HW6_IS457_8

Main topic: practice plotting techniques

```
data("mtcars")
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

Q1, use coloring for multi-level variables (ggplot)

create a scatterplot showing the relationship between horsepower(hp) and gas mileage(mpg) separately for

transmission (am) and engine type (vs). The number of cylinders in each automobile engine is represented by

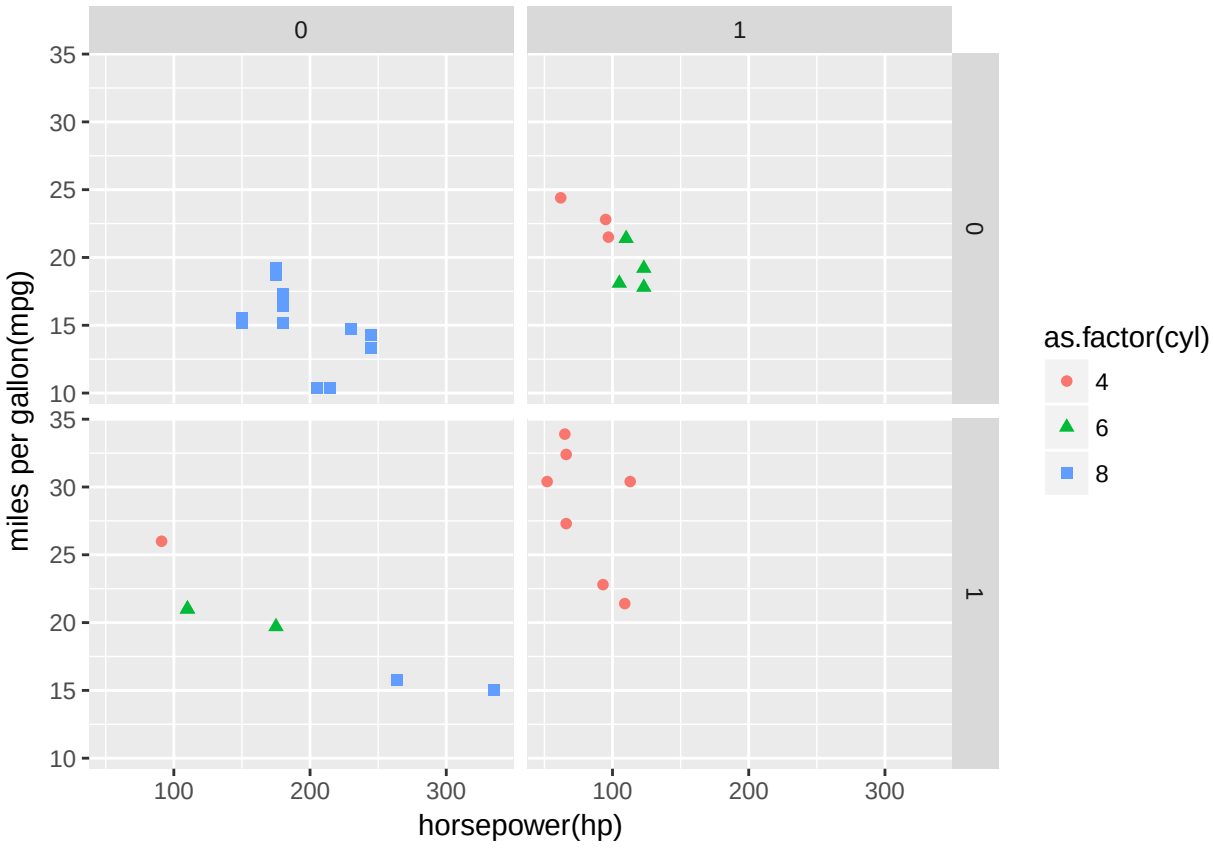
both shape(symbol) and color.

hint: your plot should consist of 4 sub-plots (facet plots);

hint: convert some variables to factor variables before plotting.

your code here

```
qplot(hp, mpg, data=mtcars, shape=as.factor(cyl), color=as.factor(cyl),
facets=am~vs,size=I(1.7),xlab="horsepower(hp)", ylab="miles per gallon(mpg)")
```


0
1
0
1
35
30
25
20
15
10
100
200
300
miles per gallon(mpg)
horsepower(hp)
as.factor(cyl)
4
6
8

your description here

top x margin :engine type(vs), y margin : transmission(am)

I observe that cars with vs = 0, am = 0, consist of only 8 cylinders and also seem to have lower miles per gallon.

cars with vs=1,am=0 do not have 8 cylinders and they seem to have higher miles per gallon.

cars with vs=0,am=1,there is a negative correlation between mpg and hp, cars with higher horsepower have lower miles per gallon

and higher number of cylinders

cars with vs=1,am=1, only have 4 cylinders and have low horsepower but have higher miles per gallon

Q2, display multivariate relationships: look up the manual page of package lattice

make a boxplot to display a variable, mpg, conditioned on different values of cylinders.

hint: mpg grouped by factors representing cylinders, your plot should consist of

(num. of distinct values of cyl) smaller boxplots.

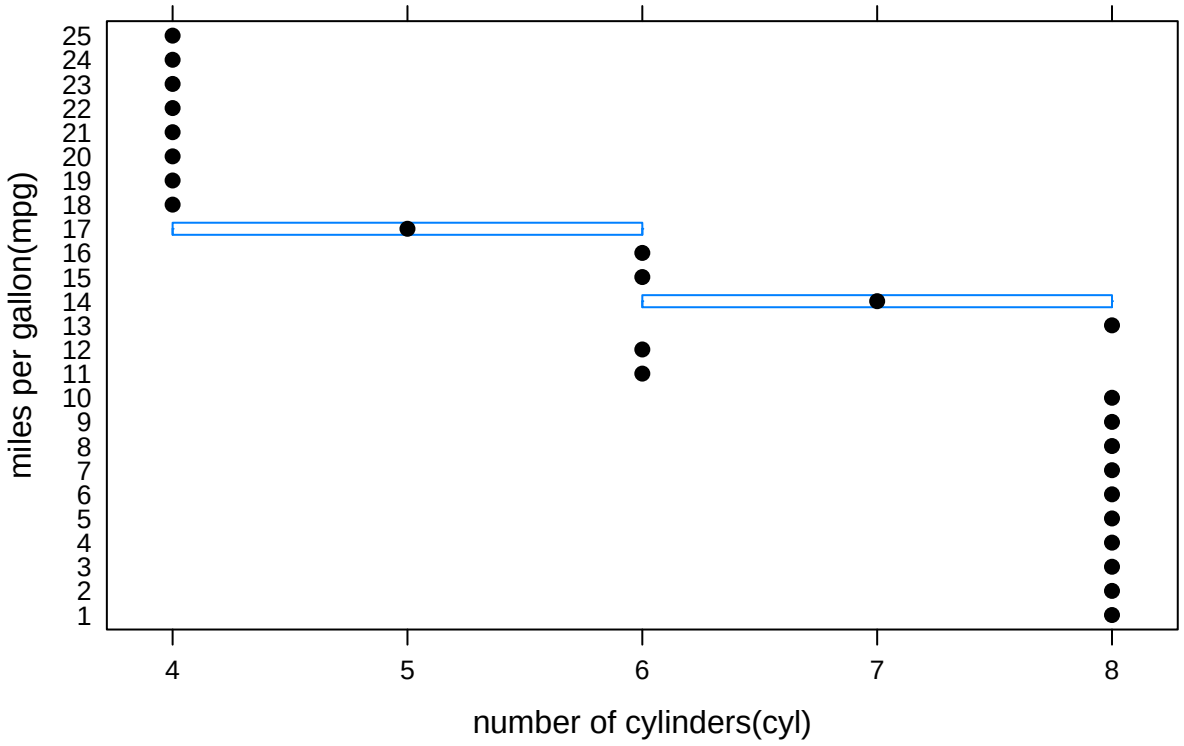
your code here

install.packages("lattice")

```
library(lattice)
```

```
## Warning: package 'lattice' was built under R version 3.4.4
```

```
bwplot(mpg ~ cyl ,data=mtcars, ylab = "miles per gallon(mpg)", main= "box plot of mtcars mpg and cyl",
xlab="number of cylinders(cyl)")
```


box plot of mtcars mpg and cyl
miles per gallon(mpg)
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
4
5
6
7
8
number of cylinders(cyl)

your description here

From The box plot we learn that for a 6 cylinder car, the maximum mpg is equal to the

minumum mpg for a 4 cylinder car.Also the maximum mpg for an 8 cylinder car is slightly less than

the median mpg for a 6 cylinder car

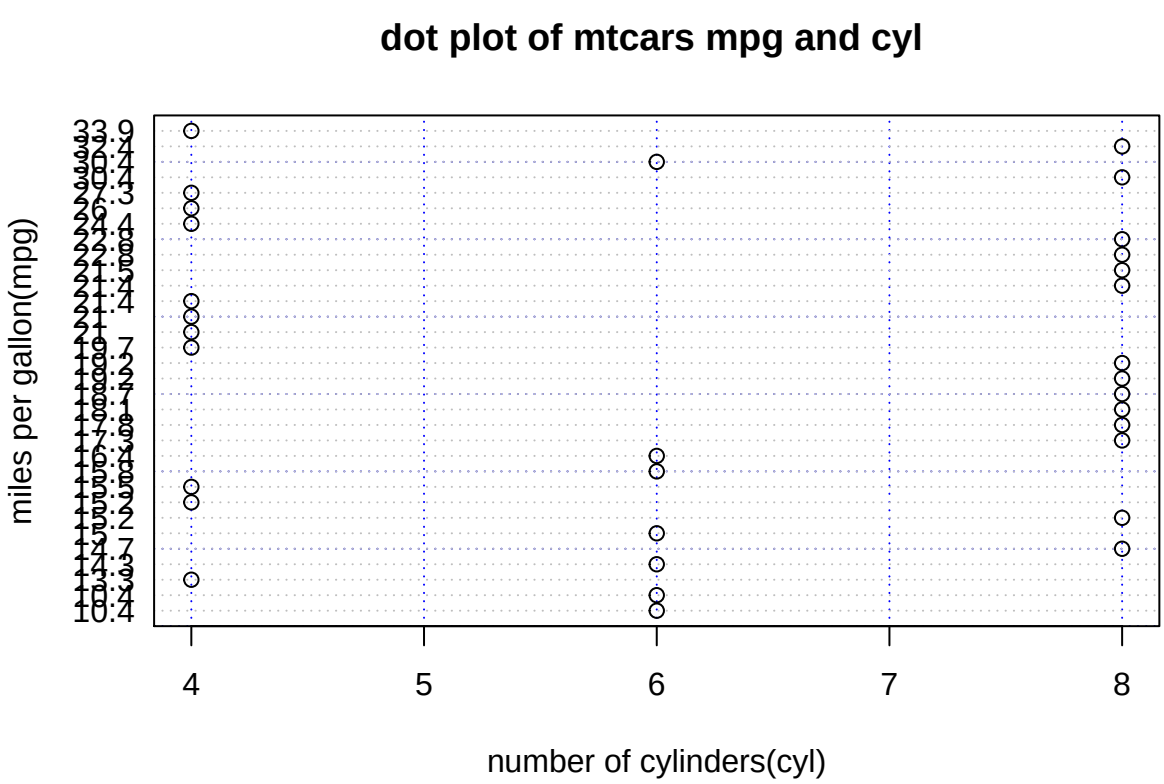
Q3, visualize a large number of labeled values using dot plots

plot the sorted mpg values with car types on the y axis, grouped by cylinders

your graph should have a grid

your code here

```
dotchart(mtcars$cyl,sort(mtcars$mpg),main="dot plot of mtcars mpg and cyl ",
xlab= "number of cylinders(cyl)", ylab="miles per gallon(mpg)",panel.first = grid(col= "blue"))
```


dot plot of mtcars mpg and cyl
miles per gallon(mpg)
4
5
6
7
8
number of cylinders(cyl)

your description here

this is a dot plot representing the sorted values of the mpg variable in the mtcars dataset

on the y axis and the cyl variable on the x axis.From this plot, we can see exactly how many observations

there are in each cylinder category and their exact mpg values

11 observations in 4 cylinders

7 observations in 6 cylinders

14 observations in 8 cylinders

Q4, comparing distributions

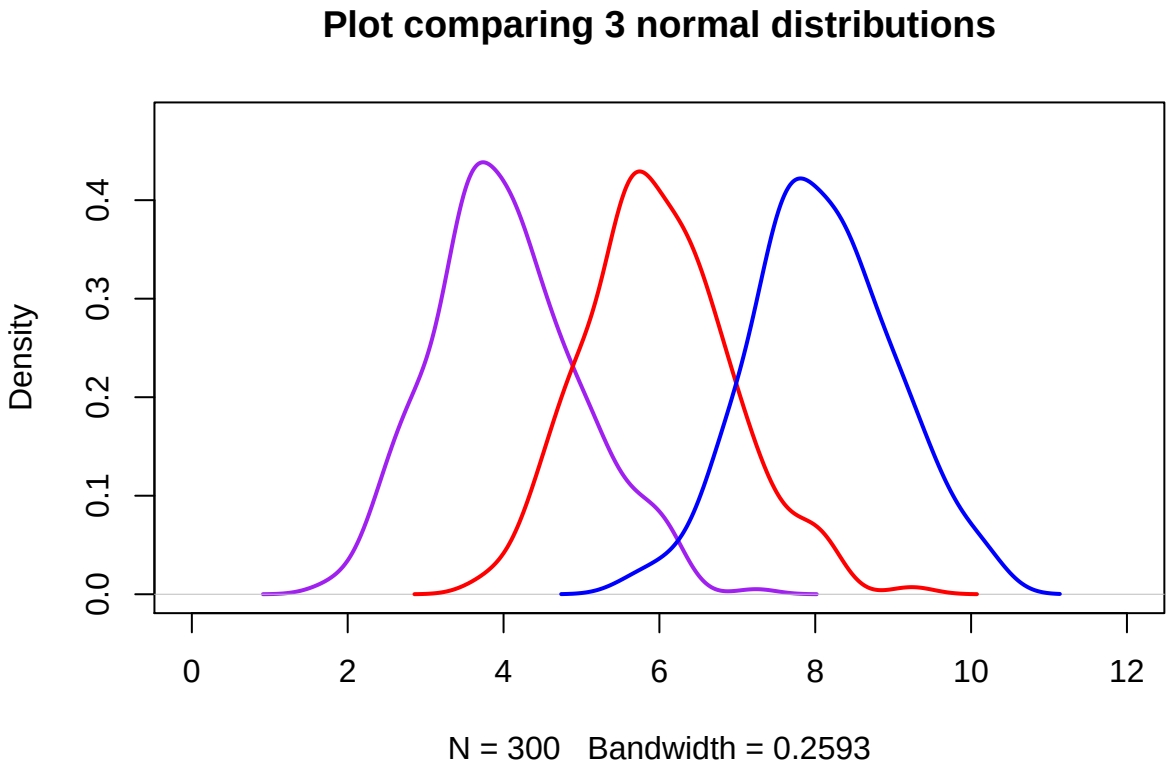
suppose we have 3 distributions as follows: rnorm(300,4), rnorm(200,6), rnorm(100, 8).

plot them on the same graph, one color each distribution, with rainbow colors.

overlapping areas should have different/darker colors.

your code here

```
set.seed(123)
plot(density(rnorm(300,4)), col="purple",xlim=c(0,12),ylim=c(0.0,0.48),lwd=2, main = "Plot comparing 3 r
set.seed(123)
lines(density(rnorm(200,6)),col="red",lwd=2)
set.seed(123)
lines(density(rnorm(100,8)),col="blue",lwd=2)
```


Plot comparing 3 normal distributions
Density
0.0
0.1
0.2
0.3
0.4
0
2
4
6
8
10
12
N = 300 Bandwidth = 0.2593

your description here

This plot contains three normal distributions each with different values for mean , n.

Q5, stack bar plots with gradient colors

use two categorical variables, cylinders and gears, to create stacked bar charts

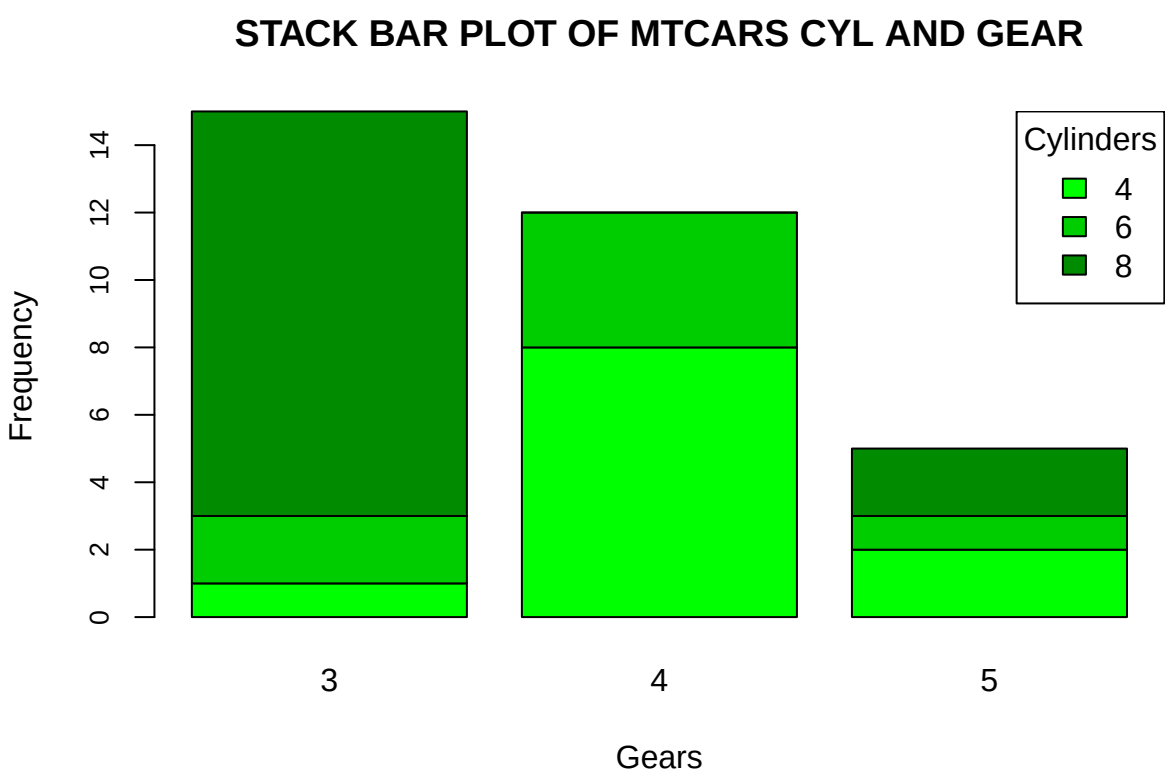
your y axis should be frequency.

use the same color but darker shades indicating higher number of cylinders.

hint: you can create a contingency table to help you plot

your code here

```
cyl_gear= table(mtcars$cyl,mtcars$gear)
barplot(cyl_gear, main="STACK BAR PLOT OF MTCARS CYL AND GEAR", xlab="Gears",ylab="Frequency",cex.axis=
col =c("green","green3","green4"))
legend("topright", title="Cylinders",legend= sort(unique(mtcars$cyl)), fill =c("green","green3","green4
```



your description here

this is a stack bar plot with gradient colors representing two categorical variables ,cylinders and gears.

it shows the frequency of each gear type and the cylinder distribution within each gear type

3 gear cars have a high frequency of 8 cylinders and smaller frequency of 6,4 cylinders

4 gear cars have a high frequency of 4 cyclinders,smaller frequency of 6 cylinders and no 8 cylinders

5 gear cars have 4,6 and 8 cylinders