1_learn_ggplot2.R

win10

2021-03-15

```
# # ############################################################################
# rm(list = ls())
# if (!is.null(dev.list()))
#   dev.off()
# options(digits = 5)
# dev.off()
# ############################################################################
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

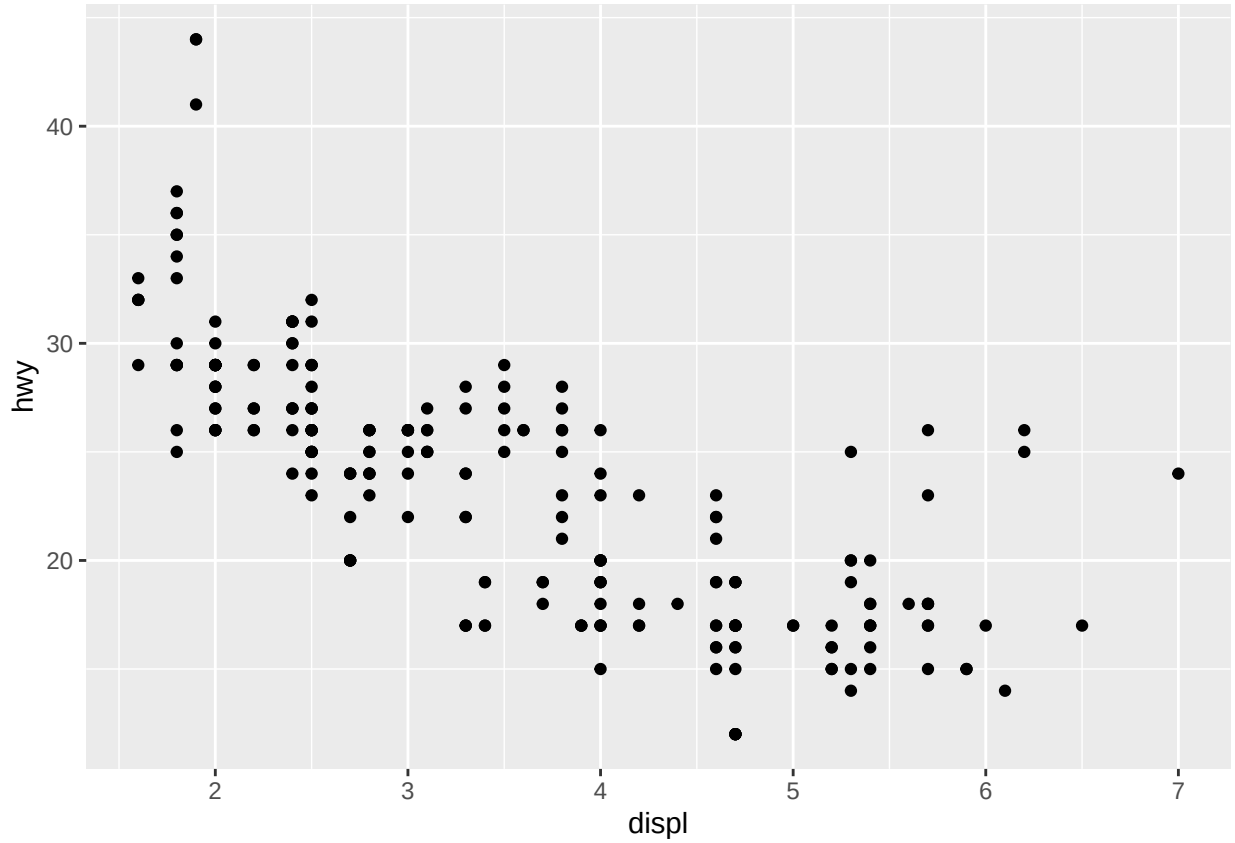
```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# ######################
mpg
```

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

```
#creates blank graph. need to add layers
ggplot(data = mpg)
```

```
#add layer with geom_point
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```




```
#get summary
summary(mpg)
```

```
##  manufacturer          model               displ           year
##  Length:234         Length:234         Min.   :1.600   Min.   :1999
##  Class :character   Class :character   1st Qu.:2.400   1st Qu.:1999
##  Mode  :character   Mode  :character   Median :3.300   Median :2004
##                                        Mean   :3.472   Mean   :2004
##                                        3rd Qu.:4.600   3rd Qu.:2008
##                                        Max.   :7.000   Max.   :2008
##       cyl           trans               drv                 cty
##  Min.   :4.000   Length:234         Length:234         Min.   : 9.00
##  1st Qu.:4.000   Class :character   Class :character   1st Qu.:14.00
##  Median :6.000   Mode  :character   Mode  :character   Median :17.00
##  Mean   :5.889                                         Mean   :16.86
##  3rd Qu.:8.000                                         3rd Qu.:19.00
##  Max.   :8.000                                         Max.   :35.00
##       hwy             fl                class
##  Min.   :12.00   Length:234         Length:234
##  1st Qu.:18.00   Class :character   Class :character
##  Median :24.00   Mode  :character   Mode  :character
##  Mean   :23.44
##  3rd Qu.:27.00
##  Max.   :44.00
```

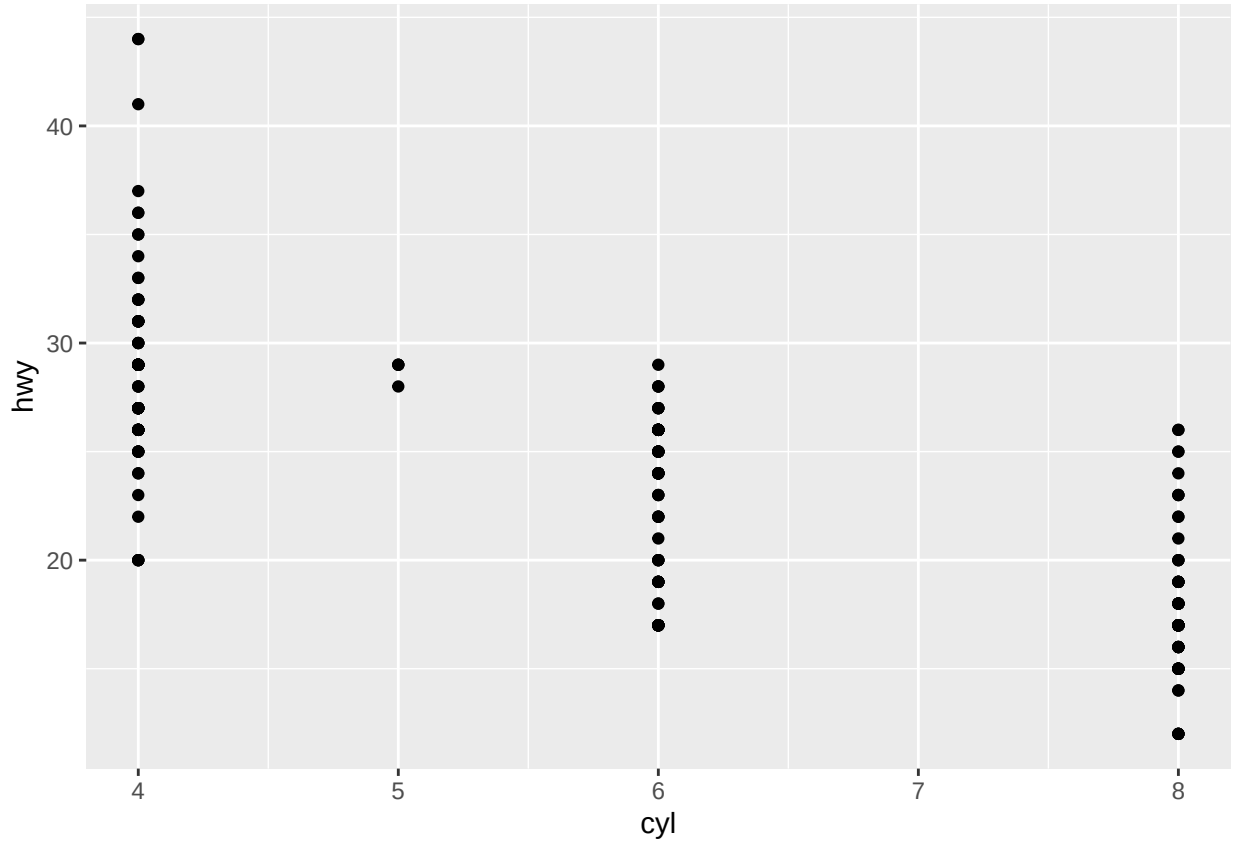
```
#get structure
str(mpg)
```

```
## tibble [234 x 11] (S3: tbl_df/tbl/data.frame)
##  $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
##  $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
##  $ displ       : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
##  $ year        : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
##  $ cyl         : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
##  $ trans       : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
##  $ drv         : chr [1:234] "f" "f" "f" "f" ...
##  $ cty         : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
##  $ hwy         : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
##  $ fl          : chr [1:234] "p" "p" "p" "p" ...
##  $ class       : chr [1:234] "compact" "compact" "compact" "compact" ...
```

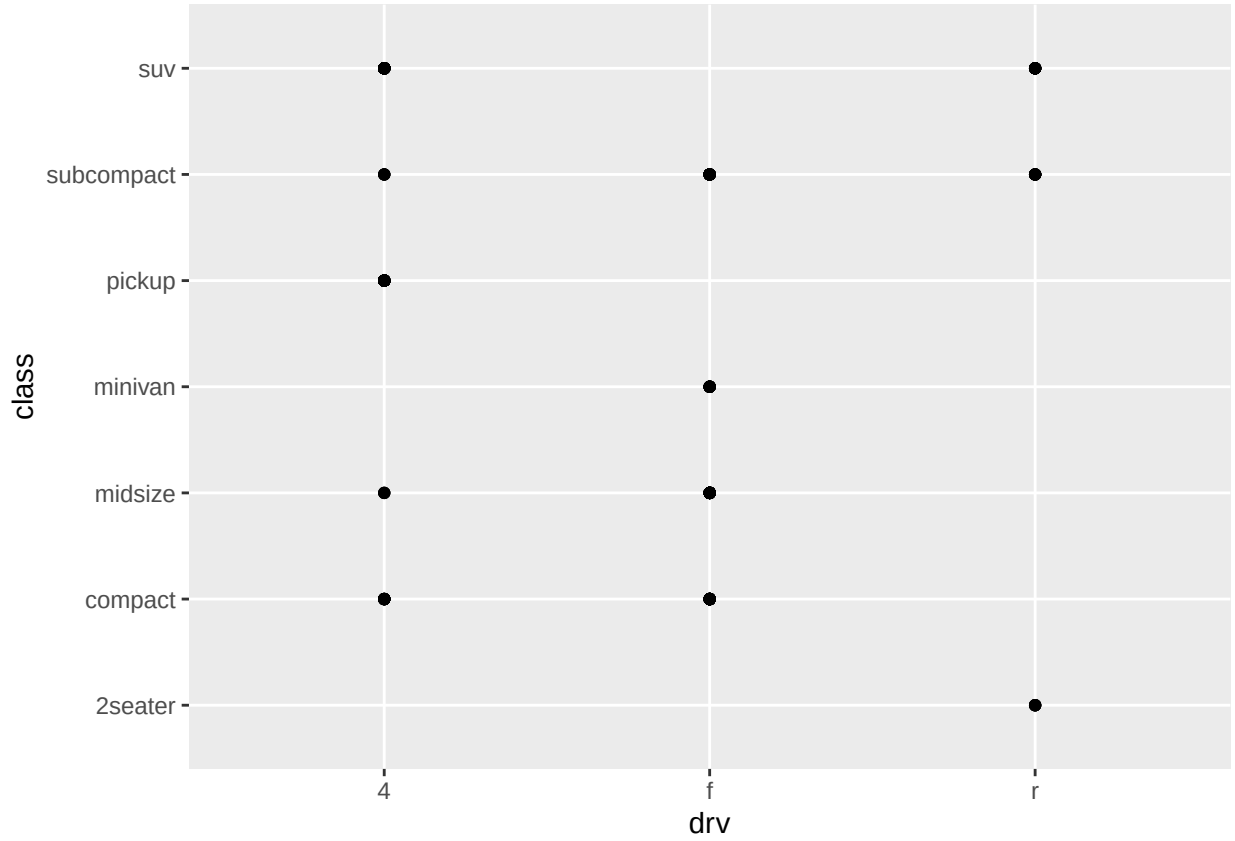
```
#get dimensions
dim(mpg)
```

```
## [1] 234  11
```

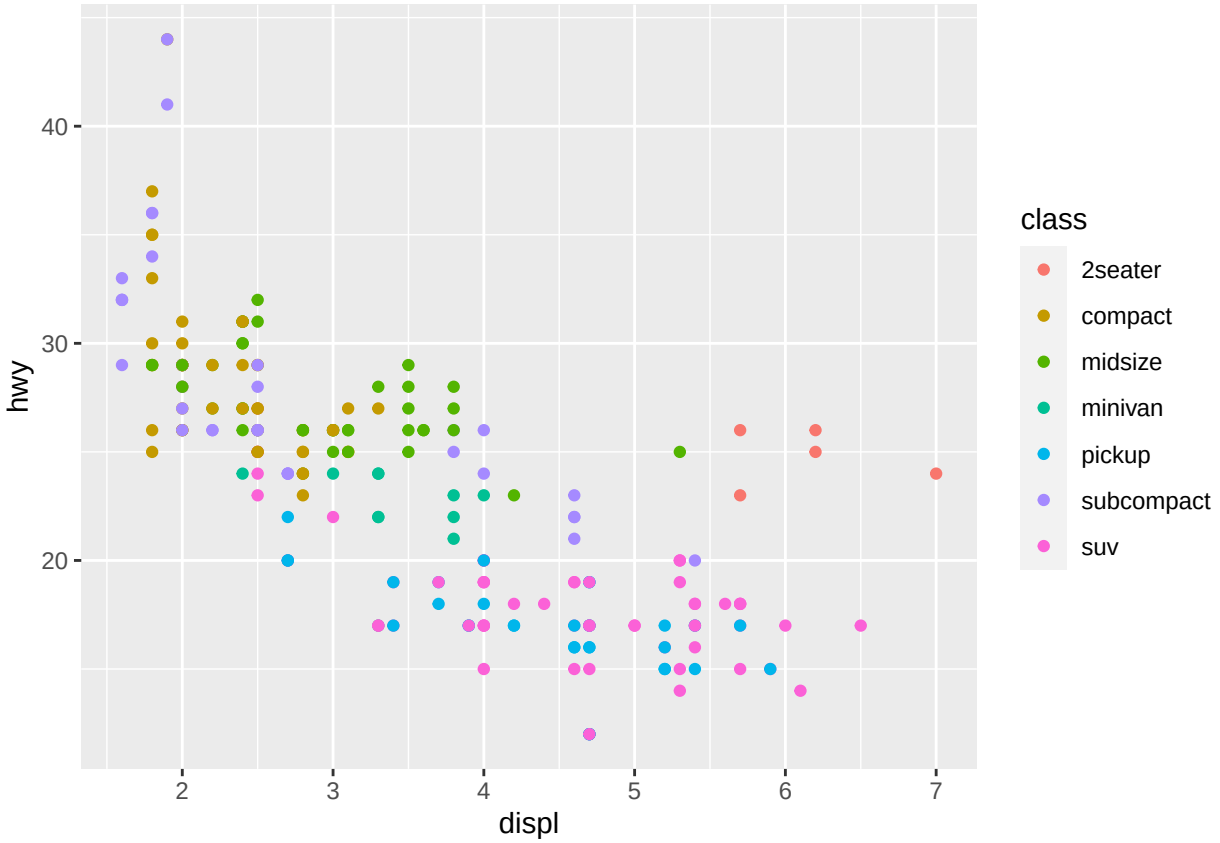
```
#new plot hwy vs cyl
ggplot(data = mpg) +
  geom_point(mapping = aes(x = cyl, y = hwy))
```



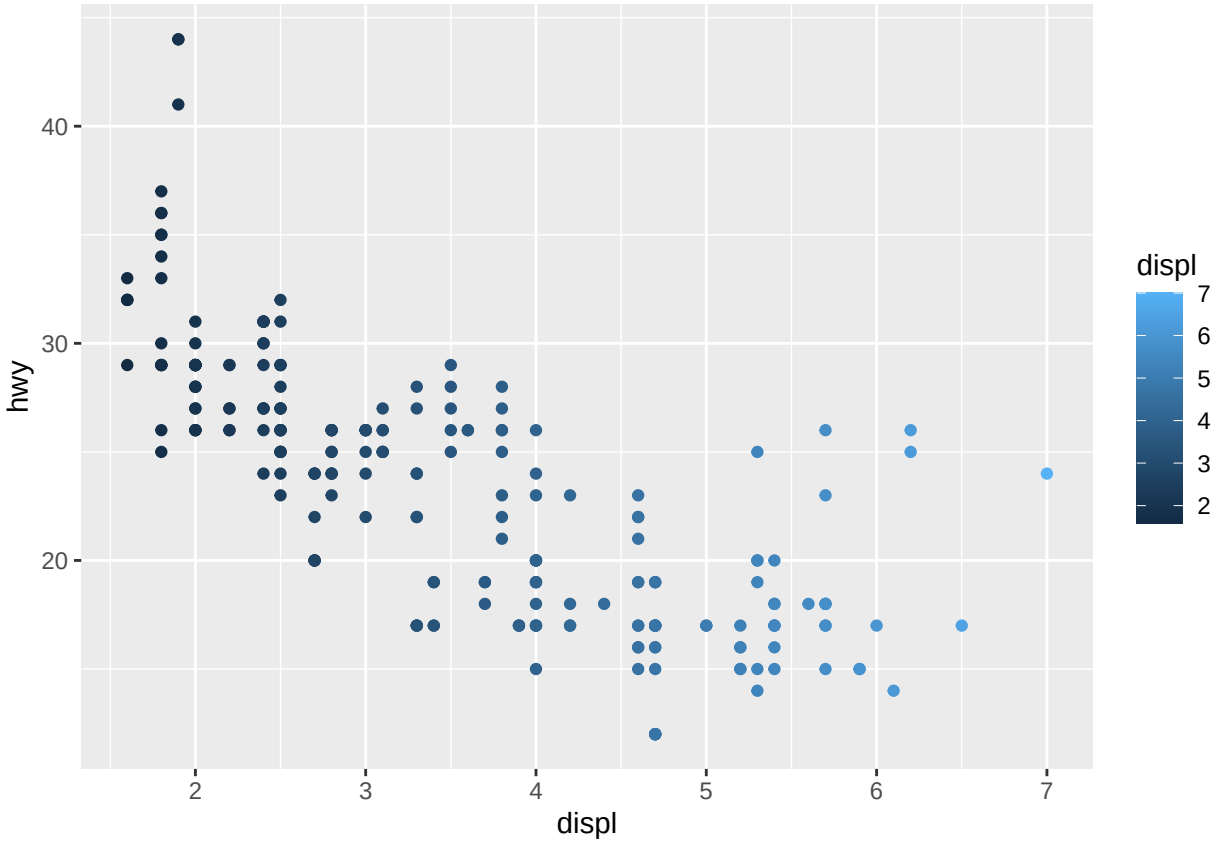

```
#new plot, class vs drv
ggplot(data = mpg) +
  geom_point(mapping = aes(x = drv, y = class))
```



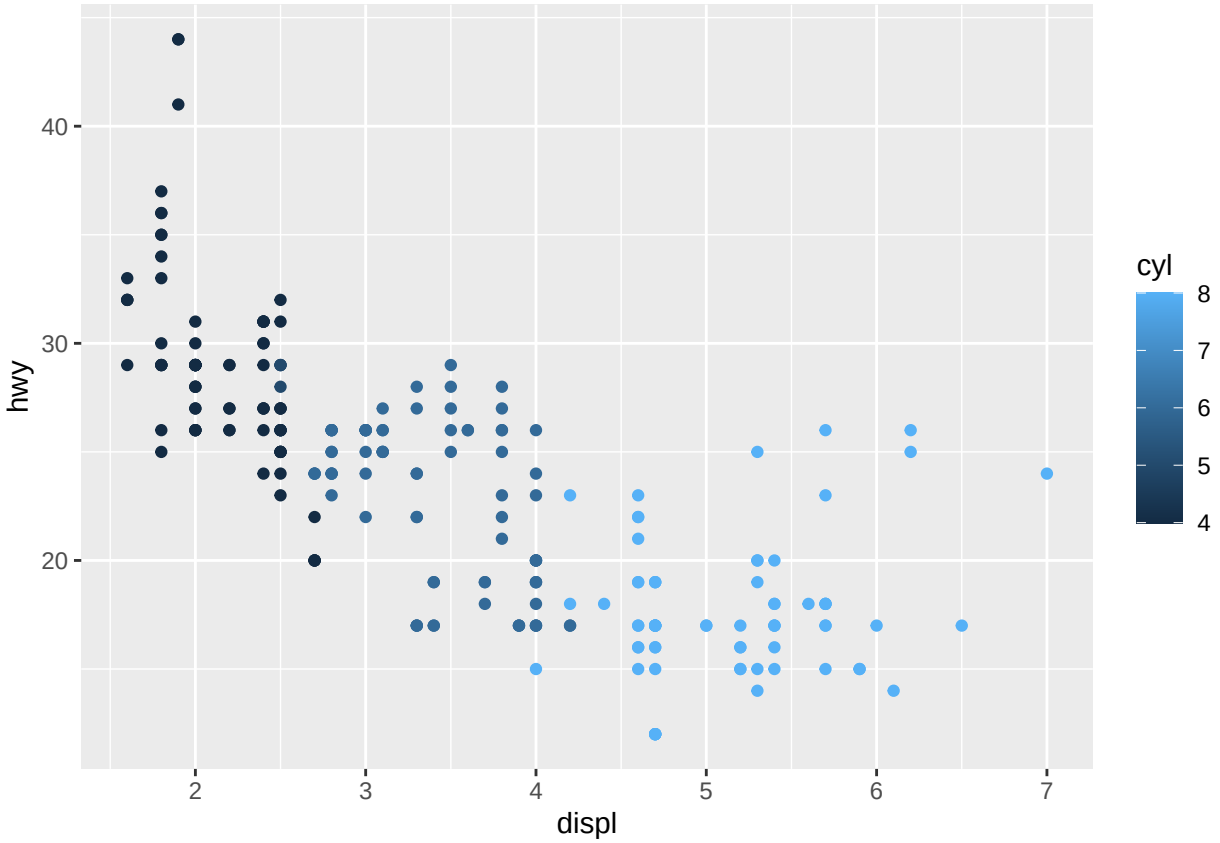

```
#drv, the type of drive train, where f = front-wheel drive, r = rear wheel drive, 4 = 4wd
#class, "type" of car
#this plot is not useful (to determine any fuel efficiency)

#add aestheic, color
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = class))
```



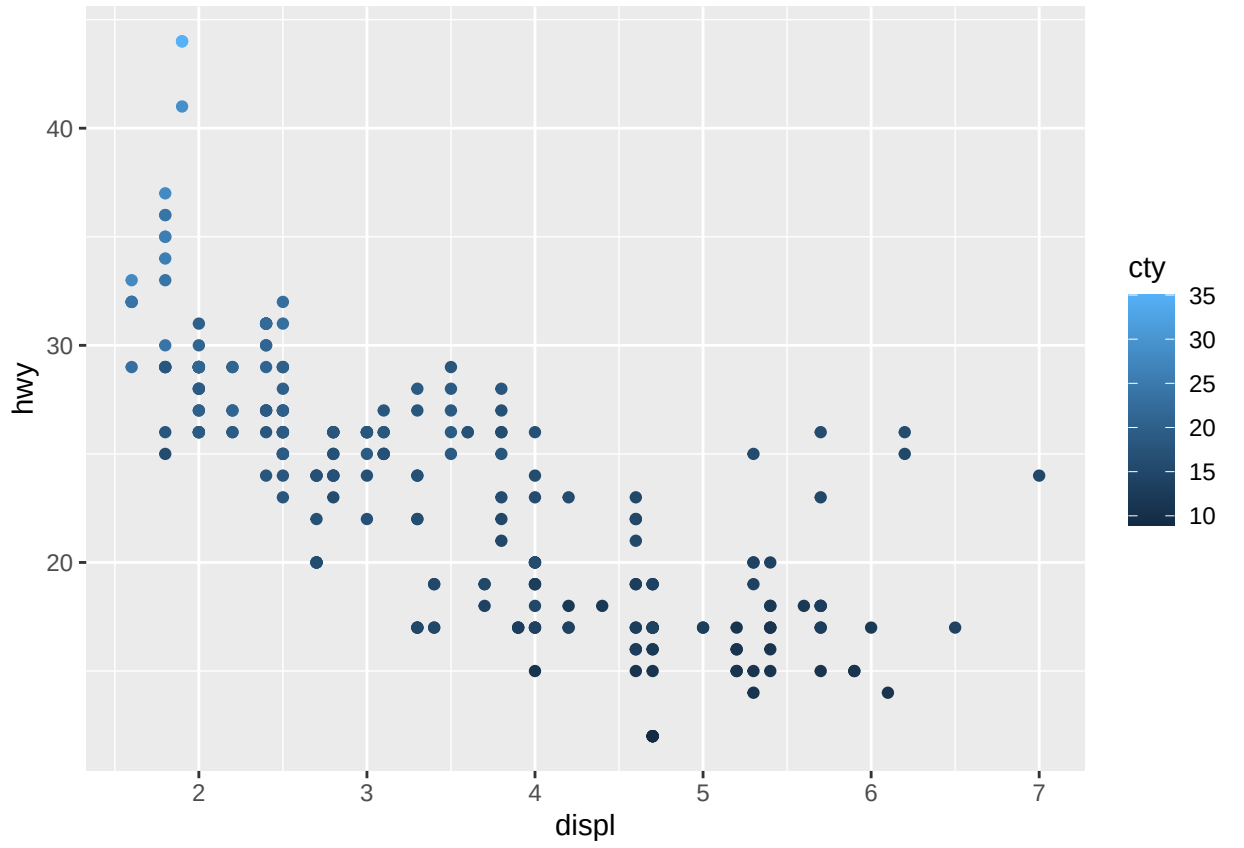

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = displ))
```



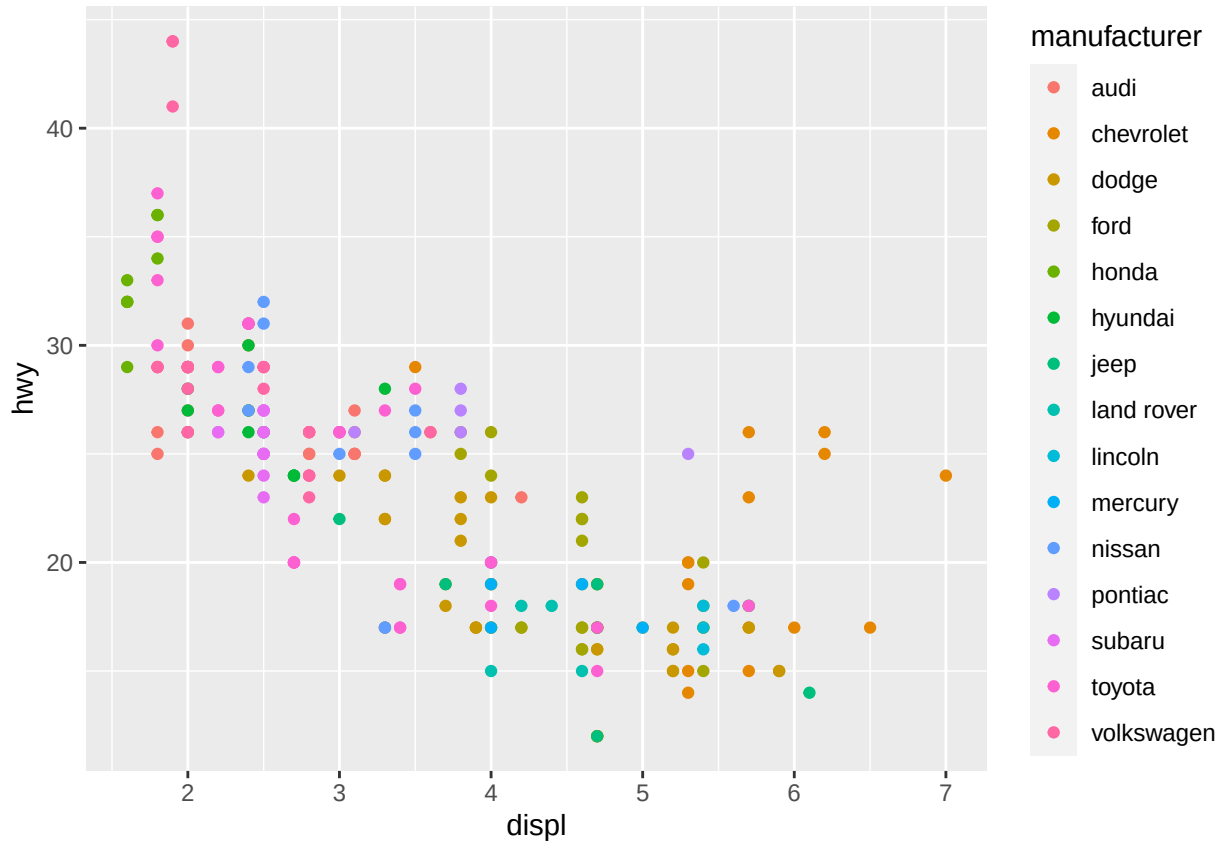

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = cyl))
```




```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = cty))
```

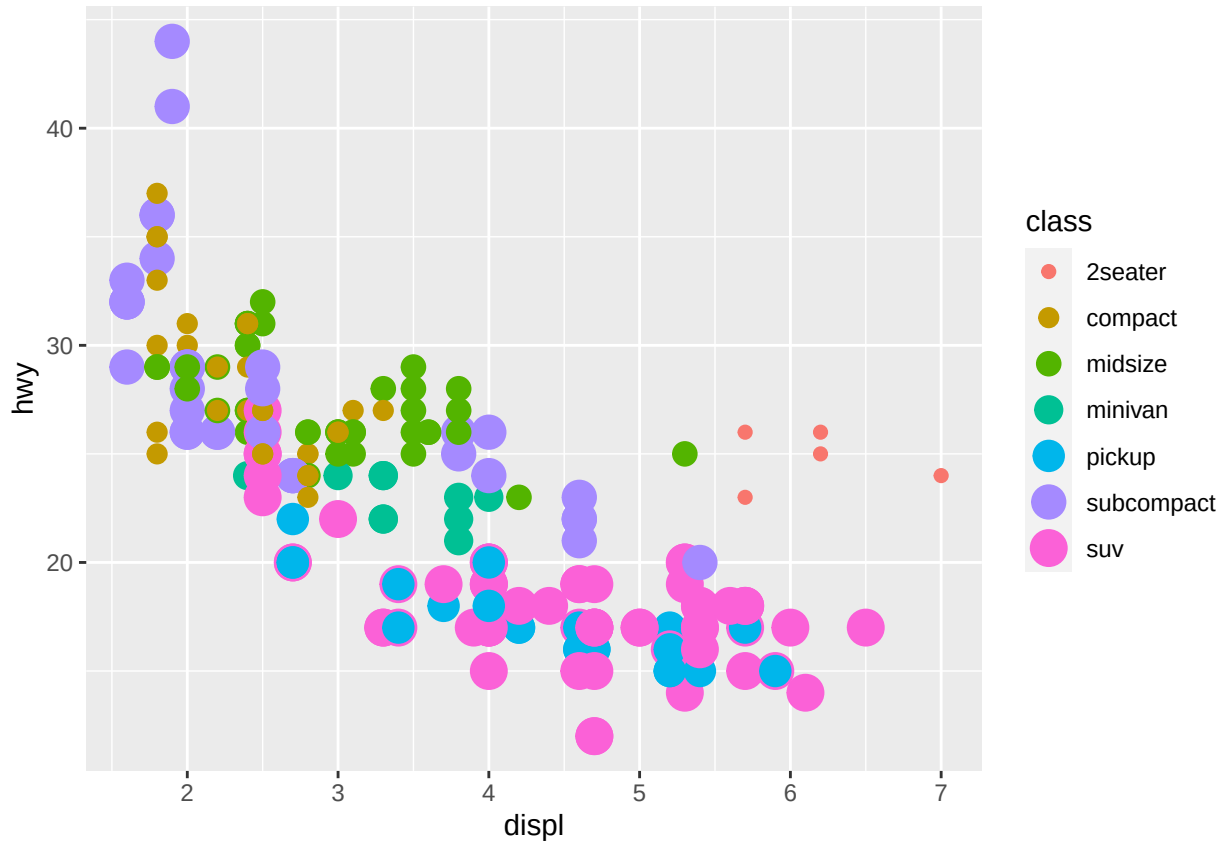



```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = manufacturer))
```




```
#add aestheic, size
ggplot(data = mpg) +
  geom_point(mapping = aes(x= displ, y= hwy, color= class, size= class))
```

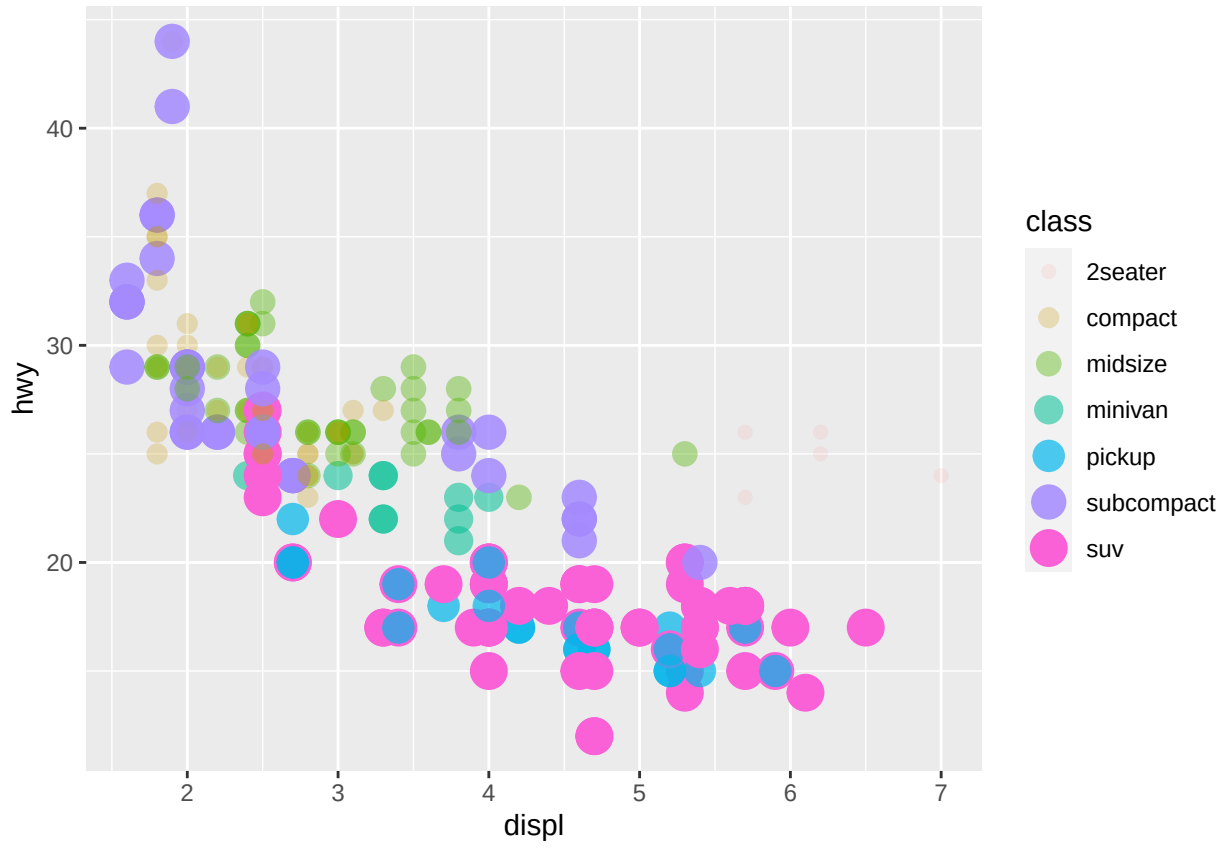
```
## Warning: Using size for a discrete variable is not advised.
```



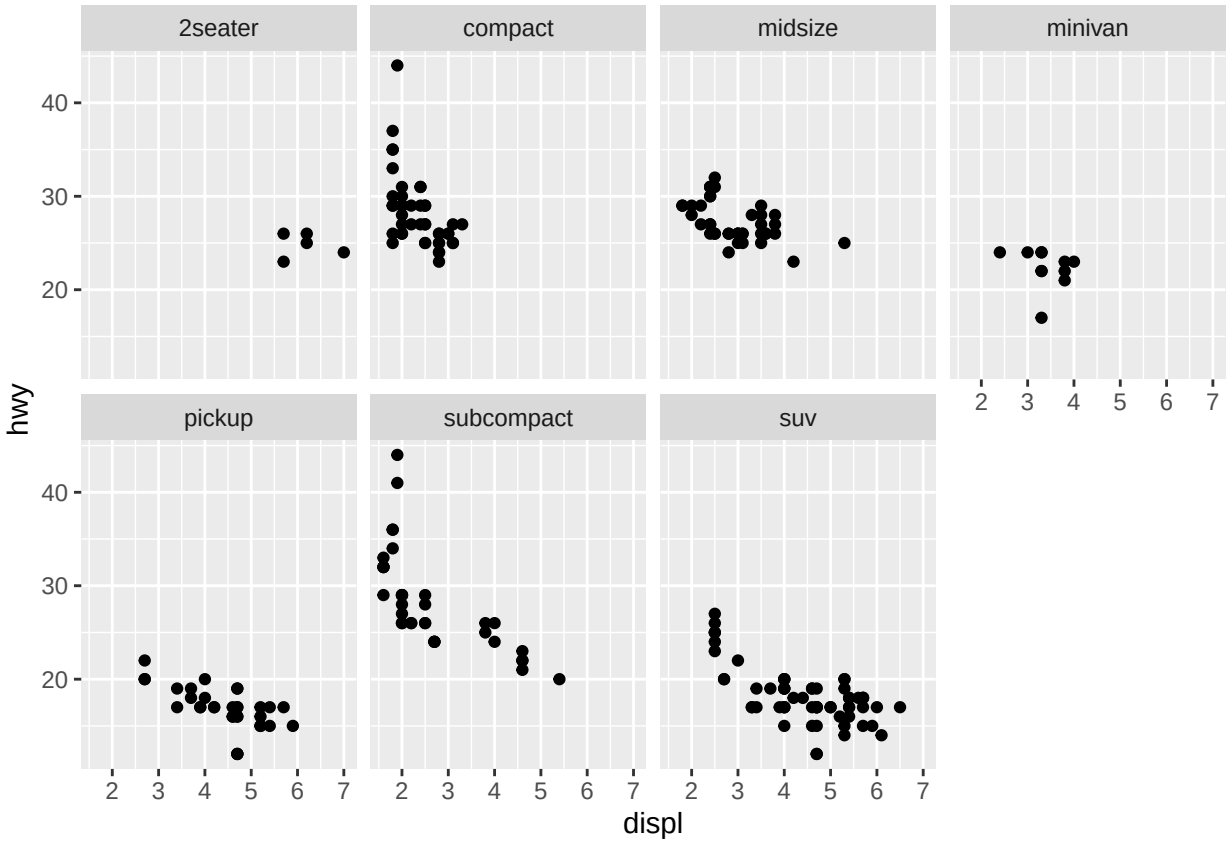

```
#add aestheic, alpha (or transperency)
ggplot(data = mpg) +
  geom_point(mapping = aes(x= displ, y= hwy, color= class, size= class, alpha= class))
```

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

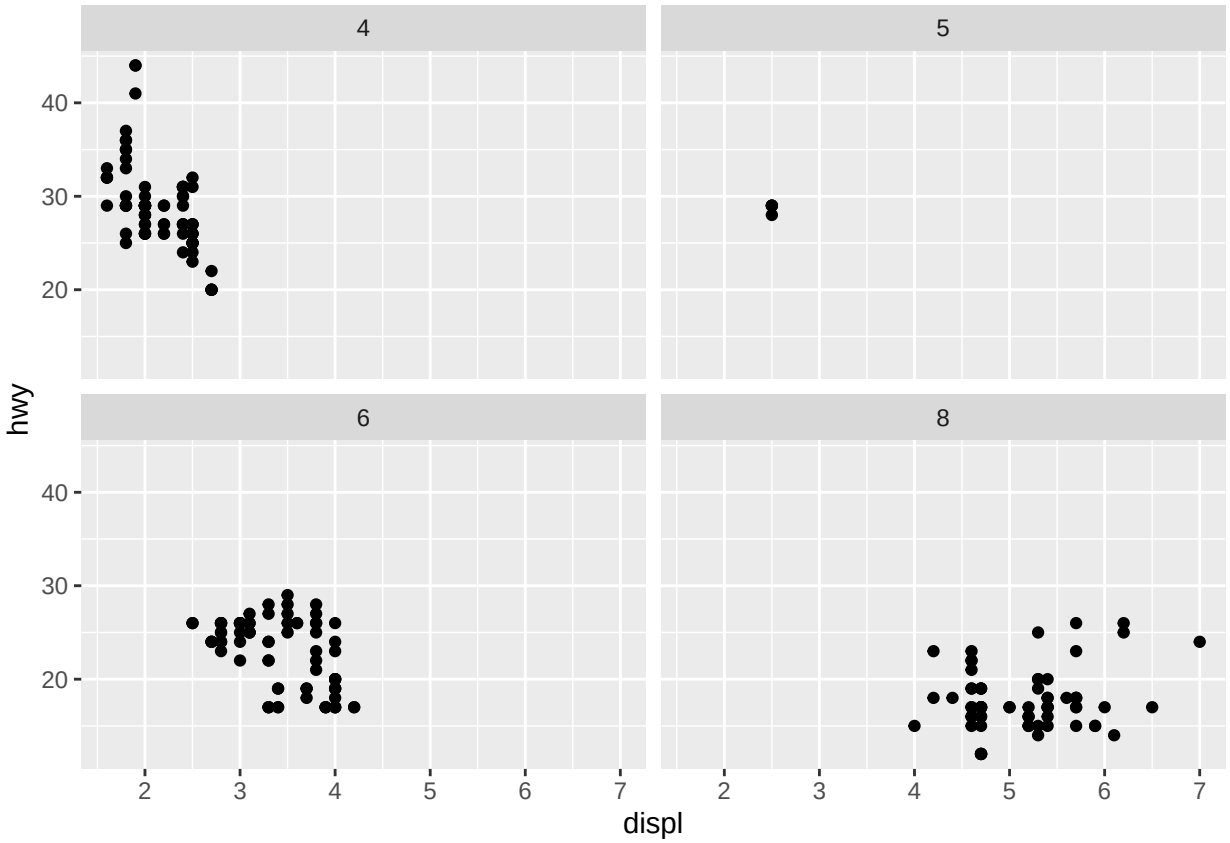
```
## Warning: Using alpha for a discrete variable is not advised.
```



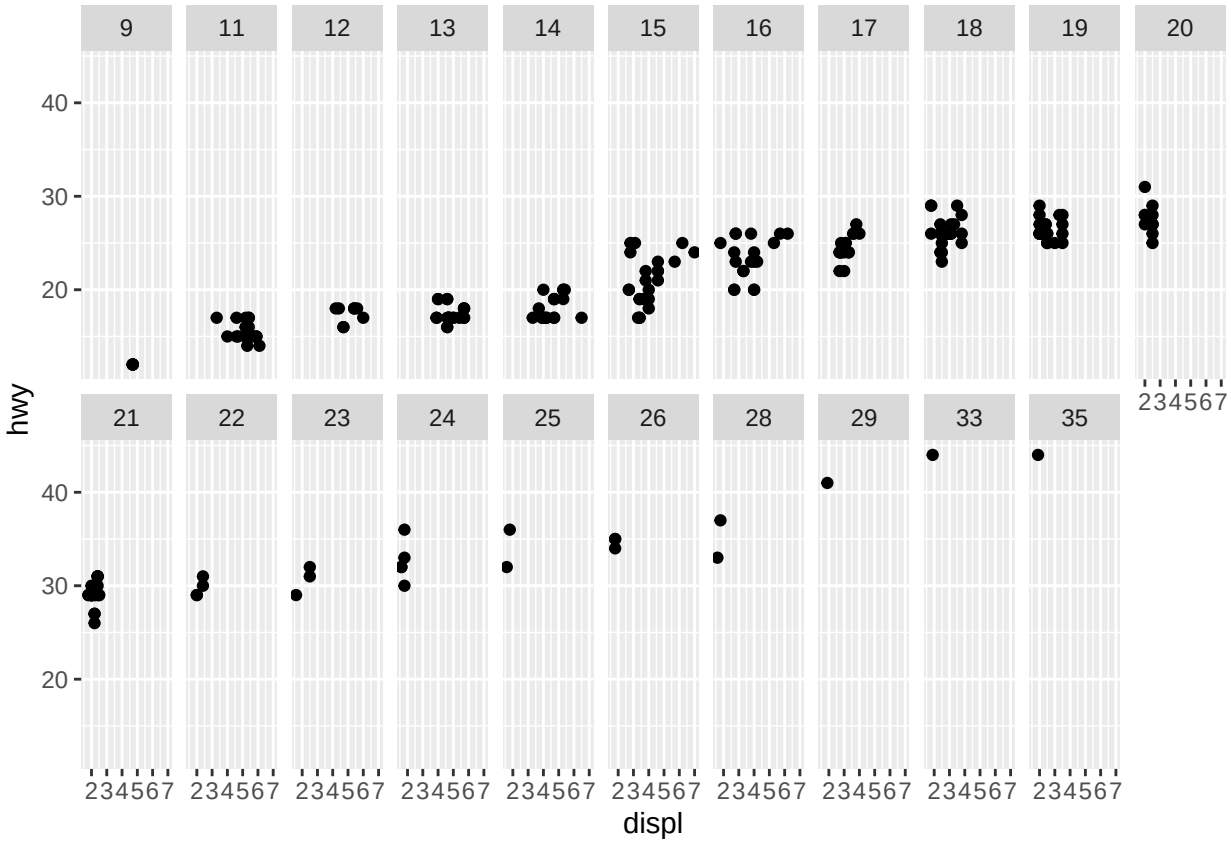

```
#facets
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(~ class, nrow = 2)
```



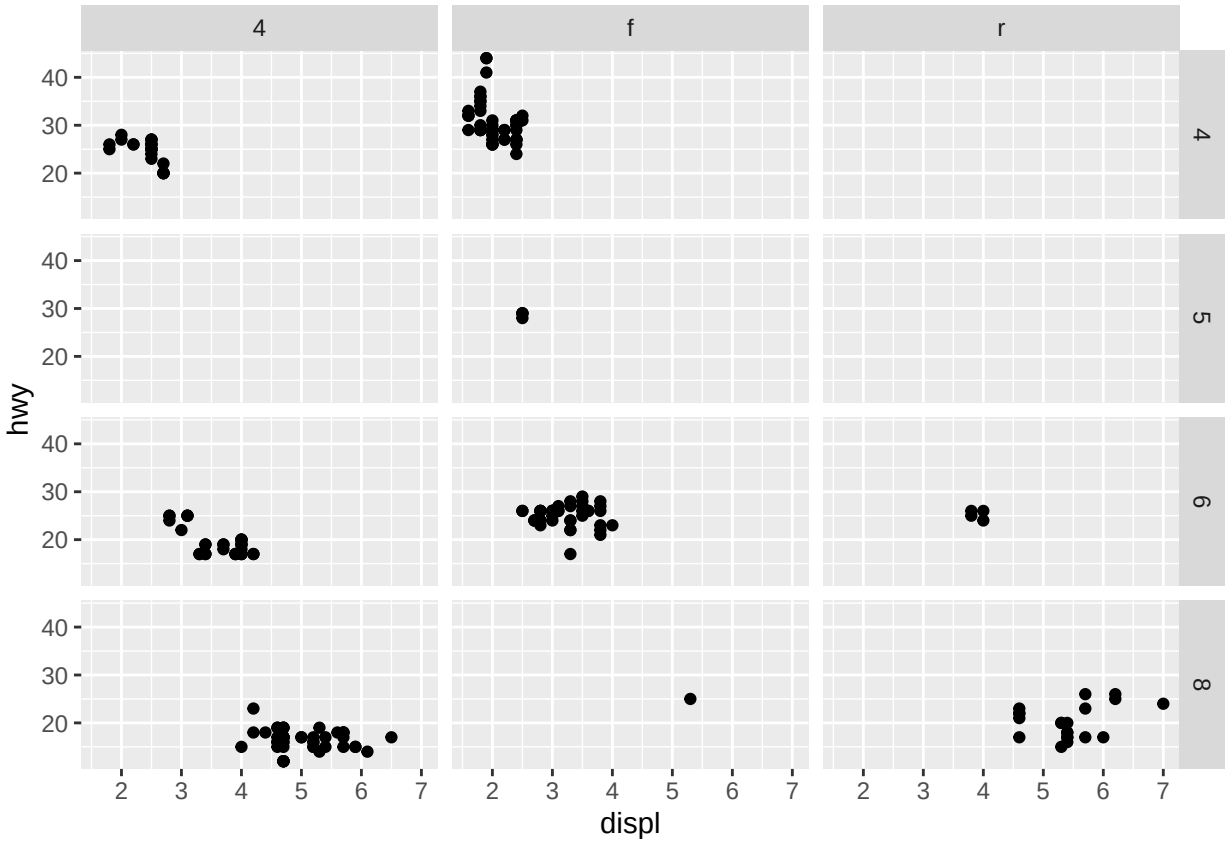

```
#facets, non-continous/categorical facet variable
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(~ cyl, nrow = 2)
```



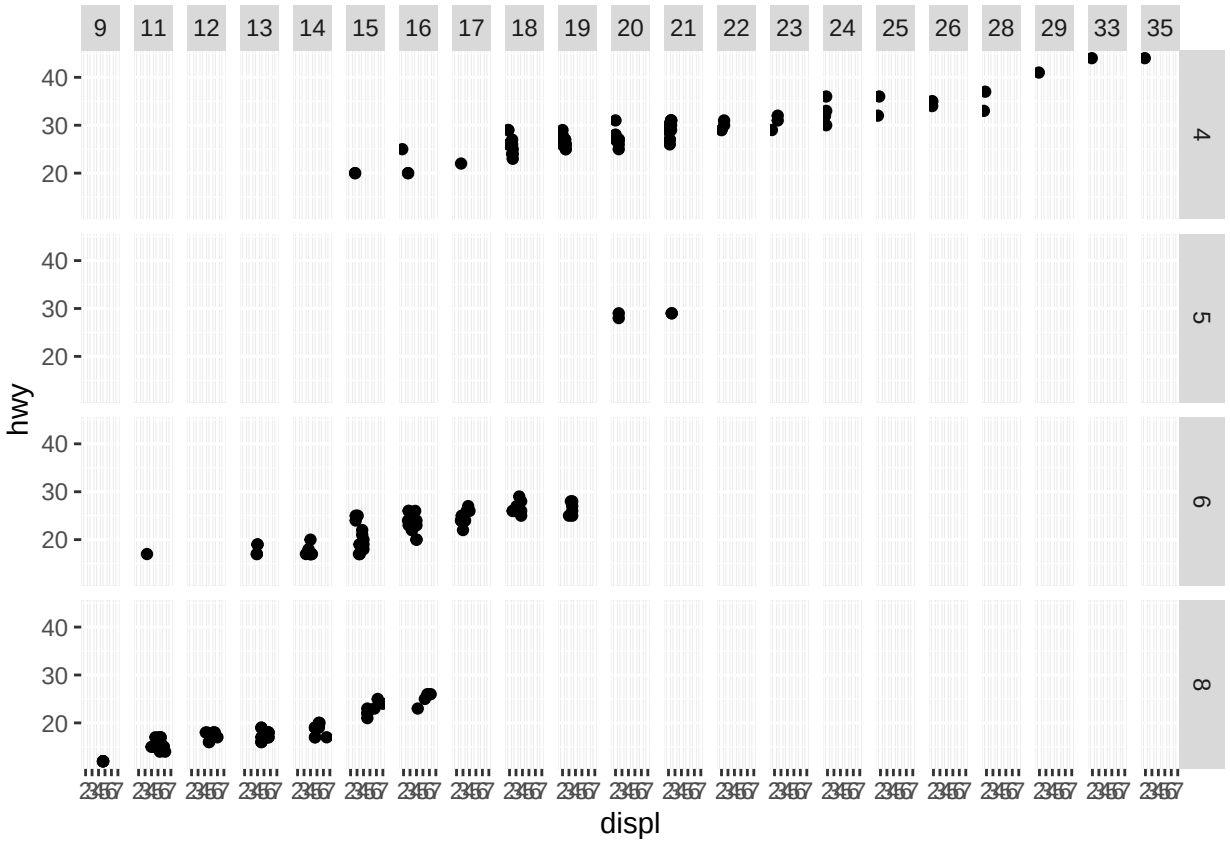

```
#facets, continous/NON-categorical facet variable
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(~ cty, nrow = 2)
```



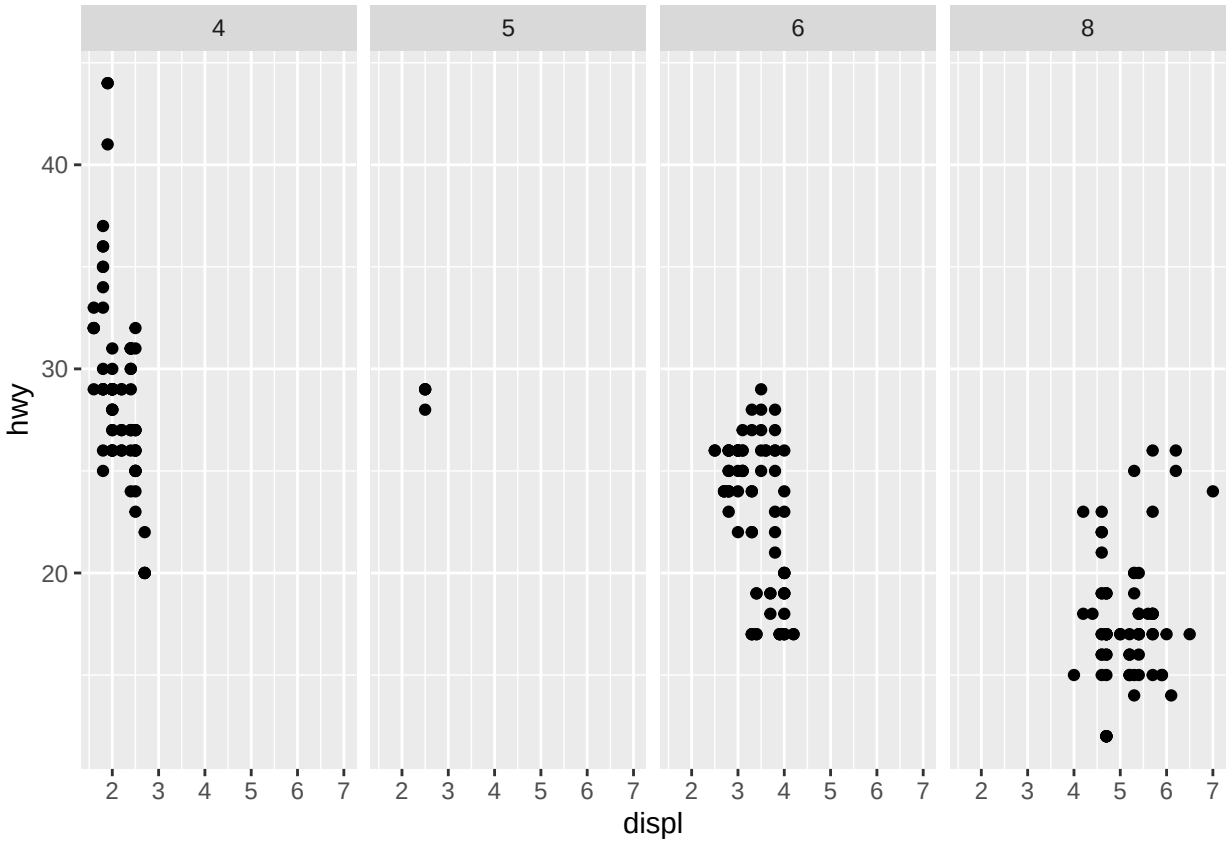

```
#facet grid, 2-additional predictor variables
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_grid(cyl ~ drv)
```



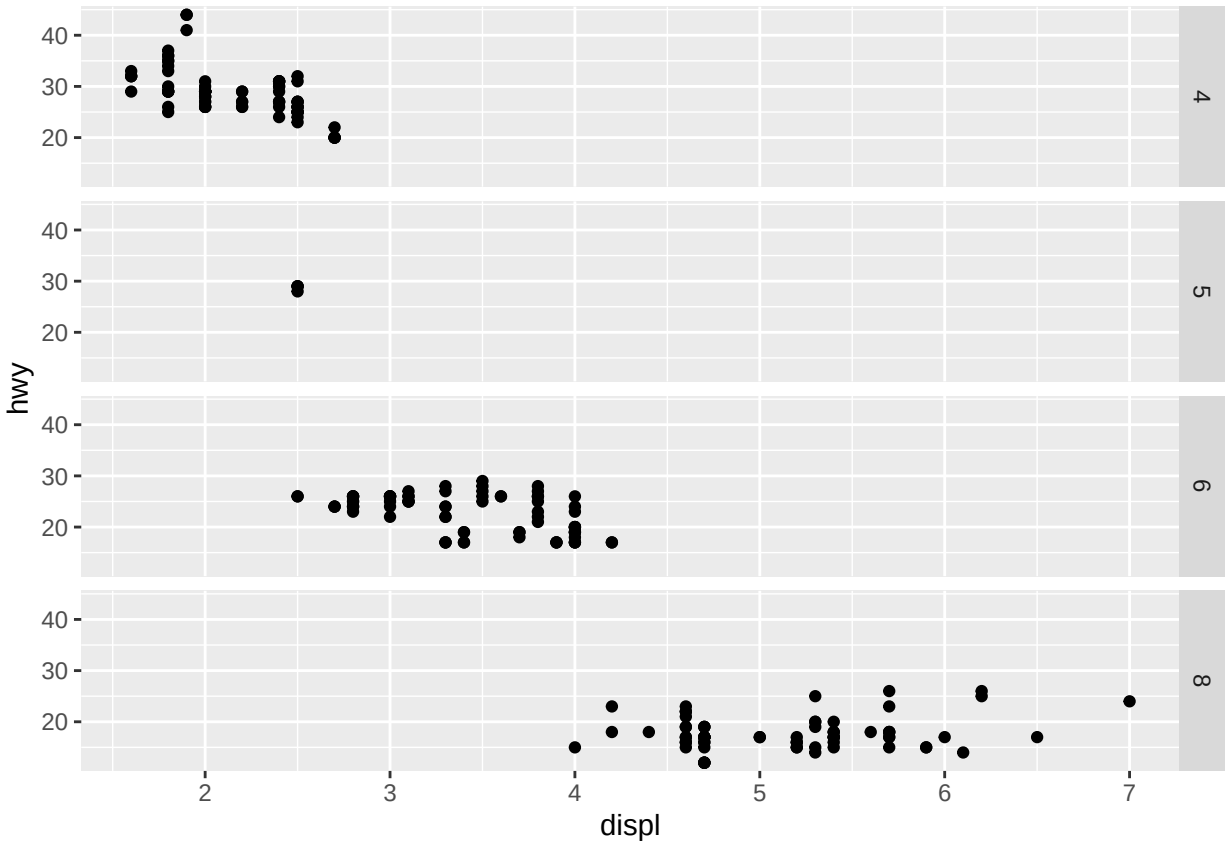

```
#facet-grid, 2-additional predictor variables
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_grid(cyl ~ cty)
```



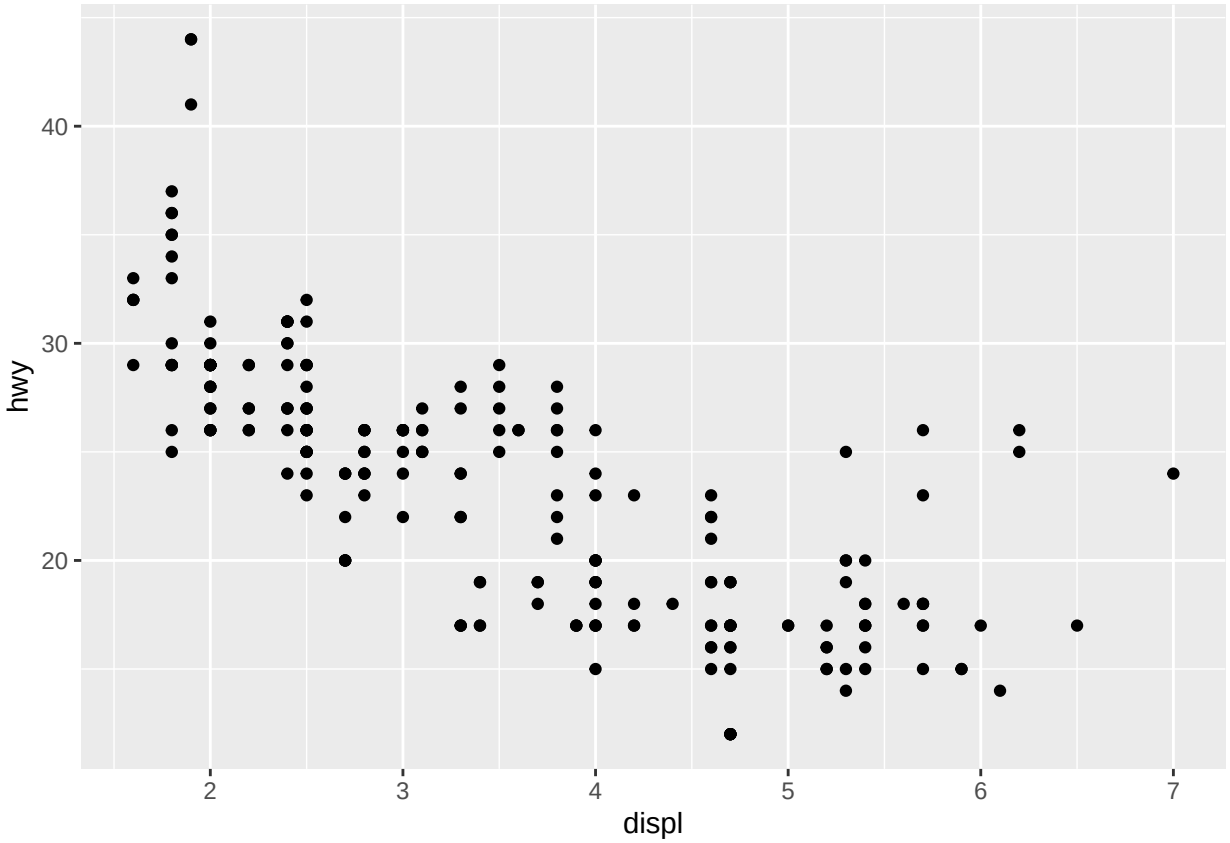

```
#facet-grid, 1-additional predictor variable
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_grid(. ~ cyl) #facet variable axis, opposite to x-axis
```




```
#facet-grid, 1-additional predictor variable
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_grid(cyl~ .) #facet variable axis, opposite to y-axis
```

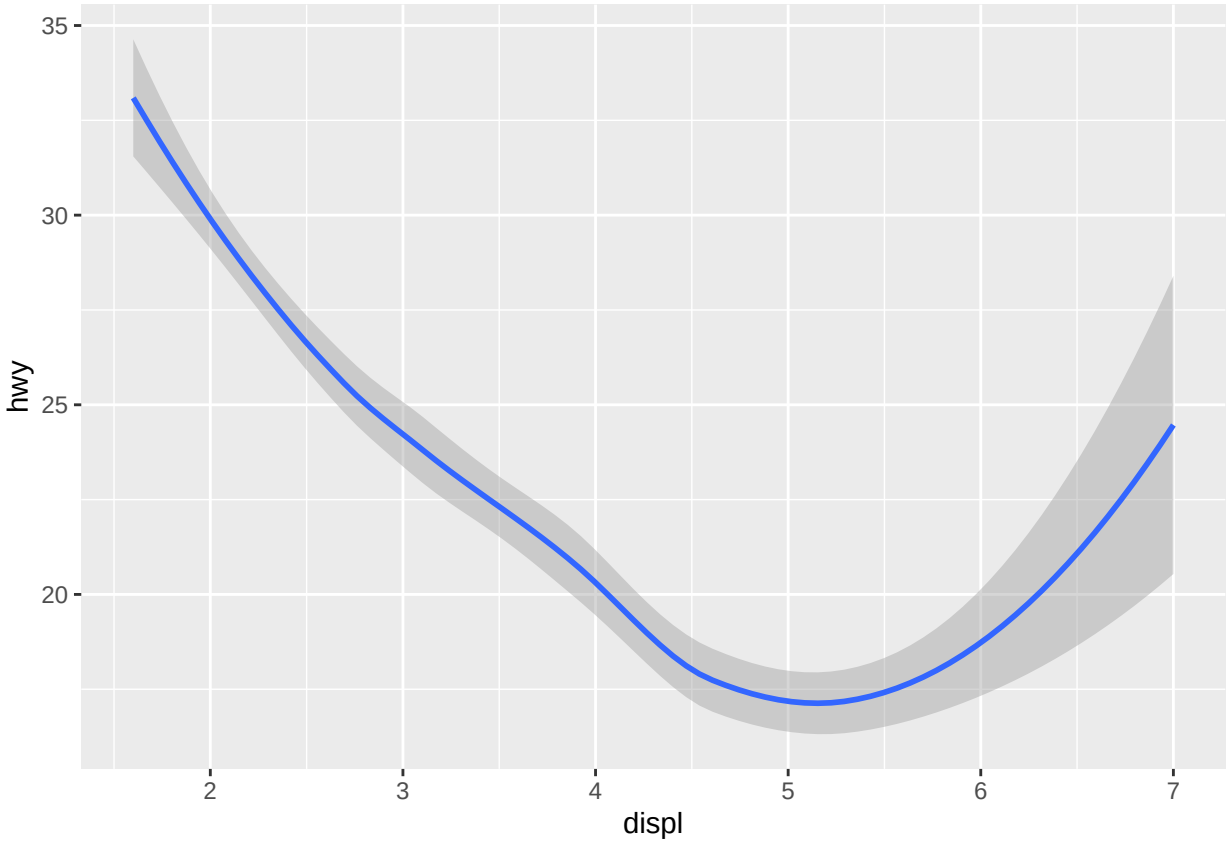



```
# scatter-plot
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```

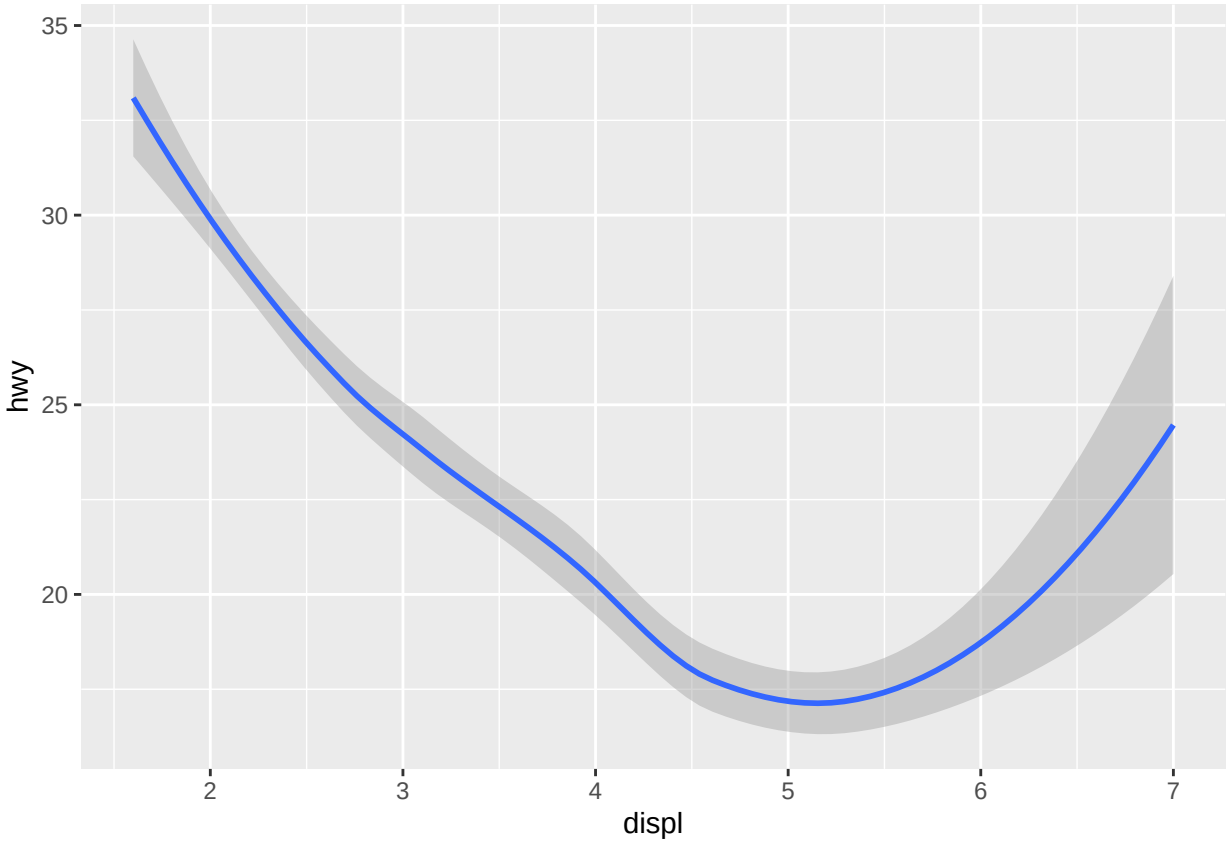



```
# smoothed line plot, method selected automatically
# refer to R-output
ggplot(data = mpg) +
  geom_smooth(mapping = aes(x = displ, y = hwy))
```

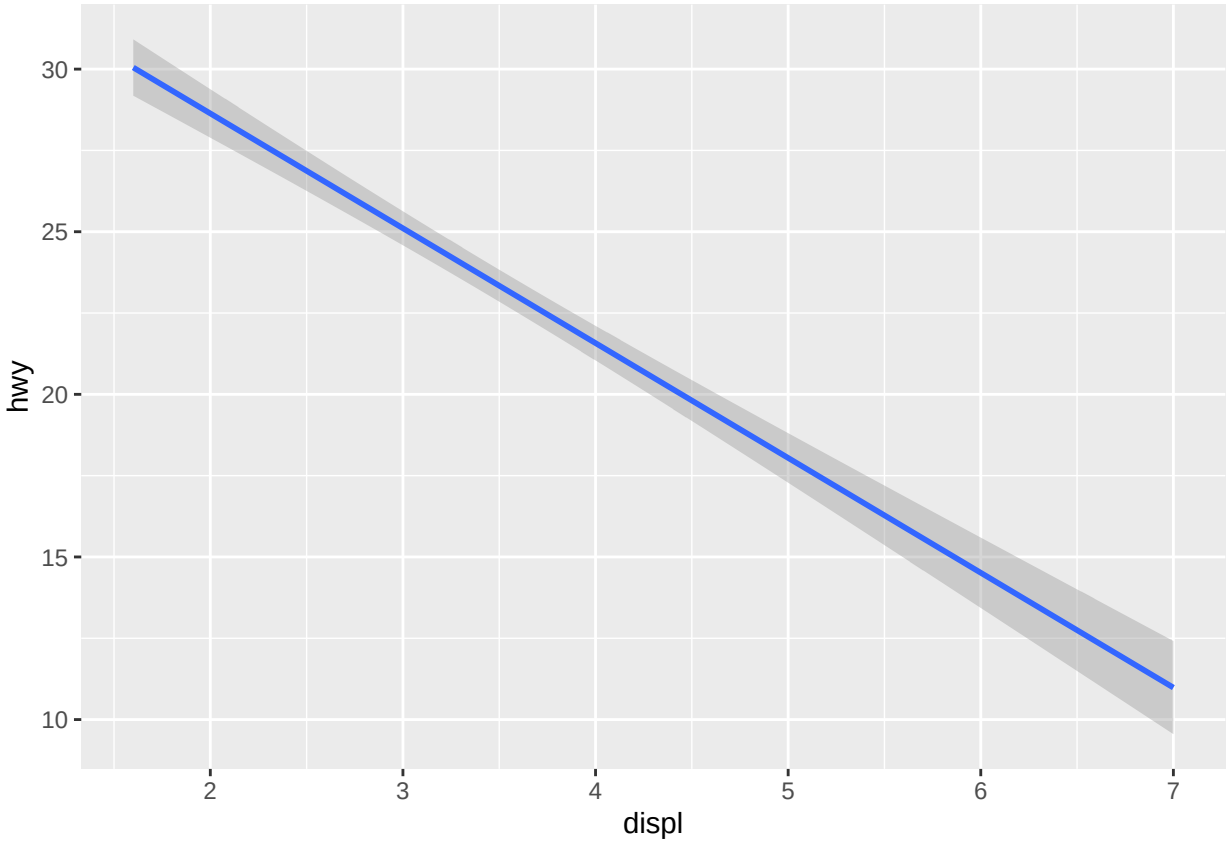
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



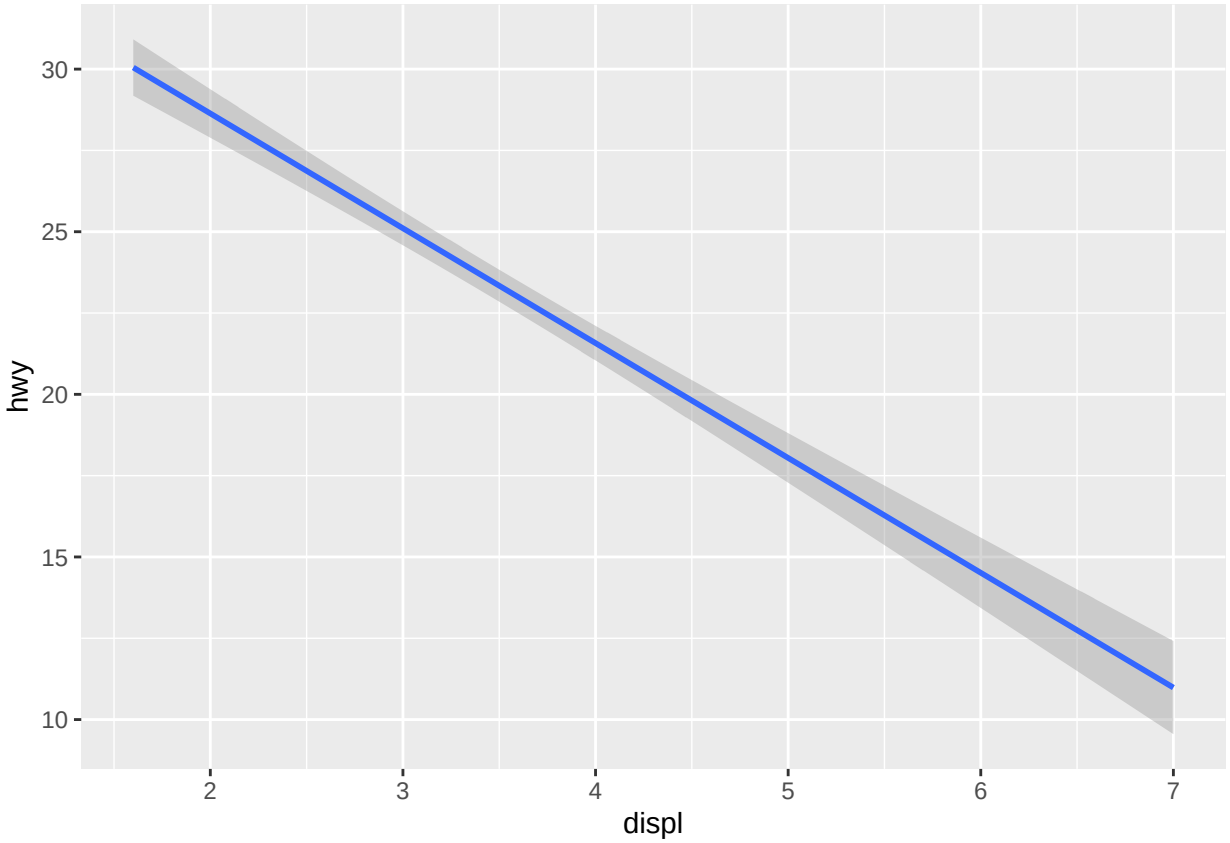

```
# smoothed line plot, explicit method specified
ggplot(data = mpg) +
  geom_smooth(mapping = aes(x = displ, y = hwy),
              method = "loess", formula= y~x)
```



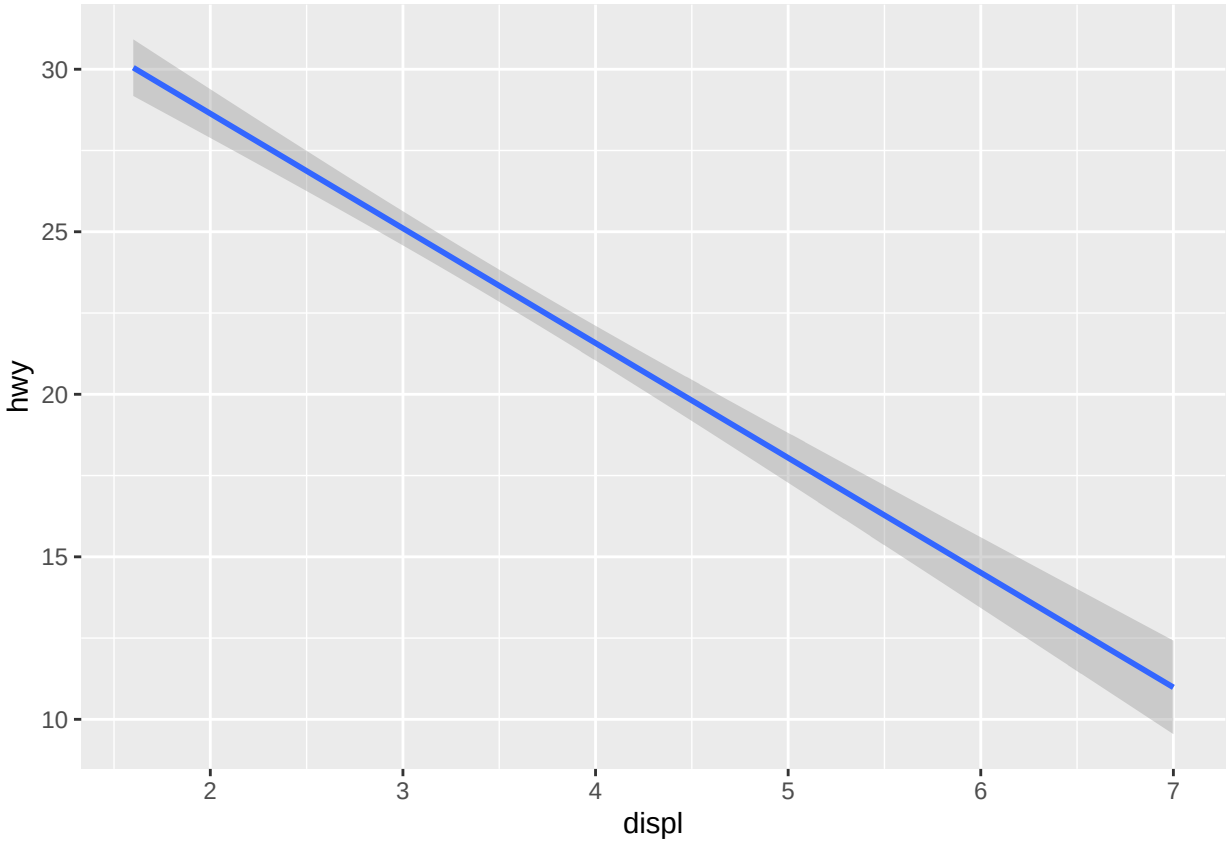

```
# smoothed line plot
ggplot(data = mpg) +
  geom_smooth(mapping = aes(x = displ, y = hwy),
              method = "gam", formula= y~x)
```




```
# smoothed line plot
ggplot(data = mpg) +
  geom_smooth(mapping = aes(x = displ, y = hwy),
              method = "glm", formula= y~x)
```

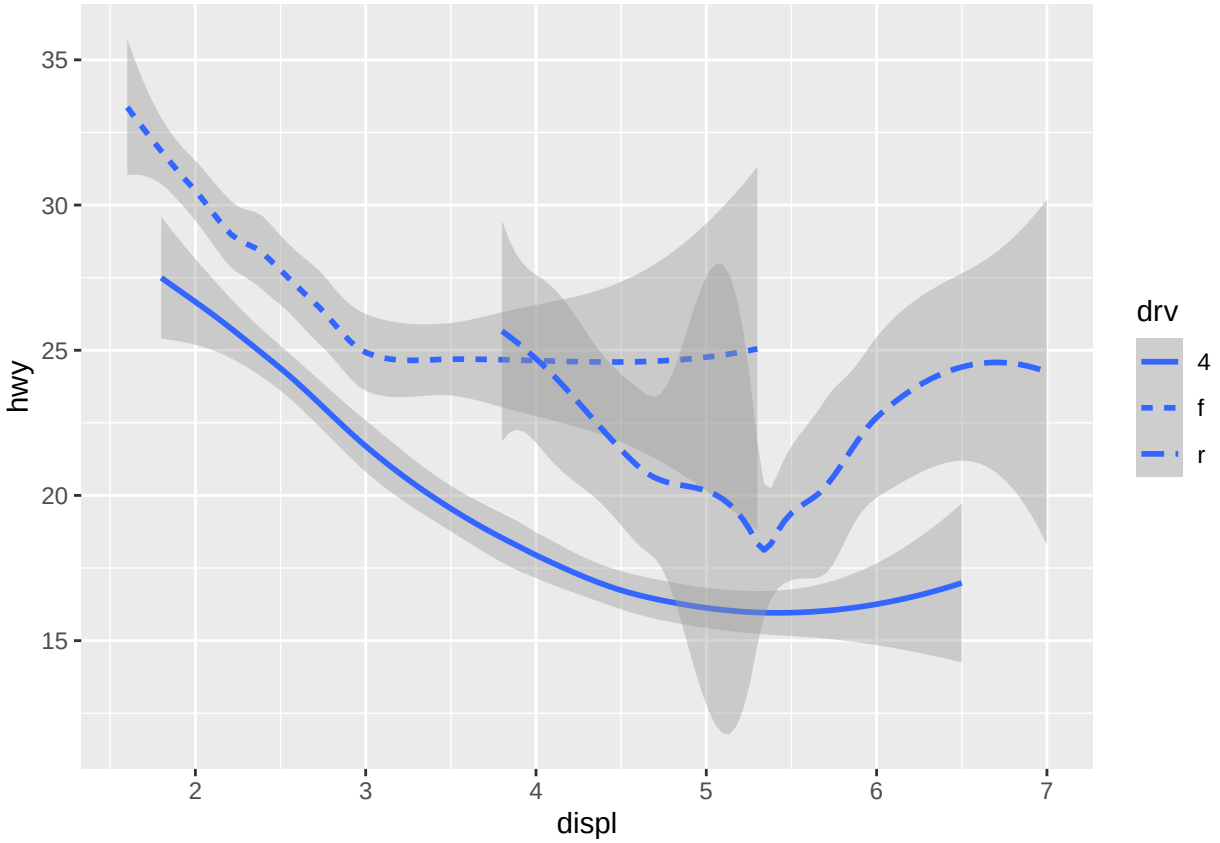



```
# smoothed line plot
ggplot(data = mpg) +
  geom_smooth(mapping = aes(x = displ, y = hwy),
              method = "lm", formula= y~x)
```




```
# add aesthetic linetype, categorical-only/non-continous
ggplot(data = mpg) +
  geom_smooth(mapping = aes(x = displ, y = hwy, linetype = drv))
```

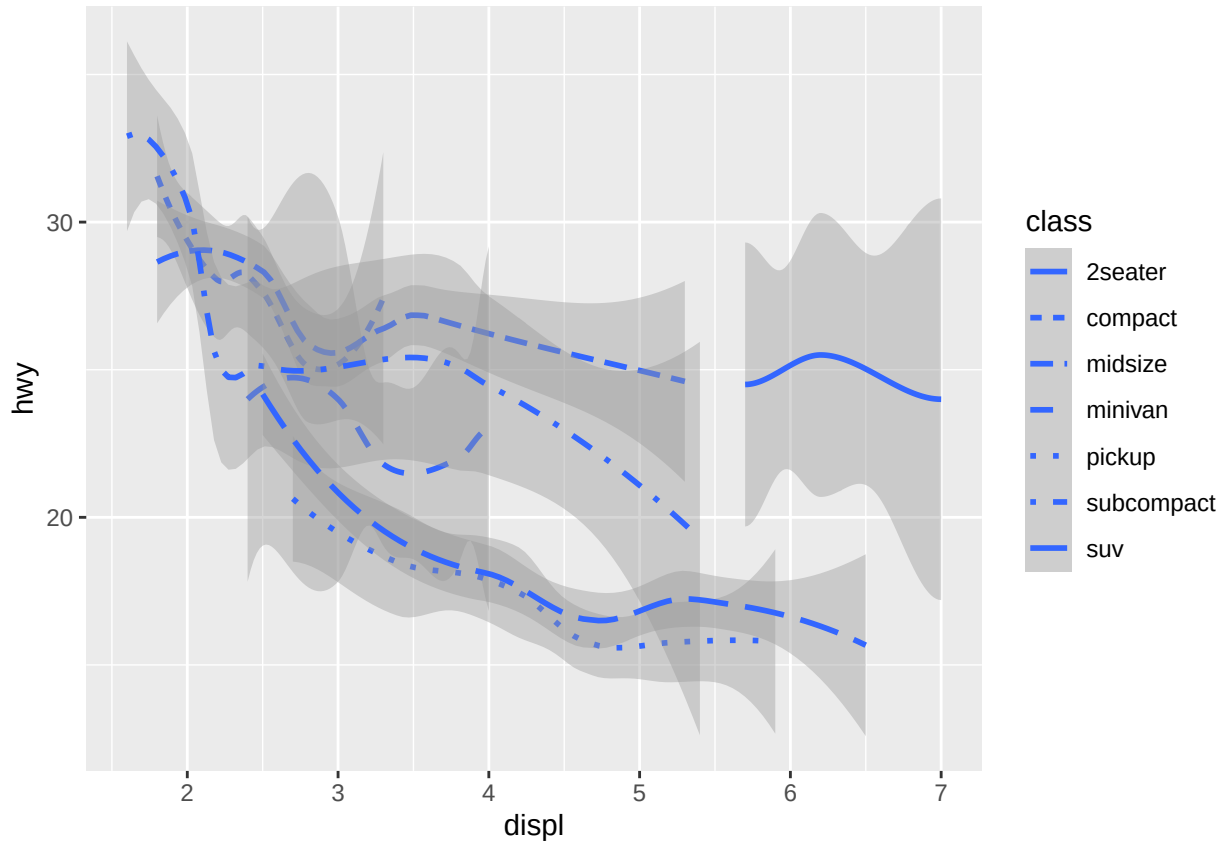
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



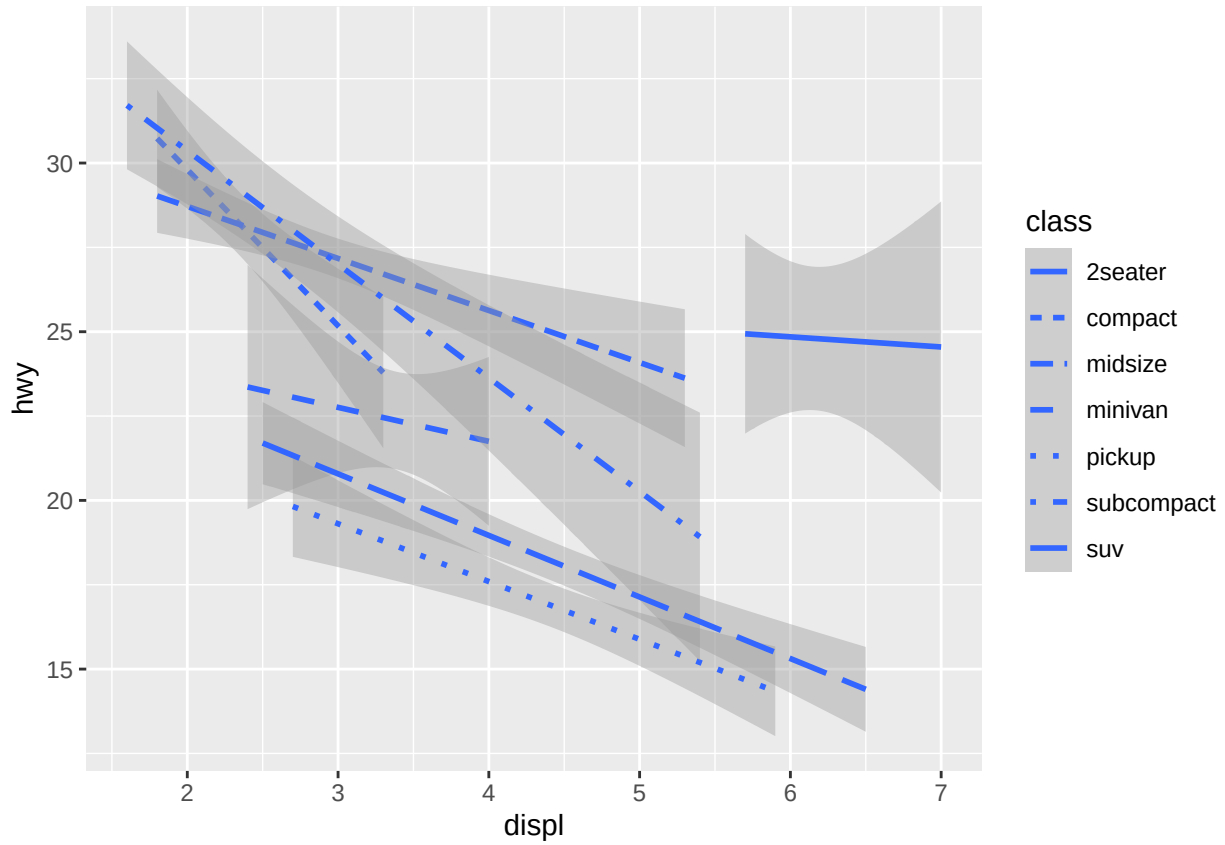

```
# add aesthetic linetype, categorical-only/non-continous
#ggplot(data = mpg) +
  #geom_smooth(mapping = aes(x = displ, y = hwy, linetype = cyl))
# Error: A continuous variable can not be mapped to linetype
# cyl is continuous-variable

# add aesthetic linetype, categorical-only/non-continous
ggplot(data = mpg) +
  geom_smooth(mapping = aes(x = displ, y = hwy, linetype = class))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 5.6935

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 0.5065

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 0.65044

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
```

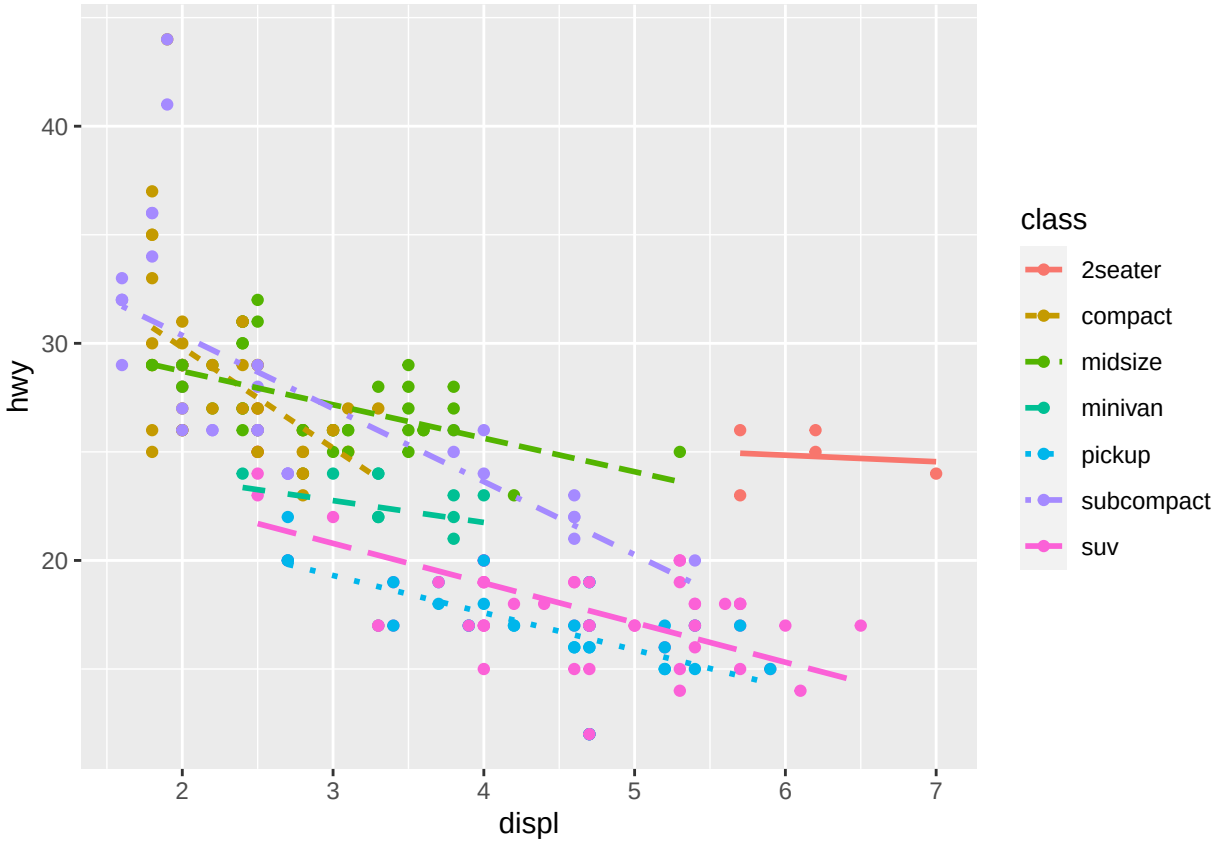
```
## as.matrix(model.frame(delete.response(terms(object)), : span too small. fewer
## data values than degrees of freedom.

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
## 5.6935

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 0.5065

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
## number 0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : There are other near
## singularities as well. 0.65044

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 4.008

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 0.708

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 0.25

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
## 4.008

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 0.708

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
## number 0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : There are other near
## singularities as well. 0.25
```



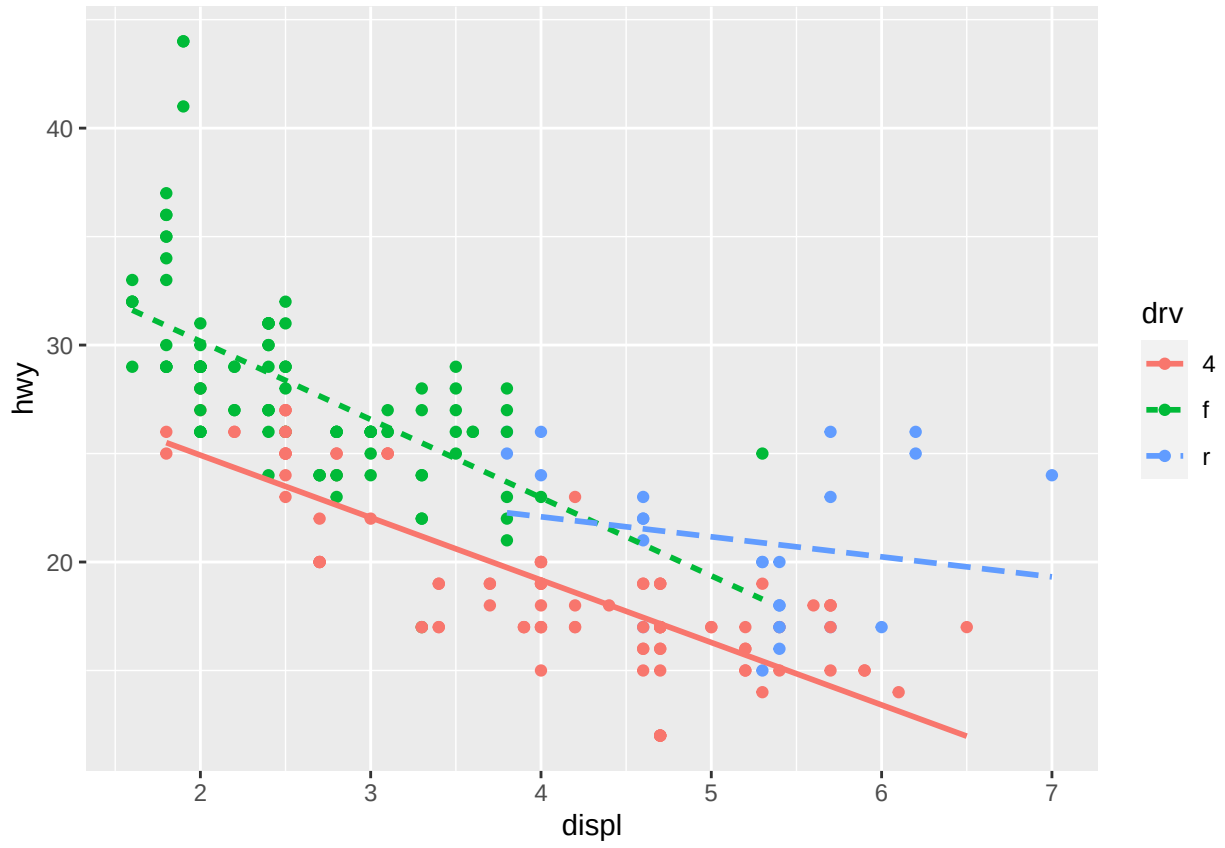

```
# add aesthetic linetype, categorical-only/non-continous
ggplot(data = mpg) +
  geom_smooth(mapping = aes(x = displ, y = hwy, linetype = class),
              method = "lm", formula= y~x, se= TRUE)
```



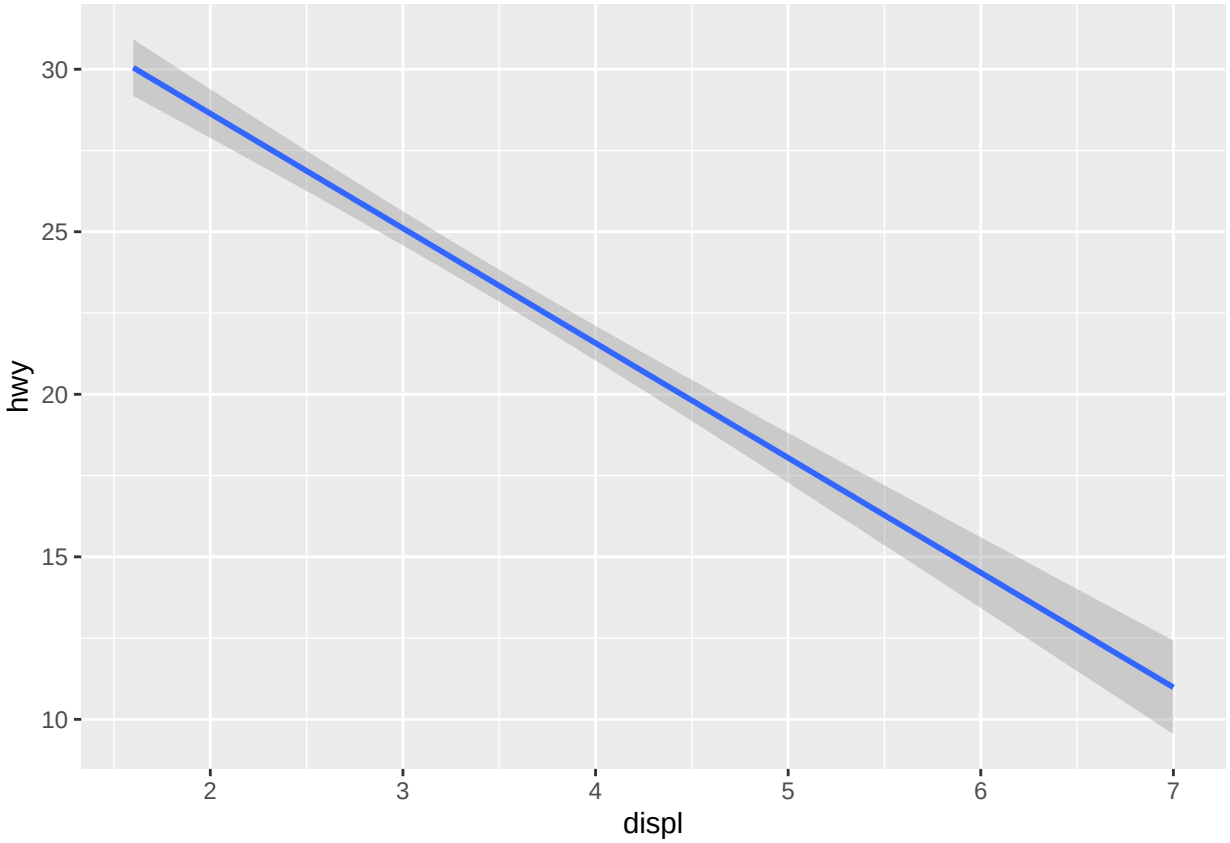

```
# add scatter-plot as another layer
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color= class)) +
  geom_smooth(mapping = aes(x = displ, y = hwy, linetype = class, color= class),
              method = "lm", formula= y~x, se= FALSE)
```



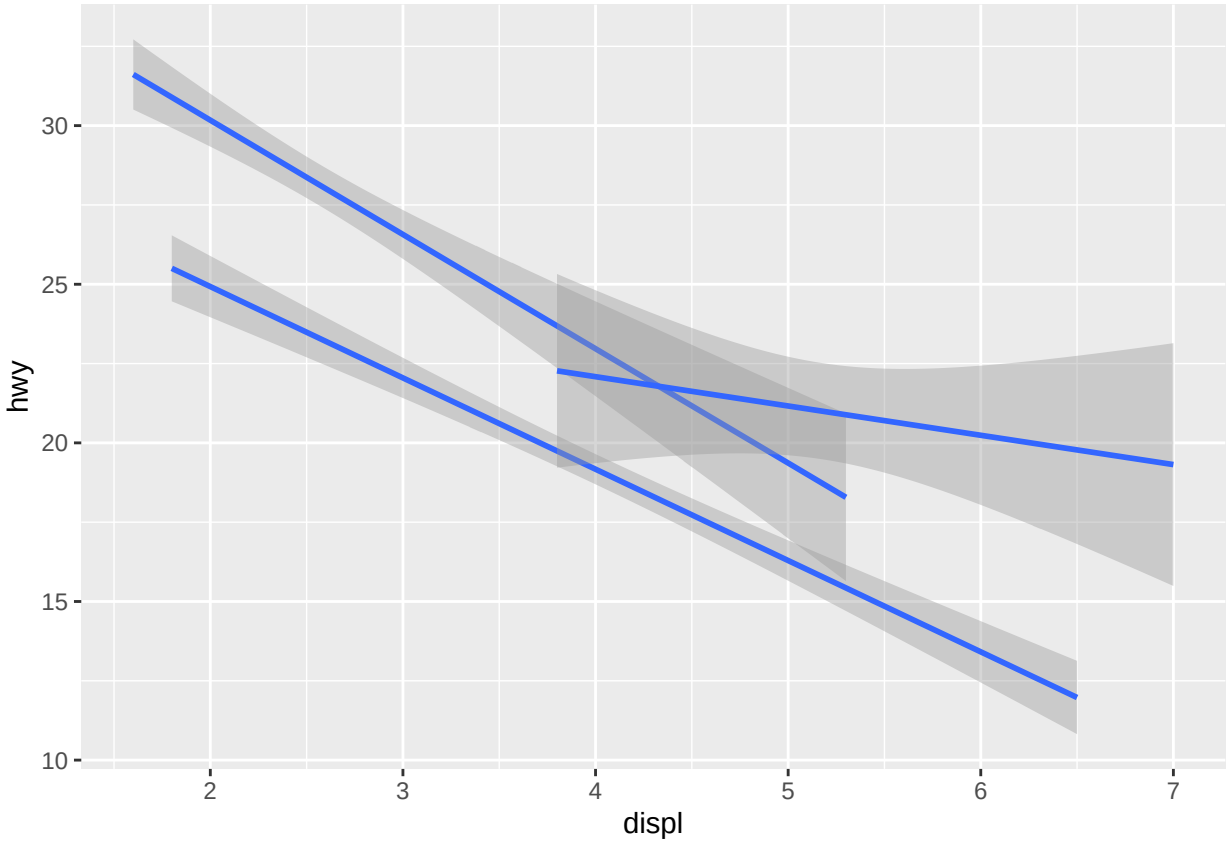

```
# add scatter-plot as another layer
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color= drv)) +
  geom_smooth(mapping = aes(x = displ, y = hwy, linetype = drv, color= drv),
              method = "lm", formula= y~x, se= FALSE)
```



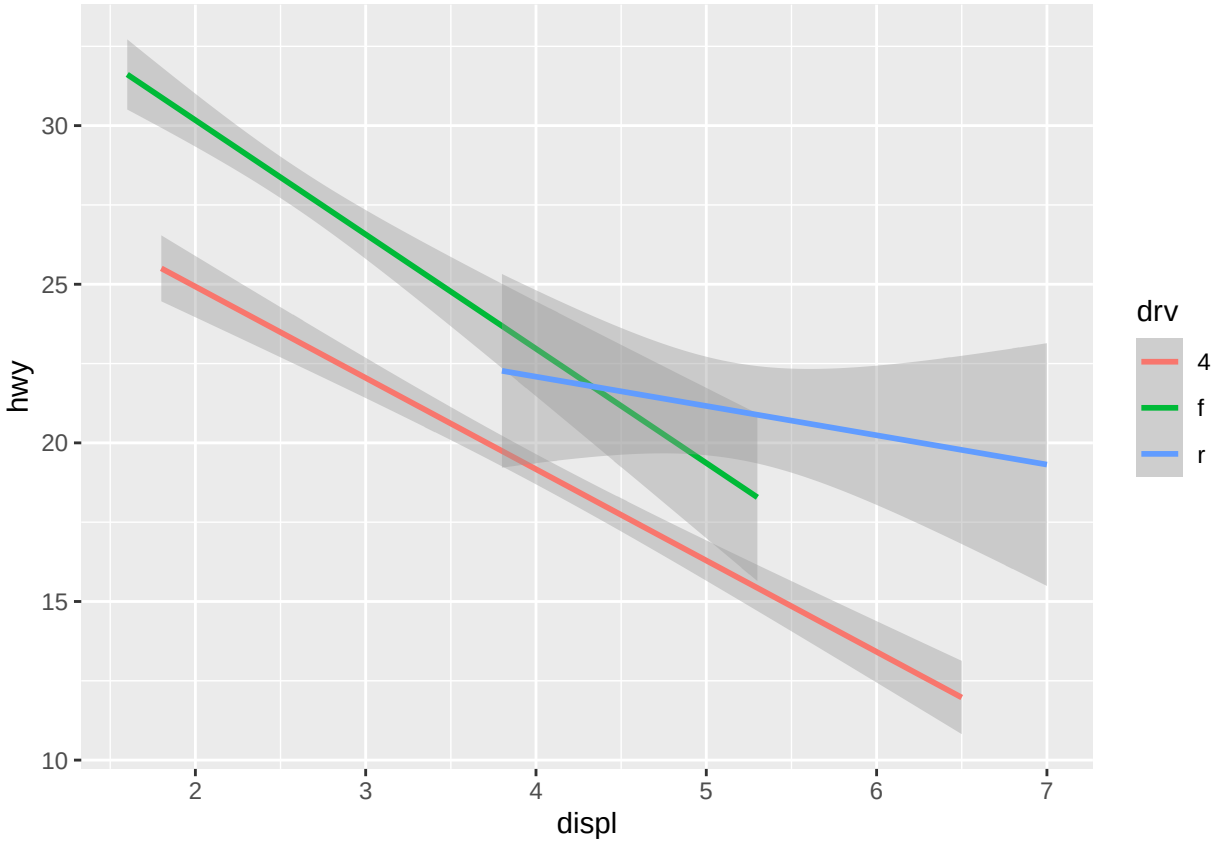

```
ggplot(data = mpg) +
  geom_smooth(mapping = aes(x = displ, y = hwy),
              method = "lm", formula= y~x, show.legend = TRUE)
```




```
ggplot(data = mpg) +
  geom_smooth(mapping = aes(x = displ, y = hwy, group = drv),
              method = "lm", formula= y~x, show.legend = TRUE)
```




```
ggplot(data = mpg) +
  geom_smooth(mapping = aes(x = displ, y = hwy, color = drv),
              show.legend = TRUE, method = "lm", formula= y~x)
```


hwy
30
25
20
15
10
2
3
4
5
6
7
displ
drv
4
f
r