L2 prep M2

Learning Spoons

2018년 4월 17일

Module 2: ggplot2

library(ggplot2)  
? mpg

## starting httpd help server ... done

mpg

## # A tibble: 234 x 11  
## manufacturer model displ year cyl trans drv cty hwy fl   
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr>  
## 1 audi a4 1.8 1999 4 auto(l~ f 18 29 p   
## 2 audi a4 1.8 1999 4 manual~ f 21 29 p   
## 3 audi a4 2 2008 4 manual~ f 20 31 p   
## 4 audi a4 2 2008 4 auto(a~ f 21 30 p   
## 5 audi a4 2.8 1999 6 auto(l~ f 16 26 p   
## 6 audi a4 2.8 1999 6 manual~ f 18 26 p   
## 7 audi a4 3.1 2008 6 auto(a~ f 18 27 p   
## 8 audi a4 quat~ 1.8 1999 4 manual~ 4 18 26 p   
## 9 audi a4 quat~ 1.8 1999 4 auto(l~ 4 16 25 p   
## 10 audi a4 quat~ 2 2008 4 manual~ 4 20 28 p   
## # ... with 224 more rows, and 1 more variable: class <chr>

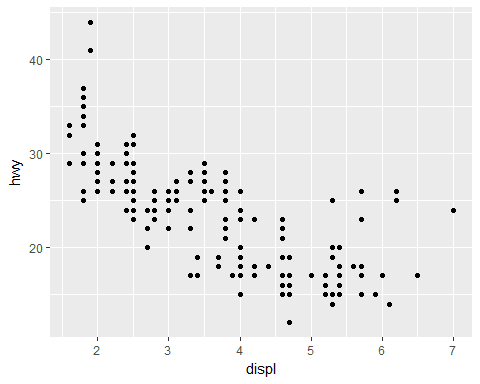
class(mpg)

## [1] "tbl\_df" "tbl" "data.frame"

# mpg has attributes for multiple classes!  
# tbl is from package "tibble" and can be handy.

bigger engine less mileage?

# engine size  
ggplot(data=mpg) +  
 geom\_point(mapping = aes(x = displ, y = hwy))



[L2.M2.Fig 1]

# template  
ggplot(data = <DATA>) +  
 <GEOM\_FUNCTION>(mapping = aes(<MAPPING>))

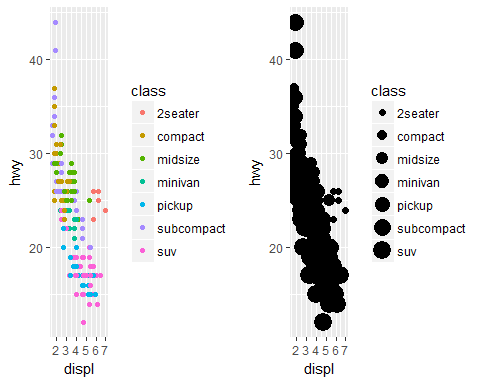
Your question confirmed?

Additional questions?

What are the six points?  
Same tendency among the different classes?  
compact, midsize, suv, 2seater, minivan, pickup, subcompact

library(gridExtra)  
a <- ggplot(data = mpg) +  
 geom\_point(mapping = aes(x = displ, y = hwy, color = class))  
b <- ggplot(data = mpg) +  
 geom\_point(mapping = aes(x = displ, y = hwy, size = class))  
grid.arrange(a, b, nrow=1, ncol=2)

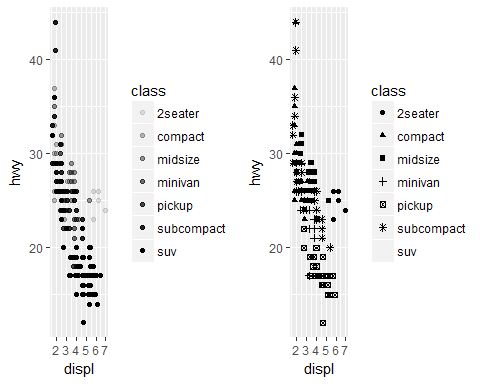
## Warning: Using size for a discrete variable is not advised.



a <- ggplot(data = mpg) +  
 geom\_point(mapping = aes(x = displ, y = hwy, alpha = class))  
b <- ggplot(data = mpg) +  
 geom\_point(mapping = aes(x = displ, y = hwy, shape = class))  
grid.arrange(a, b, nrow=1, ncol=2)

## Warning: The shape palette can deal with a maximum of 6 discrete values  
## because more than 6 becomes difficult to discriminate; you have 7.  
## Consider specifying shapes manually if you must have them.

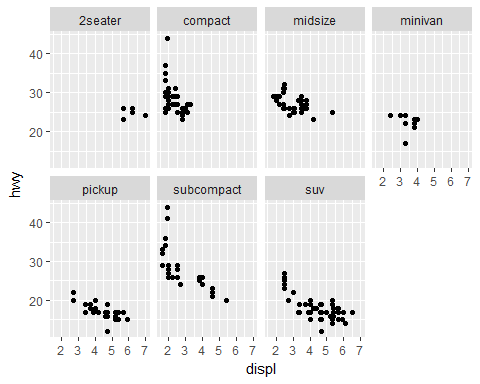
## Warning: Removed 62 rows containing missing values (geom\_point).



discussion on size > alpha > color = shape

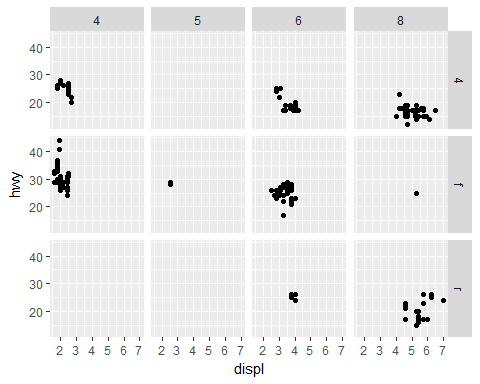
facets

ggplot(data = mpg) +   
 geom\_point(mapping = aes(x = displ, y = hwy)) +  
 facet\_wrap(~ class, nrow = 2)



[L2.M2.Fig 2]

ggplot(data = mpg) +   
 geom\_point(mapping = aes(x = displ, y = hwy)) +  
 facet\_grid(drv ~ cyl)



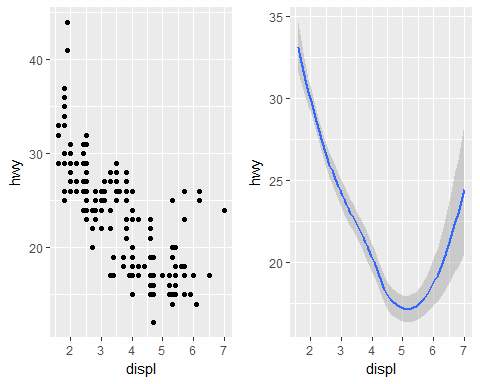
[L2.M2.Fig 3]

discussion: facet or not?

## 

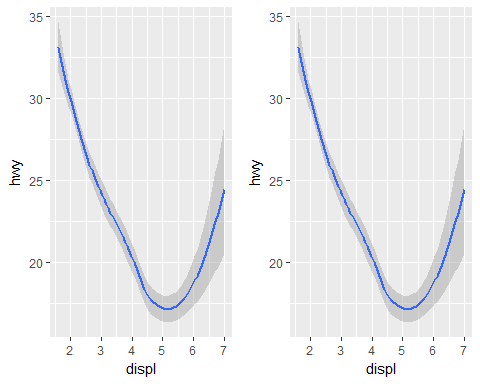
# point vs smooth  
a <- ggplot(data = mpg) +  
 geom\_point(mapping = aes(x = displ, y = hwy))  
b <- ggplot(data = mpg) +  
 geom\_smooth(mapping = aes(x = displ, y = hwy))  
grid.arrange(a, b, nrow=1, ncol=2)

## `geom\_smooth()` using method = 'loess'



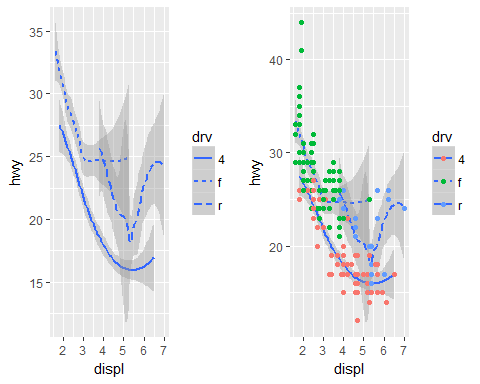
# aes extends or overwrites  
a <- ggplot(data = mpg) +  
 geom\_smooth(mapping = aes(x = displ, y = hwy))  
b <- ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +  
 geom\_smooth()  
grid.arrange(a, b, nrow=1, ncol=2)

## `geom\_smooth()` using method = 'loess'  
## `geom\_smooth()` using method = 'loess'



# point + smooth  
a <- ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +  
 geom\_smooth(aes(linetype = drv))  
b <- ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +  
 geom\_smooth(aes(linetype = drv)) +  
 geom\_point(aes(color = drv)) # line type ignored  
grid.arrange(a, b, nrow=1, ncol=2)

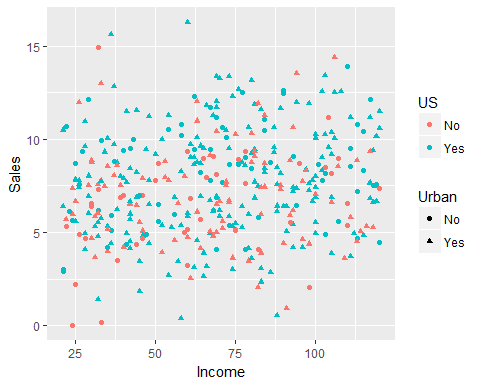
## `geom\_smooth()` using method = 'loess'  
## `geom\_smooth()` using method = 'loess'



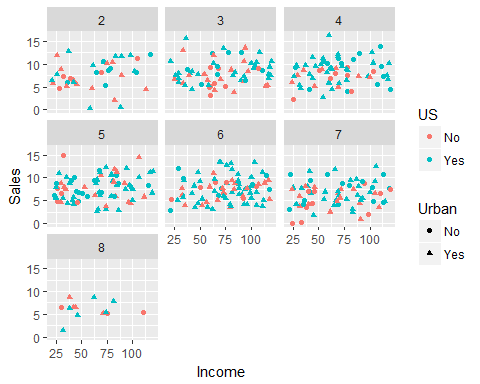
library(ggplot2)  
library(ISLR)   
str(Carseats)

## 'data.frame': 400 obs. of 11 variables:  
## $ Sales : num 9.5 11.22 10.06 7.4 4.15 ...  
## $ CompPrice : num 138 111 113 117 141 124 115 136 132 132 ...  
## $ Income : num 73 48 35 100 64 113 105 81 110 113 ...  
## $ Advertising: num 11 16 10 4 3 13 0 15 0 0 ...  
## $ Population : num 276 260 269 466 340 501 45 425 108 131 ...  
## $ Price : num 120 83 80 97 128 72 108 120 124 124 ...  
## $ ShelveLoc : Factor w/ 3 levels "Bad","Good","Medium": 1 2 3 3 1 1 3 2 3 3 ...  
## $ Age : num 42 65 59 55 38 78 71 67 76 76 ...  
## $ Education : num 17 10 12 14 13 16 15 10 10 17 ...  
## $ Urban : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 1 2 2 1 1 ...  
## $ US : Factor w/ 2 levels "No","Yes": 2 2 2 2 1 2 1 2 1 2 ...

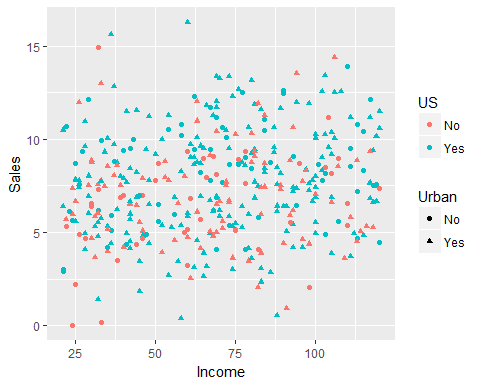
a <- ggplot(data = Carseats, aes(x = Income, y = Sales)) +   
 geom\_point(aes(shape = Urban, color = US))  
print(a)



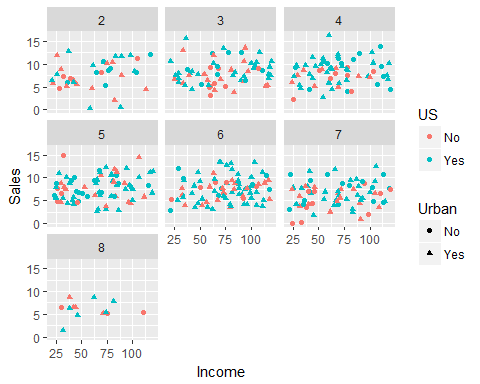
a + facet\_wrap(~ floor(Age/10))



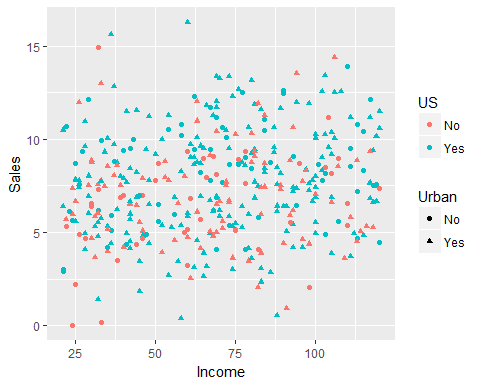
print(a)



doFacetWrap <- TRUE  
a <- ggplot(data = Carseats, aes(x = Income, y = Sales)) +   
 geom\_point(aes(shape = Urban, color = US))  
if (doFacetWrap) {  
 a <- a + facet\_wrap(~ floor(Age/10))  
}  
print(a)



doFacetWrap <- FALSE  
a <- ggplot(data = Carseats, aes(x = Income, y = Sales)) +   
 geom\_point(aes(shape = Urban, color = US))  
if (doFacetWrap) {  
 a <- a + facet\_wrap(~ floor(Age/10))  
}  
print(a)



# Peak - final project

appendix

# C, C++  
for(int i=1; i <=10, i=i+1) {  
 printf("value of i: %d\n", i);  
}  
  
# MATLAB  
for i=1:10 {  
 disp("value of i: %d", i)  
}  
  
# Python  
for i in range(1, 11):  
 print "value of i: %d" % (i)  
  
## R  
for (i in 1:10) {  
 print(paste("value of i:", i))  
}