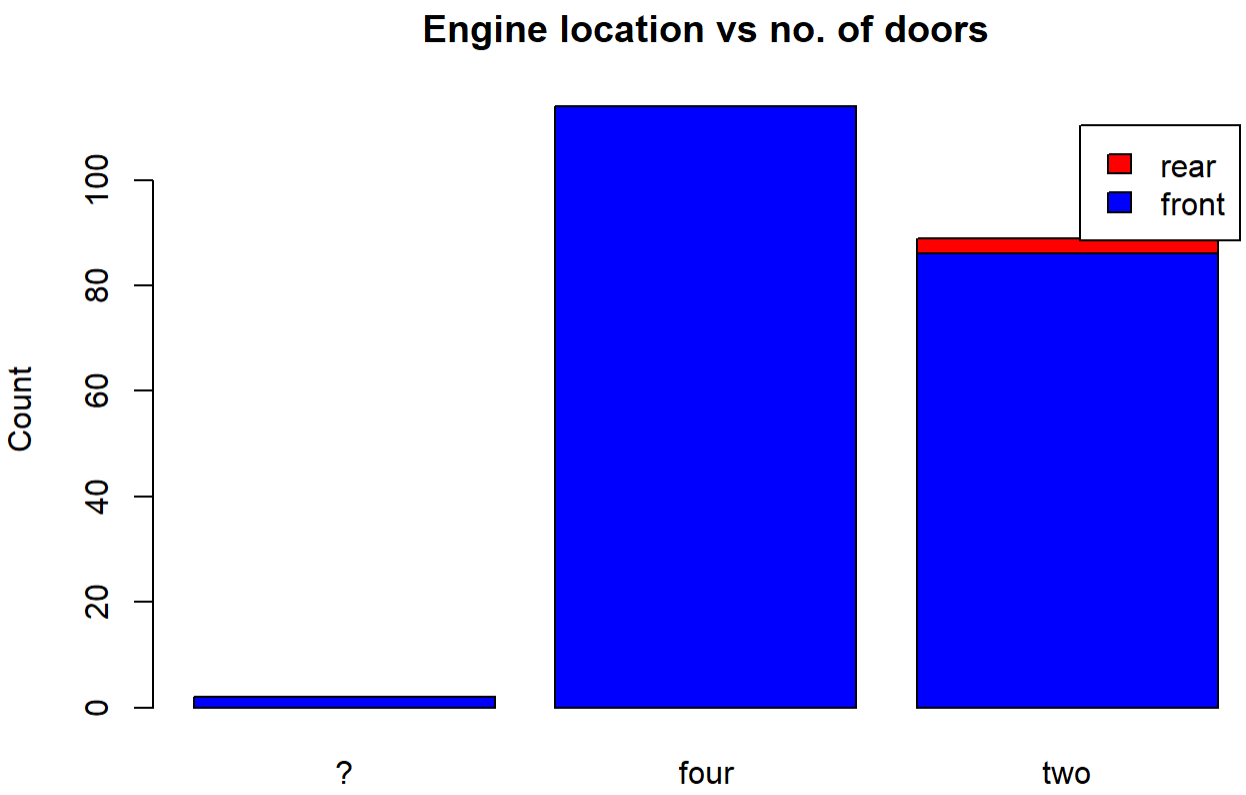
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

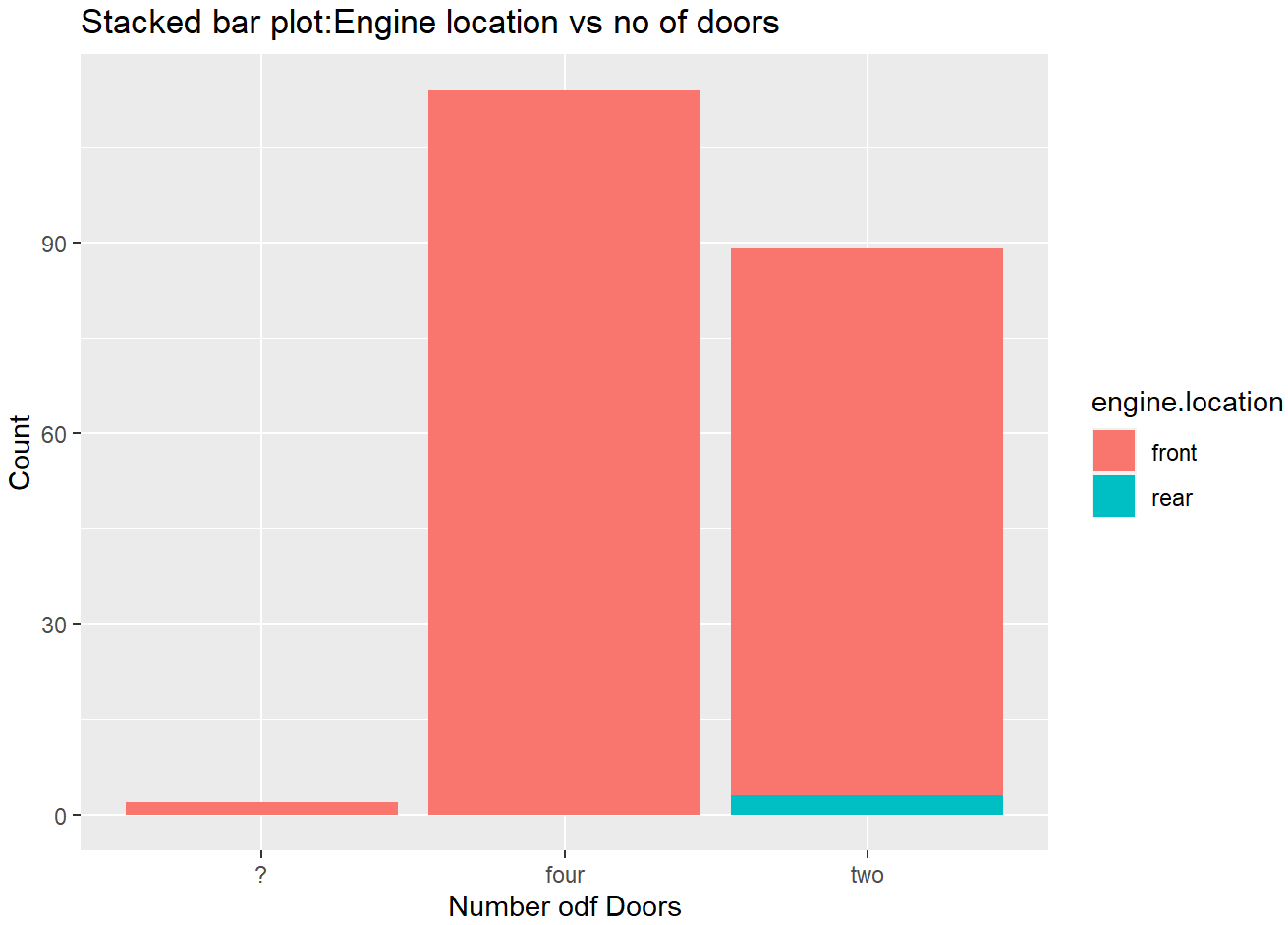
```
#Bivariate Analysis
#1. Load the data.csv dataset into R environment and clean the dataset.
data_clean<-read.csv("C:/Users/batch1/Downloads/data(1).csv")


#2.Create a contingency table for "engine-location" and "num-of-doors" and store in object.
(Use: table())
# Categorical vs. Categorical
cont_table<-table(data_clean$engine.location,data_clean$num.of.doors)
cont_table
```

```
##
##           ? four two
##   front   2  114  86
##   rear    0    0   3
```

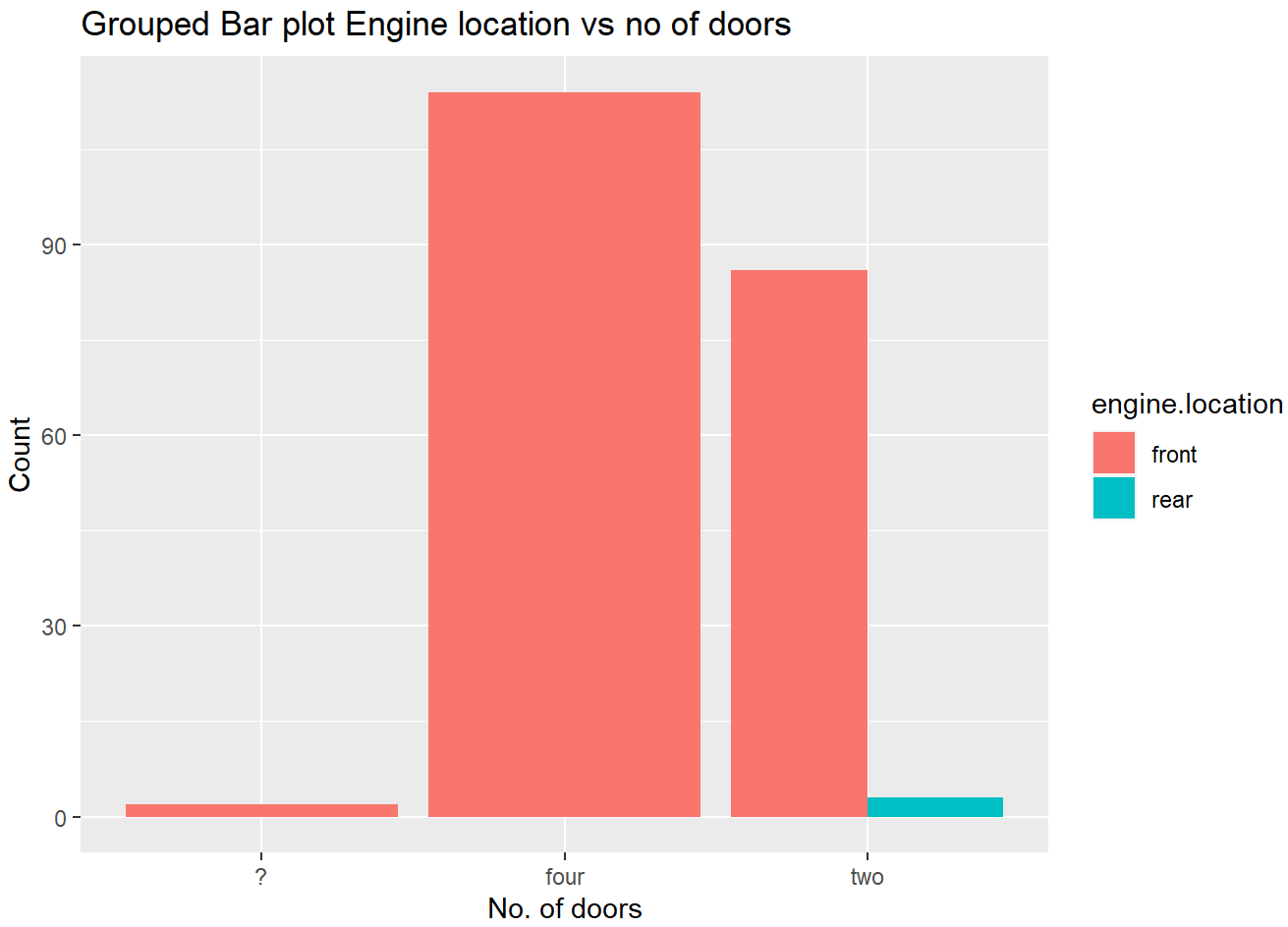
```
#3.1. Display a Stacked bar chart, include Xlab, Ylab, main title.  (barplot(), data from Q2 object)
barplot(cont_table,beside=FALSE,ylab="Count",main="Engine location vs no. of doors",col=c("blue","red"),legend=rownames(cont_table))
```



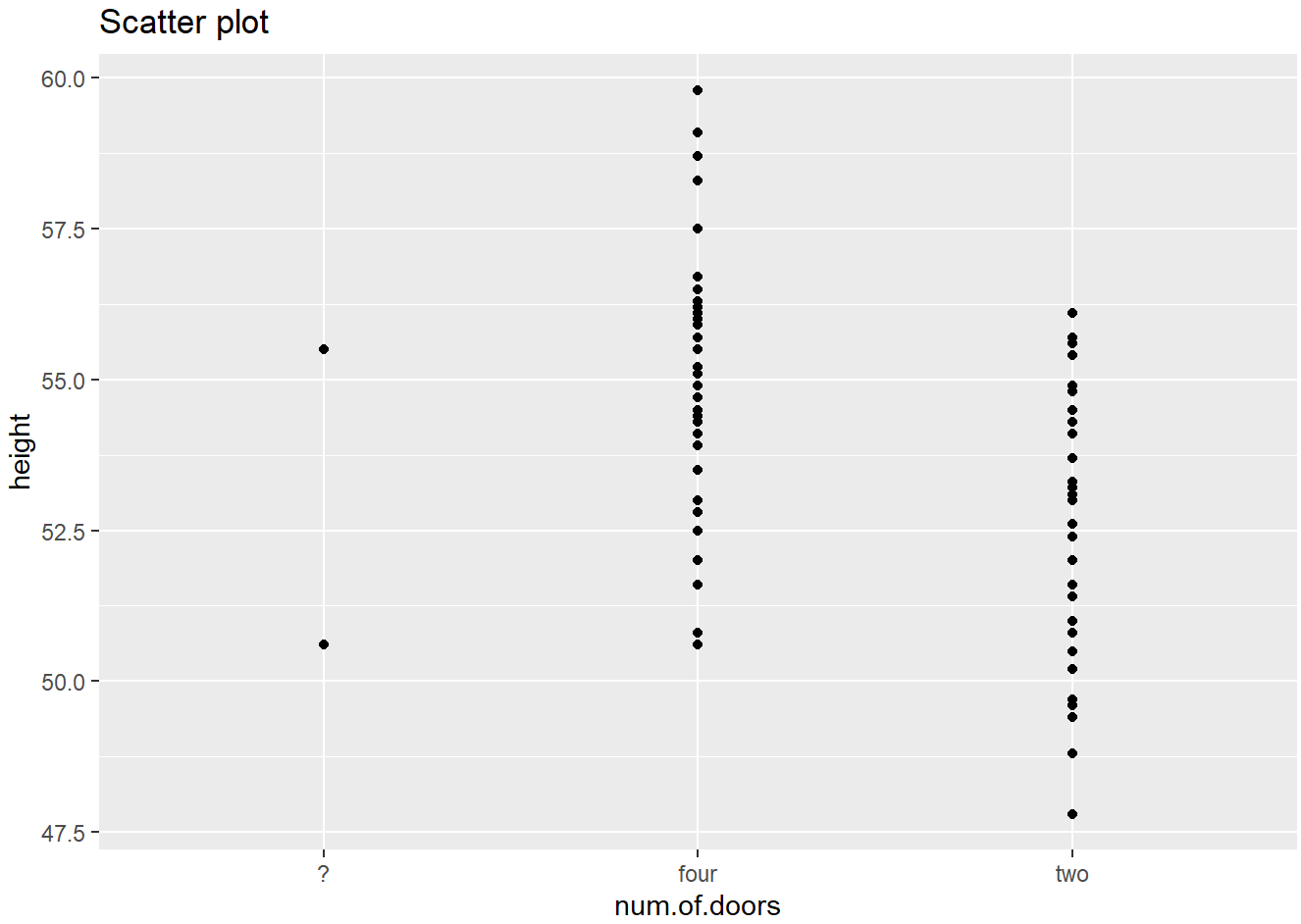

```
#3.2. Display a Stacked bar plot. (ggplot() use fill, geom_bar(position="stack"))
ggplot(data_clean,aes(x=num.of.doors,fill=engine.location))+geom_bar(position = "stack")+labs(x="Number odf Doors",y="Count",title = "Stacked bar plot:Engine location vs no of doors")
```



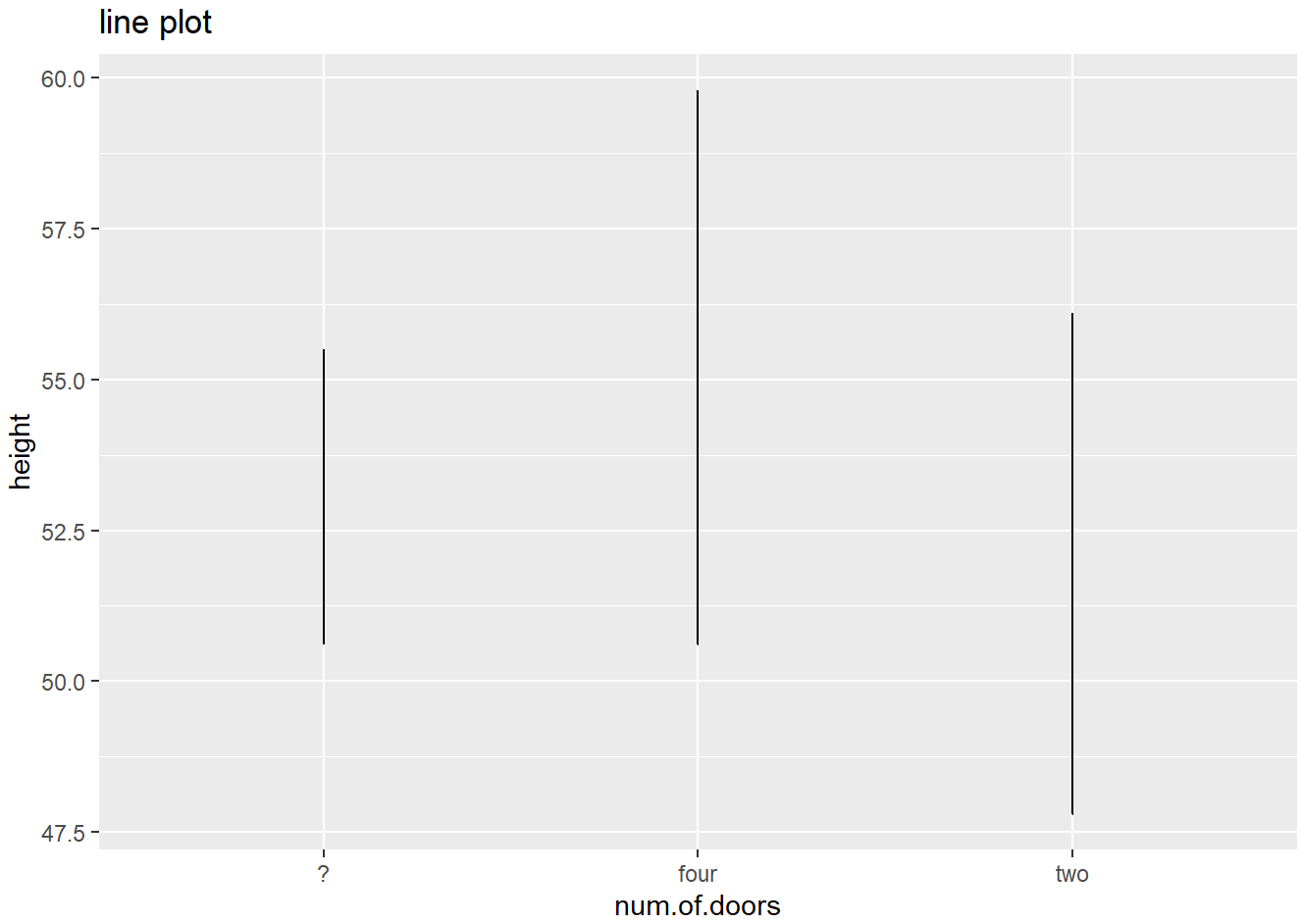

```
#3.3. Display a grouped bar plot or side-by-side plot. (ggplot() use fill, geom_bar(position
="dodge"))
ggplot(data_clean,aes(x=num.of.doors,fill=engine.location))+geom_bar(position="dodge")+labs(x
="No. of doors",y="Count",title="Grouped Bar plot Engine location vs no of doors")
```



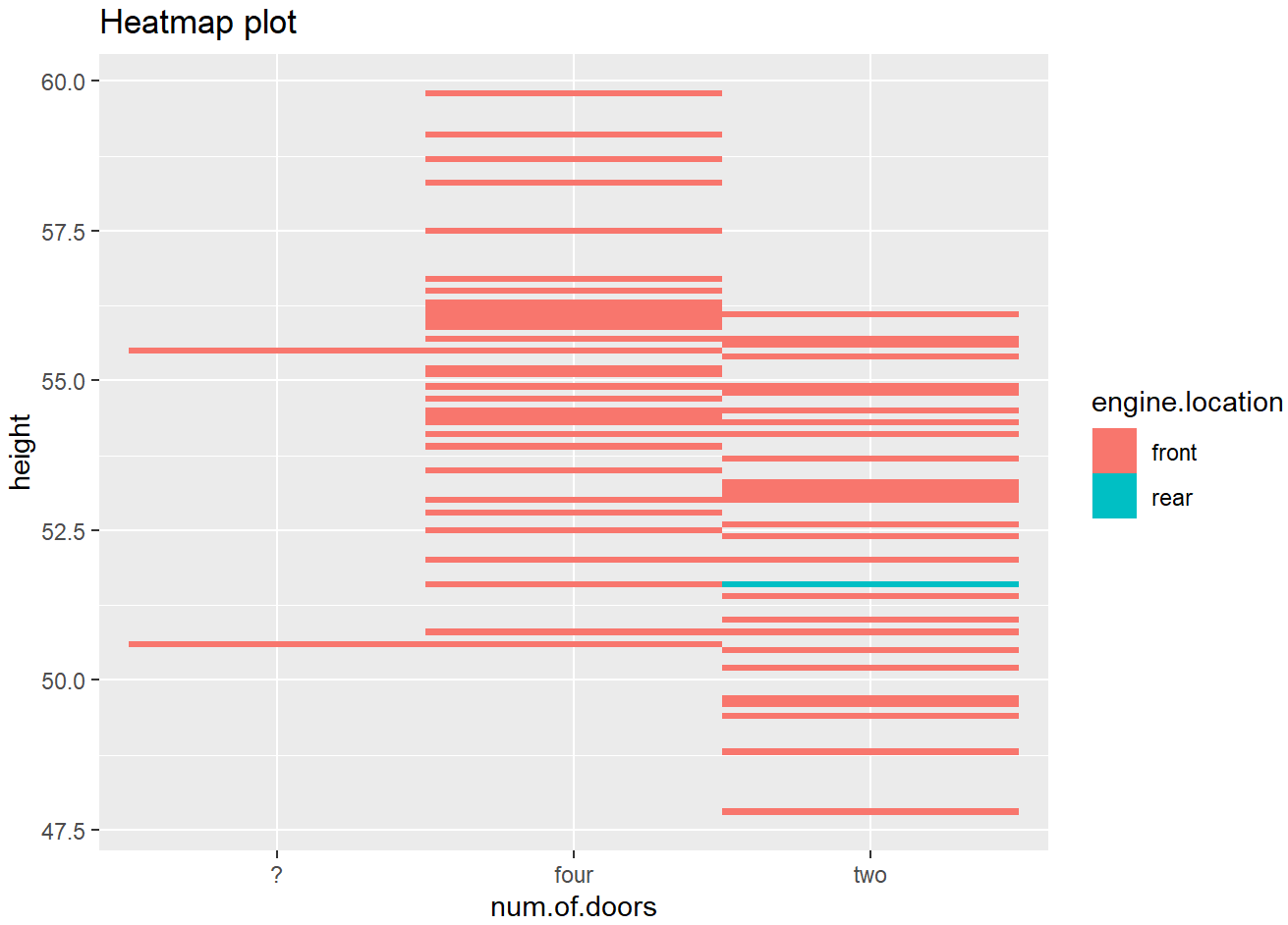

```
#4. Quantitative vs. Quantitative
#4.1. Display a Scatter plot (ggplot, geom_point())
ggplot(data_clean,aes(x=num.of.doors,y=height))+geom_point()+labs(x="num.of.doors",y="heigh
t",title="Scatter plot")
```



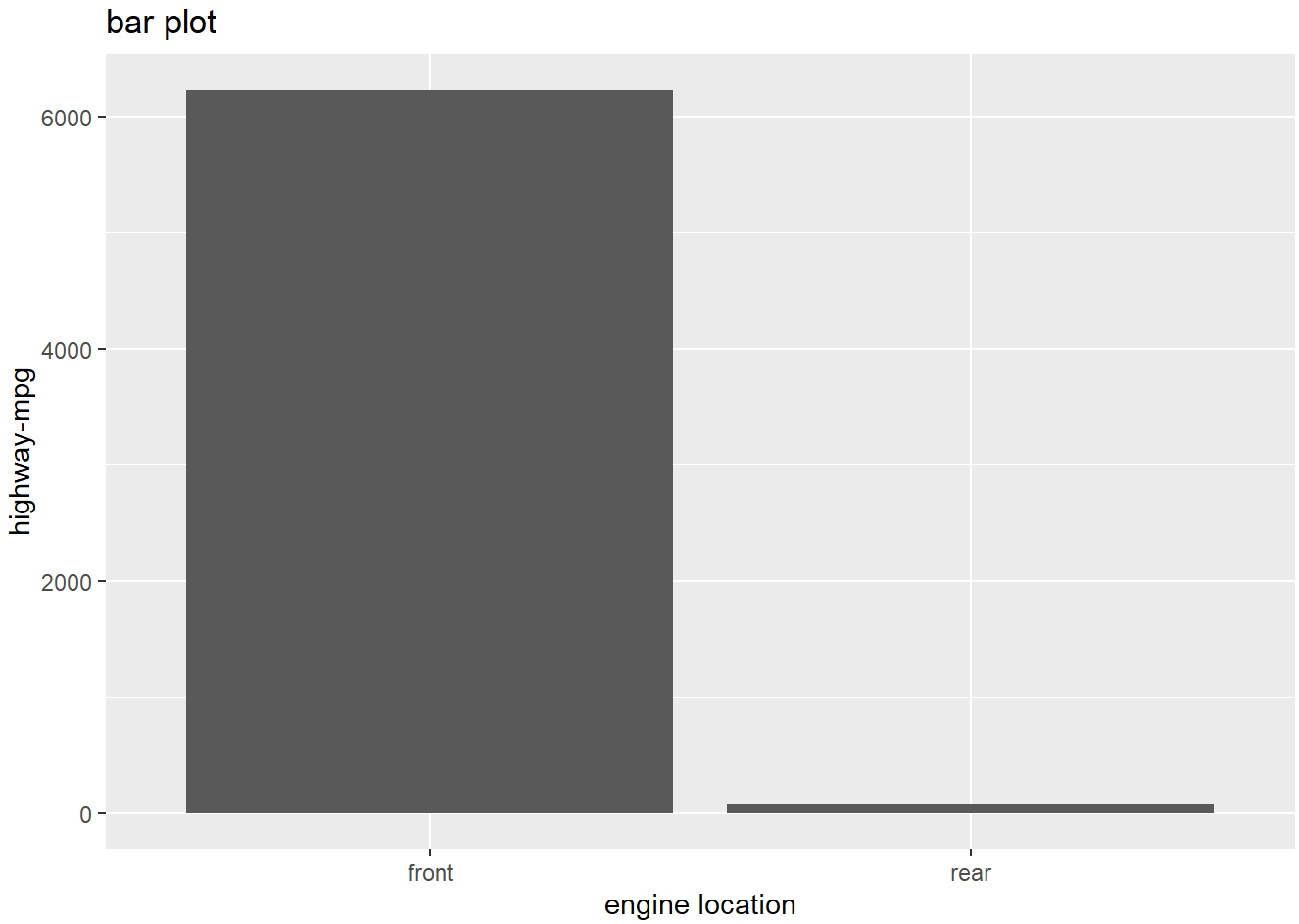

```
#4.2. Display a Line plot (ggplot, geom_line())
ggplot(data_clean,aes(x=num.of.doors,y=height))+geom_line()+labs(x="num.of.doors",y="heigh
t",title="line plot")
```



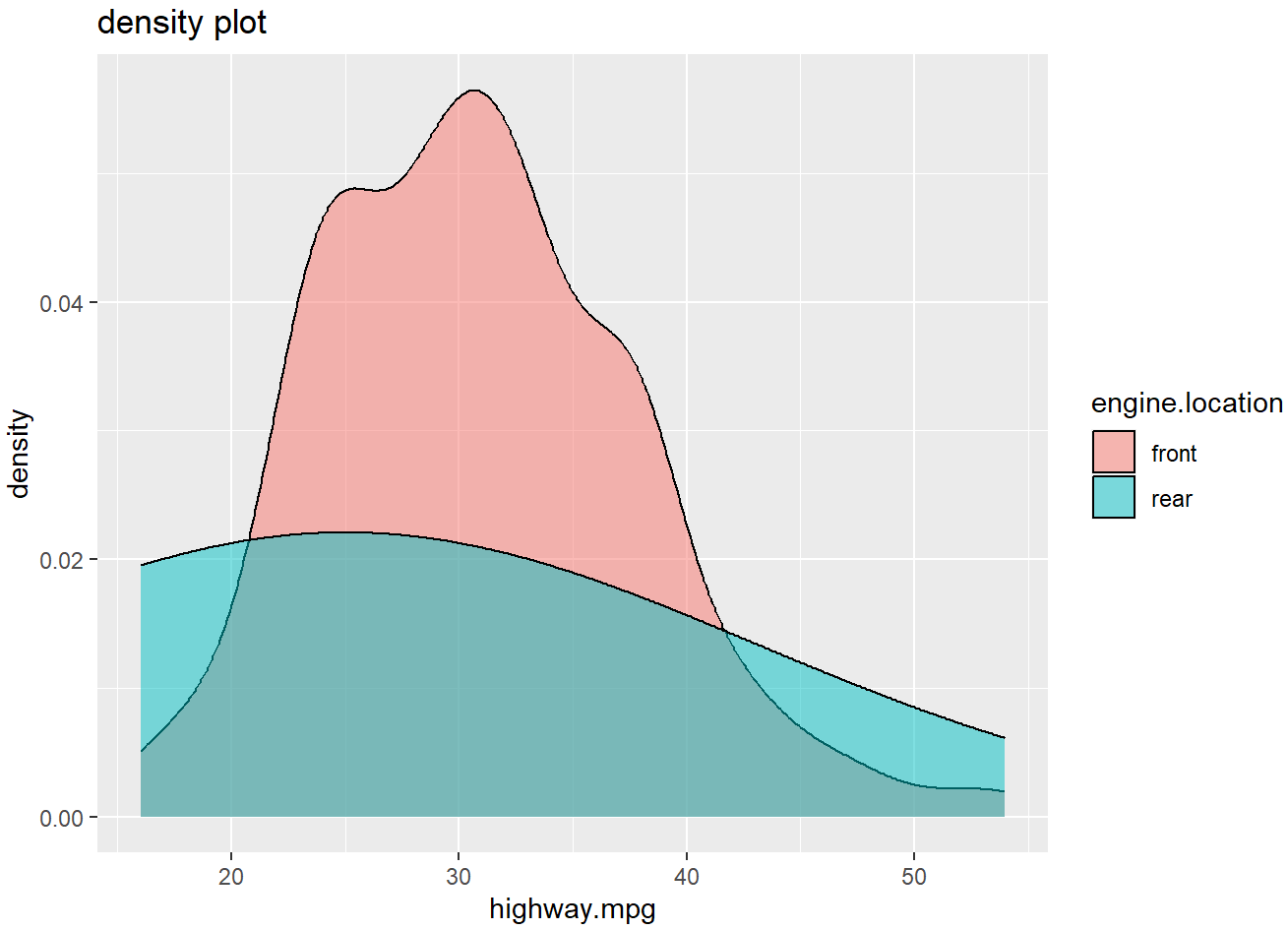

```
#4.3. Display a heatmap
ggplot(data_clean,aes(x=num.of.doors,y=height,fill=engine.location))+geom_tile()+labs(x="nu
m.of.doors",y="height",title="Heatmap plot")
```



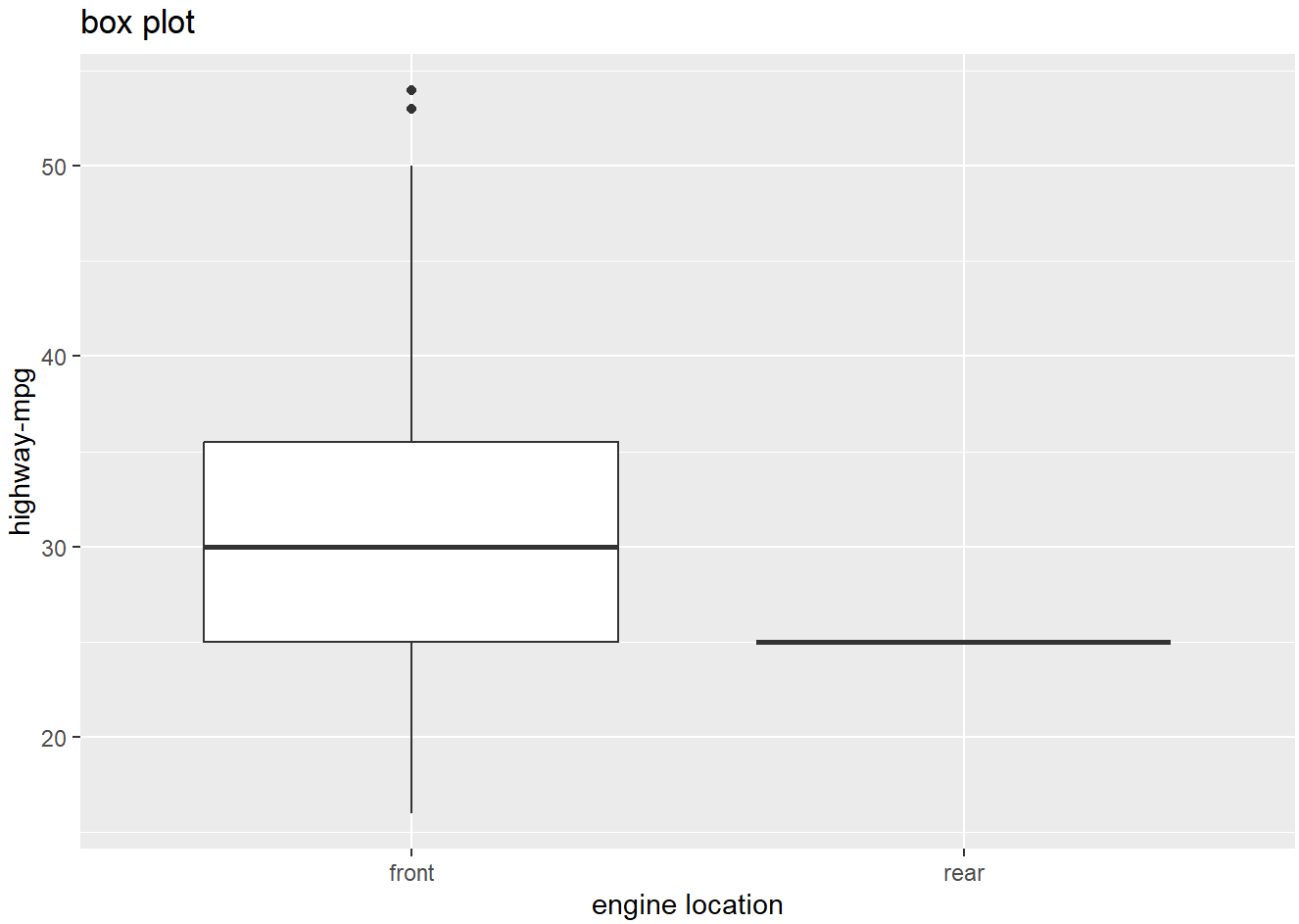

```
#5. Categorical vs. Quantitative
#5.1. Display bar chart. (ggplot, geom_bar())
ggplot(data_clean,aes(x=engine.location,y=highway.mpg))+geom_bar(stat = "identity")+labs(x="e
ngine location",y="highway-mpg",title="bar plot")
```



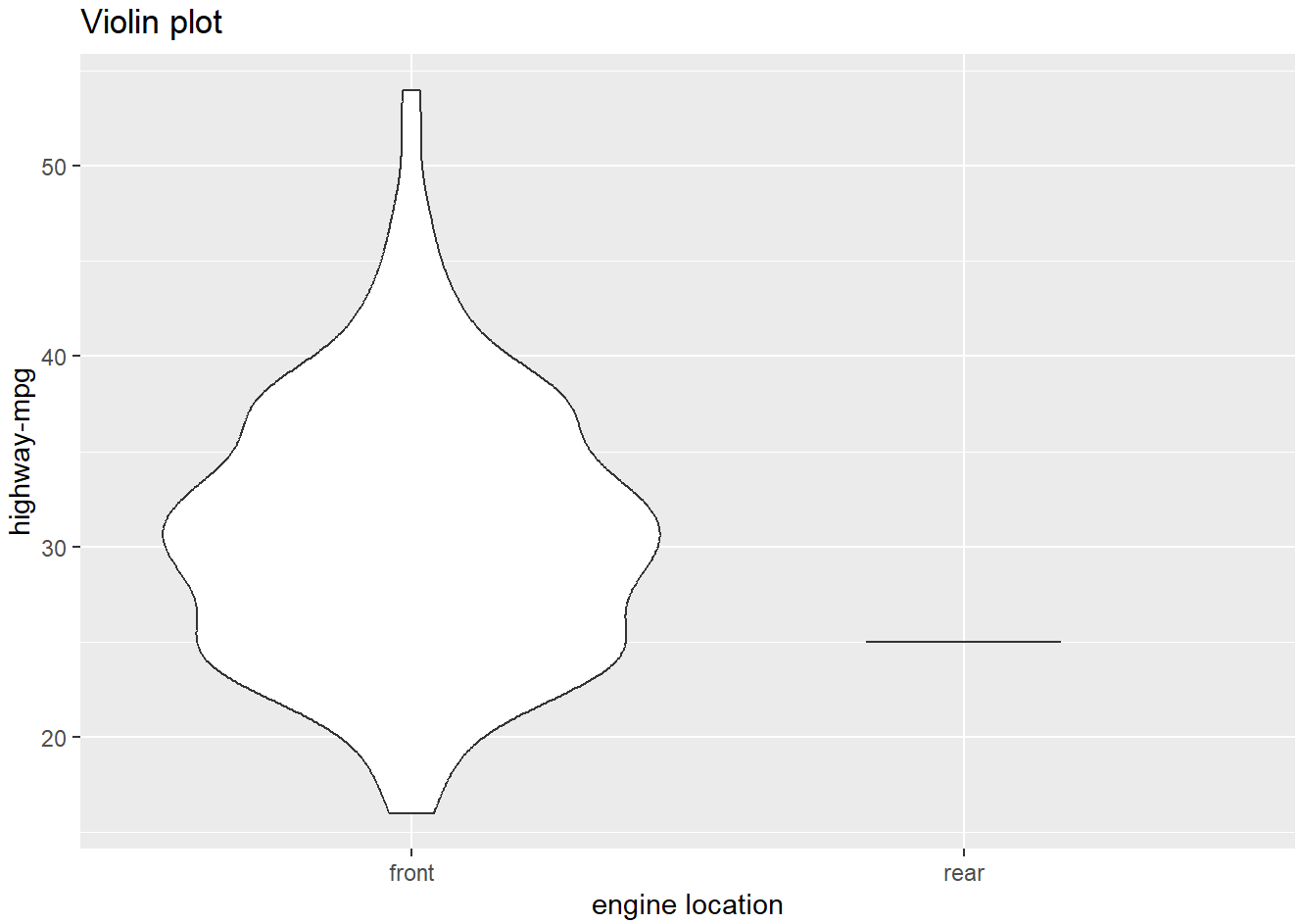

```
#5.2. Display density plot. (ggplot use fill, geom_density())
ggplot(data_clean,aes(x=highway.mpg,fill=engine.location))+geom_density(alpha=0.5)+labs(x="hi
ghway.mpg",title="density plot")
```



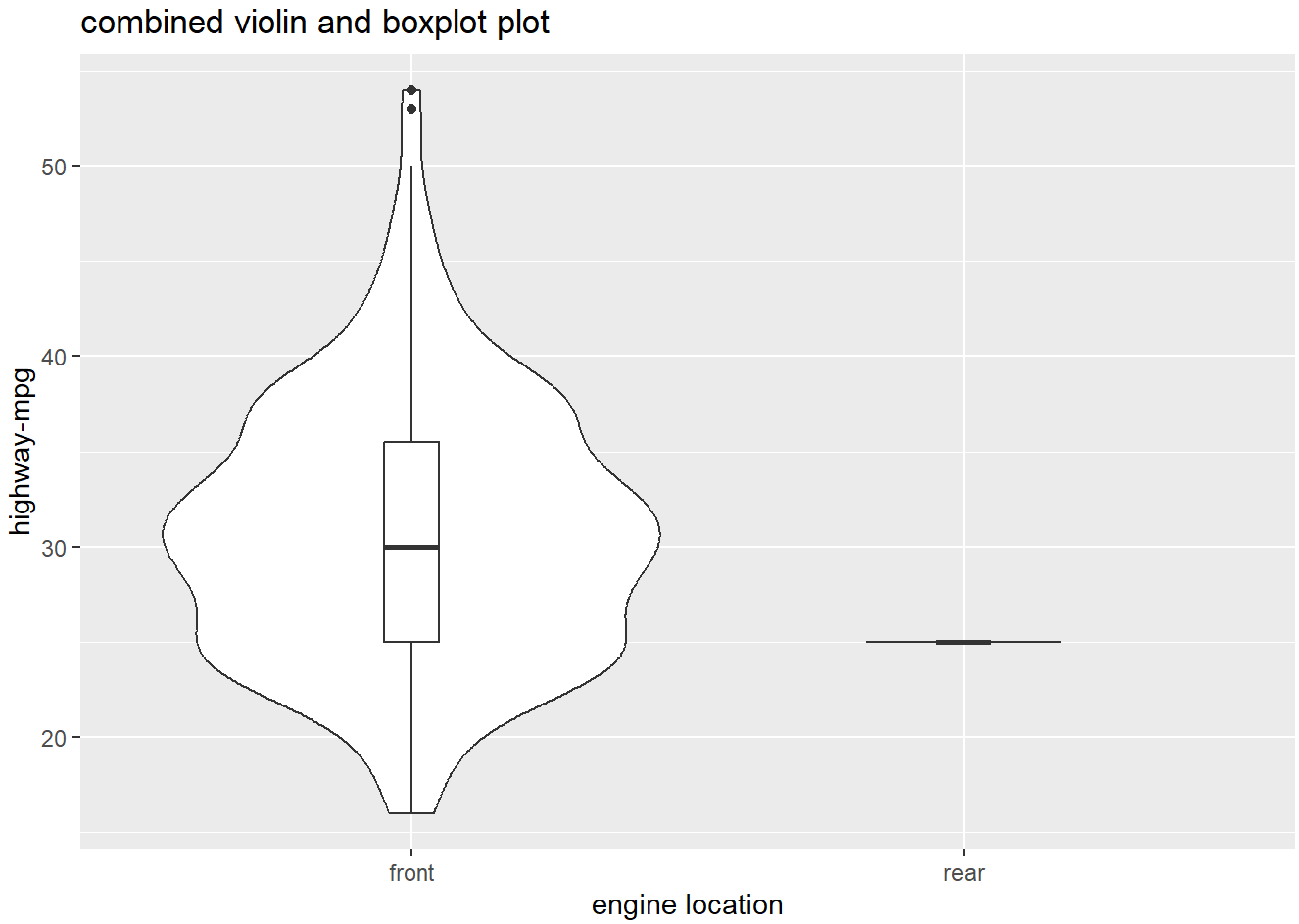

```
#5.3. Display box plot. (ggplot, geom_boxplot())
ggplot(data_clean,aes(x=engine.location,y=highway.mpg))+geom_boxplot()+labs(x="engine locatio
n",y="highway-mpg",title="box plot")
```



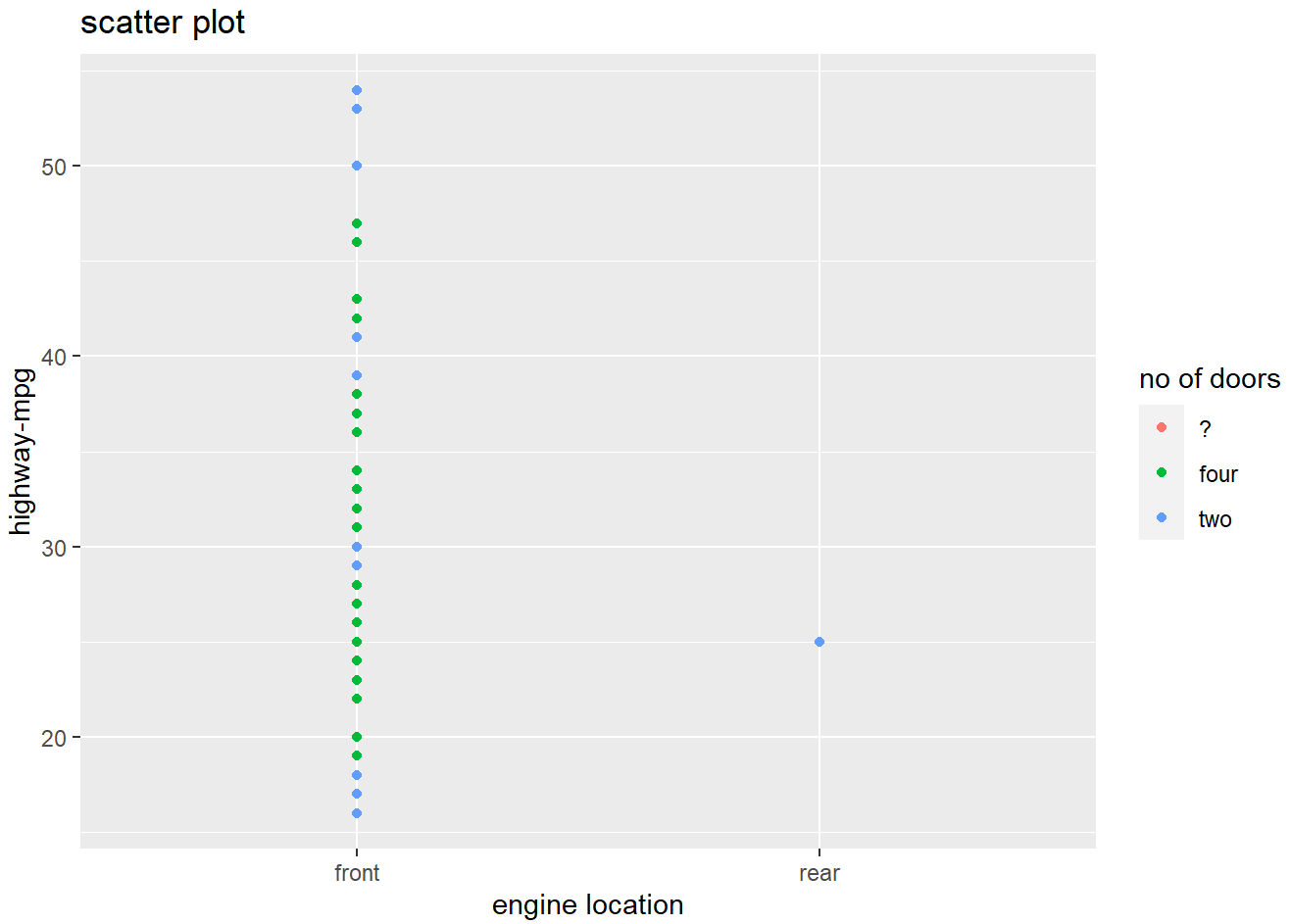

```
#5.4. Display violin plot. (ggplot, geom_violin())
ggplot(data_clean,aes(x=engine.location,y=highway.mpg))+geom_violin()+labs(x="engine locatio
n",y="highway-mpg",title="Violin plot")
```




```
#5.5. Plot the distribution using violin and boxplots (ggplot, geom_violin, geom_boxplot)#
ggplot(data_clean,aes(x=engine.location,y=highway.mpg))+geom_violin()+geom_boxplot(width =
0.1)+labs(x="engine location",y="highway-mpg",title="combined violin and boxplot plot")
```



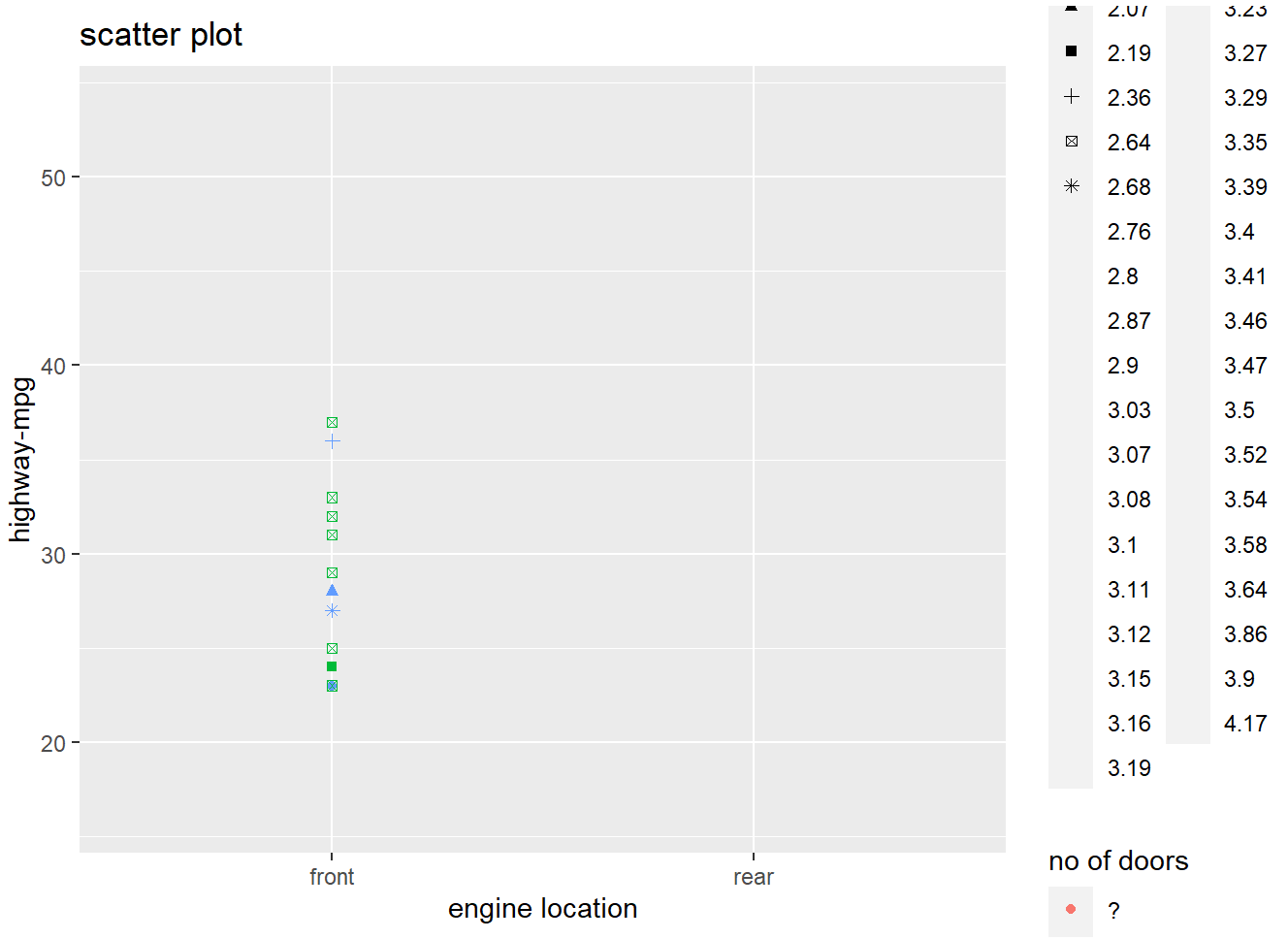

```
#Multivariate Analysis
#1. Display a Scatter plot. (ggplot use color as third variable, geom_point())
ggplot(data_clean,aes(x=engine.location,y=highway.mpg,color=num.of.doors))+geom_point()+labs
(x="engine location",y="highway-mpg",color="no of doors",title="scatter plot")
```




```
#2. Display a Scatter plot. (ggplot use color as third variable, shape as fourth variable, ge
om_point())
ggplot(data_clean,aes(x=engine.location,y=highway.mpg,color=num.of.doors,shape=stroke))+geom_
point()+labs(x="engine location",y="highway-mpg",color="no of doors",shape="stroke",title="sc
atter plot")
```

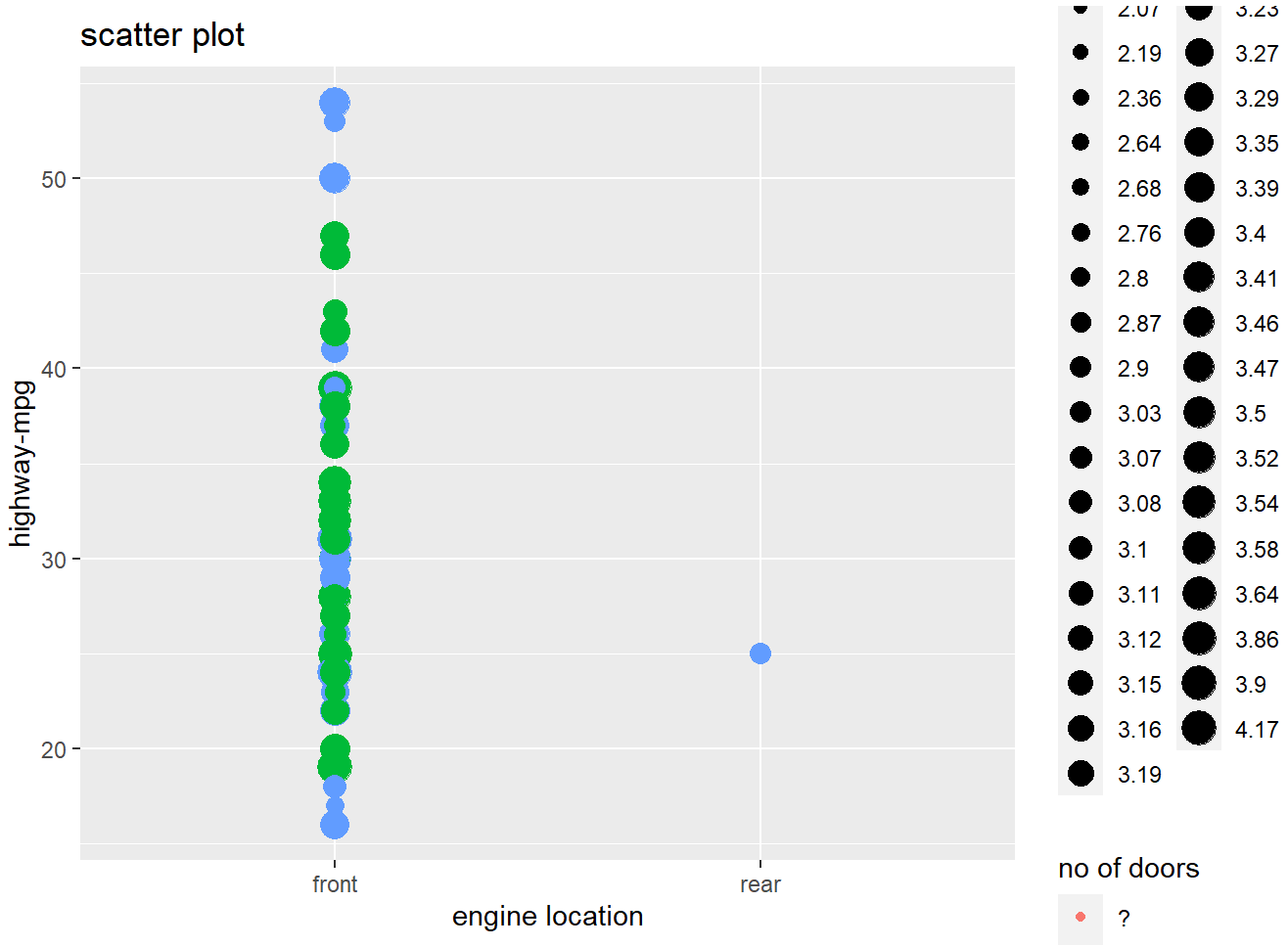
```
## Warning: The shape palette can deal with a maximum of 6 discrete values because more
## than 6 becomes difficult to discriminate
## ℹ you have requested 37 values. Consider specifying shapes manually if you need
##   that many have them.
```

```
## Warning: Removed 184 rows containing missing values (`geom_point()`).
```

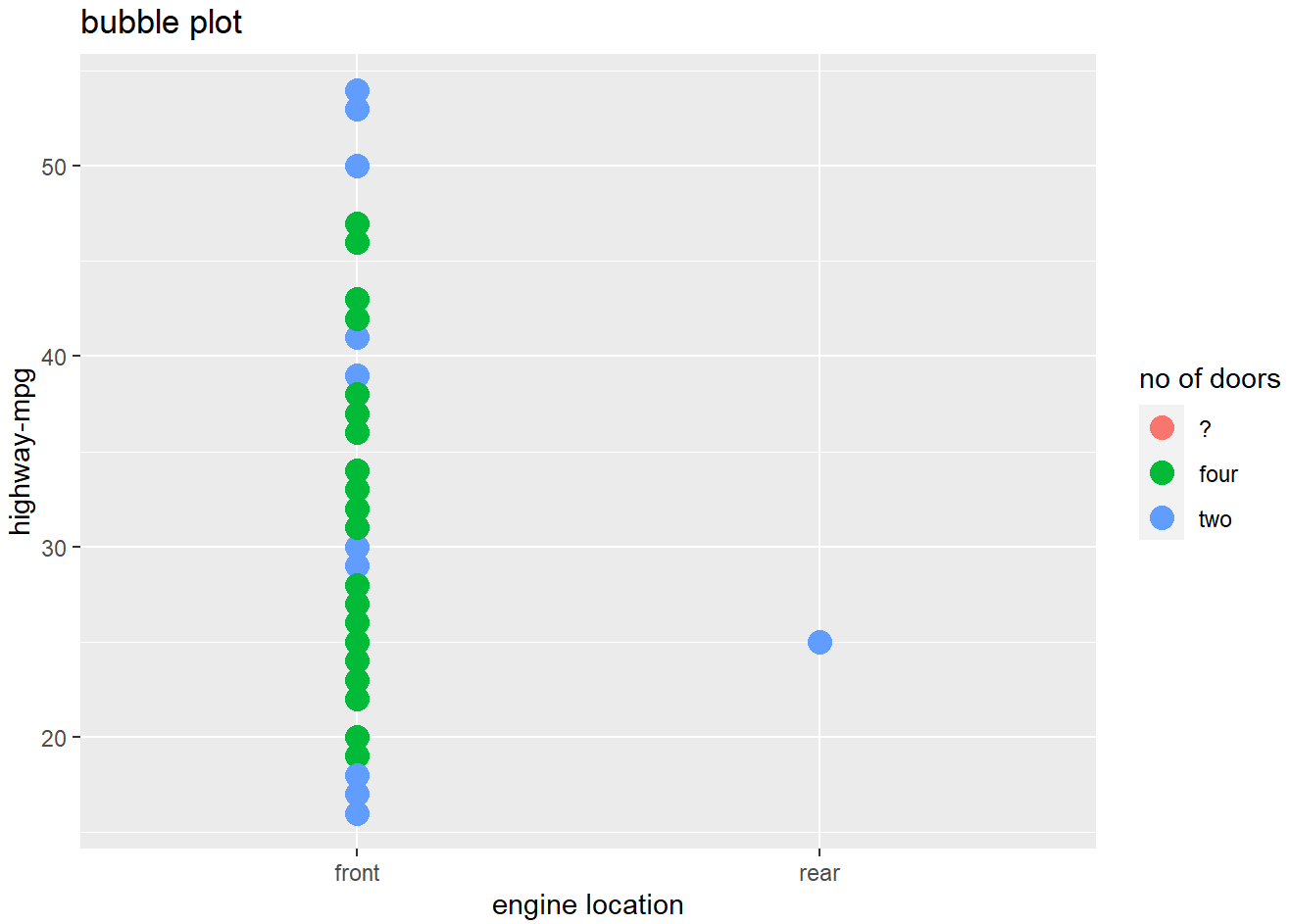



```
#3. Display a Scatter plot. (ggplot use color as third variable, size as fourth variable, geo
m_point())
ggplot(data_clean,aes(x=engine.location,y=highway.mpg,color=num.of.doors,size=stroke))+geom_p
oint()+labs(x="engine location",y="highway-mpg",color="no of doors",size="stroke",title="scat
ter plot")
```

```
## Warning: Using size for a discrete variable is not advised.
```




```
#4. Display a bubble plot. (ggplot (use x, y , color), geom_point(size))
ggplot(data_clean,aes(x=engine.location,y=highway.mpg,color=num.of.doors,size=stroke))+geom_p
oint(size=4)+labs(x="engine location",y="highway-mpg",color="no of doors",size="stroke",title
="bubble plot")
```




```
#5. Display a graph into sub-graph (ggplot(x), geom_histogram(), facet_warp())
#ggplot(data_clean,aes(x=engine.location)+geom_histogram(binwidth=1,fill="blue",color="blac
k")+facet_grid(num.of.doors~stroke)+labs(x="engine location",y="highway-mpg",title="facegrid
plot")
#6. Display a graph into sub-graph (ggplot(x), geom_histogram(), facet_grid())
#ggplot(data_clean,aes(x=engine.location)+geom_histogram(binwidth=1,fill="blue",color="blac
k")+facet_grid(num.of.doors~stroke)+labs(x="engine location",y="highway-mpg",title="facegrid
plot")
```