RWorksheet_Siatan#6

Floreda Mae Siatan

2022-12-1

```
library(dplyr)
```

```
##
## 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
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
library(tinytex)
```

```
data(mpg)
data_set <- as.data.frame(mpg)
nrow(mpg)
```

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

```
ncol(mpg)
```

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

```
mostModel <- data_set %>% group_by(manufacturer) %>% count()
mostModel
```

```
## # A tibble: 15 x 2
## # Groups:   manufacturer [15]
##    manufacturer     n
##    <chr>        <int>
##  1 audi            18
```

```
##  2 chevrolet       19
##  3 dodge           37
##  4 ford            25
##  5 honda            9
##  6 hyundai         14
##  7 jeep             8
##  8 land rover       4
##  9 lincoln          3
## 10 mercury          4
## 11 nissan          13
## 12 pontiac          5
## 13 subaru          14
## 14 toyota          34
## 15 volkswagen      27
```

```
colnames(mostModel) <- c("Manufacturer","Counts")
mostModel
```

```
## # A tibble: 15 x 2
## # Groups:   Manufacturer [15]
##    Manufacturer Counts
##    <chr>         <int>
##  1 audi             18
##  2 chevrolet        19
##  3 dodge            37
##  4 ford             25
##  5 honda             9
##  6 hyundai          14
##  7 jeep              8
##  8 land rover        4
##  9 lincoln           3
## 10 mercury           4
## 11 nissan           13
## 12 pontiac           5
## 13 subaru           14
## 14 toyota           34
## 15 volkswagen       27
```

```
mostVariation<- data_set %>% group_by(model) %>% count()
mostVariation
```

```
## # A tibble: 38 x 2
## # Groups:   model [38]
##    model                  n
##    <chr>              <int>
##  1 4runner 4wd            6
##  2 a4                     7
##  3 a4 quattro             8
##  4 a6 quattro             3
##  5 altima                 6
##  6 c1500 suburban 2wd     5
##  7 camry                  7
##  8 camry solara           7
```

```
##  9 caravan 2wd           11
## 10 civic                  9
## # ... with 28 more rows
```

```
colnames(mostVariation) <- c("Model","Counts")
mostVariation
```

```
## # A tibble: 38 x 2
## # Groups:   Model [38]
##    Model              Counts
##    <chr>               <int>
##  1 4runner 4wd             6
##  2 a4                      7
##  3 a4 quattro              8
##  4 a6 quattro              3
##  5 altima                  6
##  6 c1500 suburban 2wd      5
##  7 camry                   7
##  8 camry solara            7
##  9 caravan 2wd            11
## 10 civic                   9
## # ... with 28 more rows
```

```
UniqueModel <- data_set %>% group_by(manufacturer, model) %>% distinct() %>% count()
UniqueModel
```

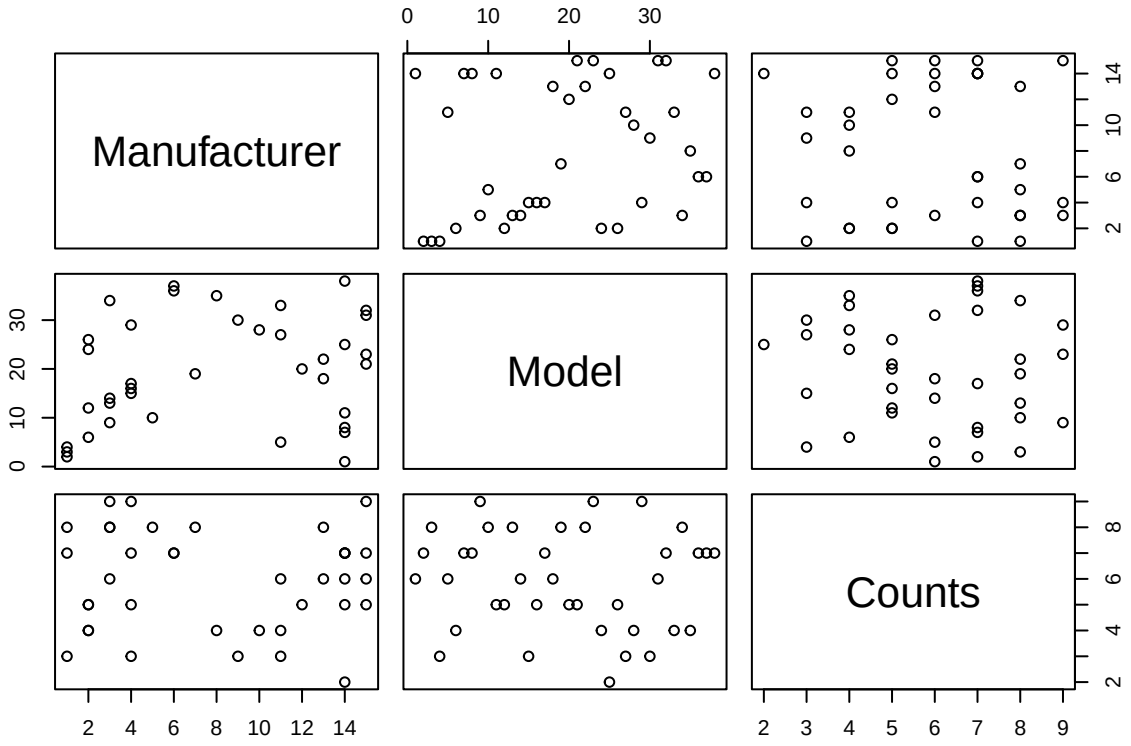
```
## # A tibble: 38 x 3
## # Groups:   manufacturer, model [38]
##    manufacturer model                 n
##    <chr>        <chr>             <int>
##  1 audi         a4                    7
##  2 audi         a4 quattro            8
##  3 audi         a6 quattro            3
##  4 chevrolet    c1500 suburban 2wd    4
##  5 chevrolet    corvette              5
##  6 chevrolet    k1500 tahoe 4wd       4
##  7 chevrolet    malibu                5
##  8 dodge        caravan 2wd           9
##  9 dodge        dakota pickup 4wd     8
## 10 dodge        durango 4wd           6
## # ... with 28 more rows
```

```
colnames(UniqueModel) <- c("Manufacturer", "Model","Counts")
UniqueModel
```

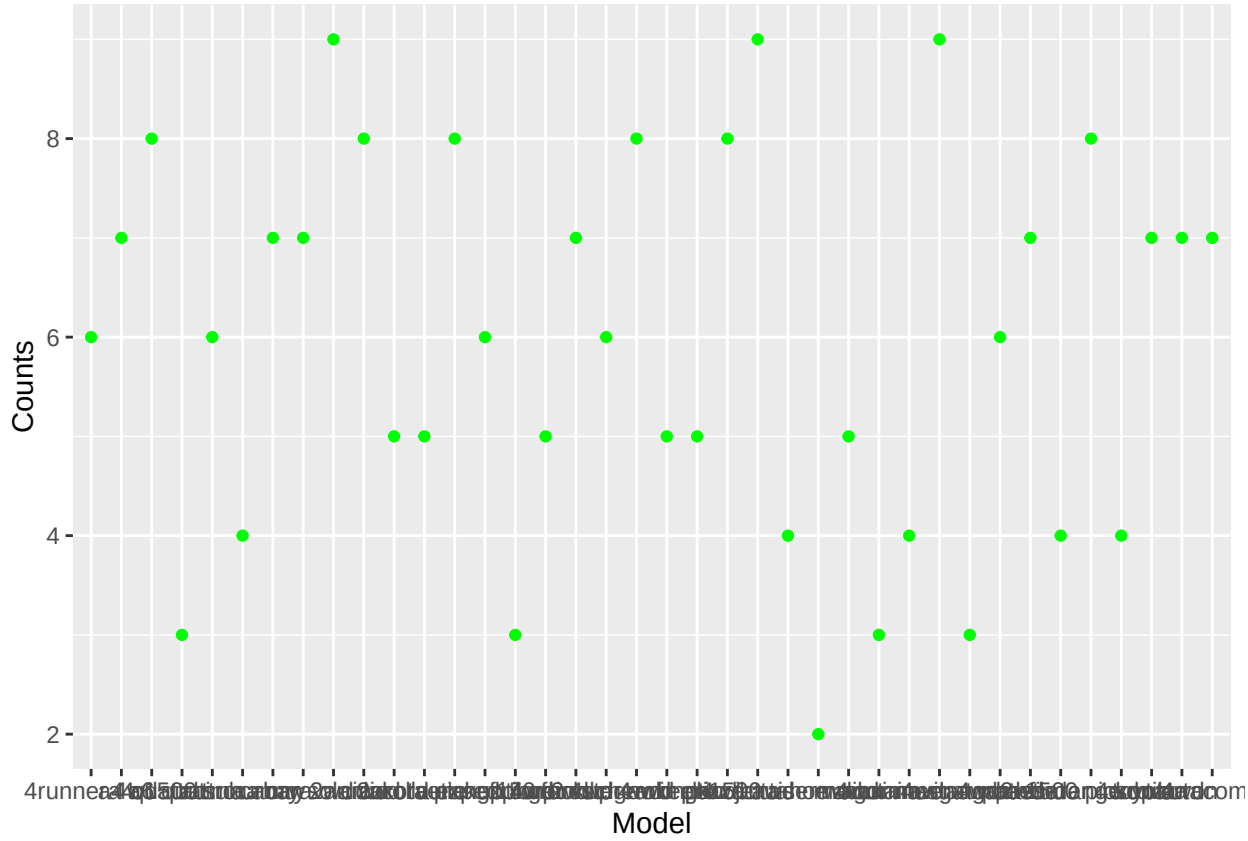
```
## # A tibble: 38 x 3
## # Groups:   Manufacturer, Model [38]
##    Manufacturer Model              Counts
##    <chr>        <chr>               <int>
##  1 audi         a4                      7
##  2 audi         a4 quattro              8
##  3 audi         a6 quattro              3
```

```
##  4 chevrolet    c1500 suburban 2wd      4
##  5 chevrolet    corvette                5
##  6 chevrolet    k1500 tahoe 4wd         4
##  7 chevrolet    malibu                  5
##  8 dodge        caravan 2wd             9
##  9 dodge        dakota pickup 4wd       8
## 10 dodge        durango 4wd             6
## # ... with 28 more rows
```

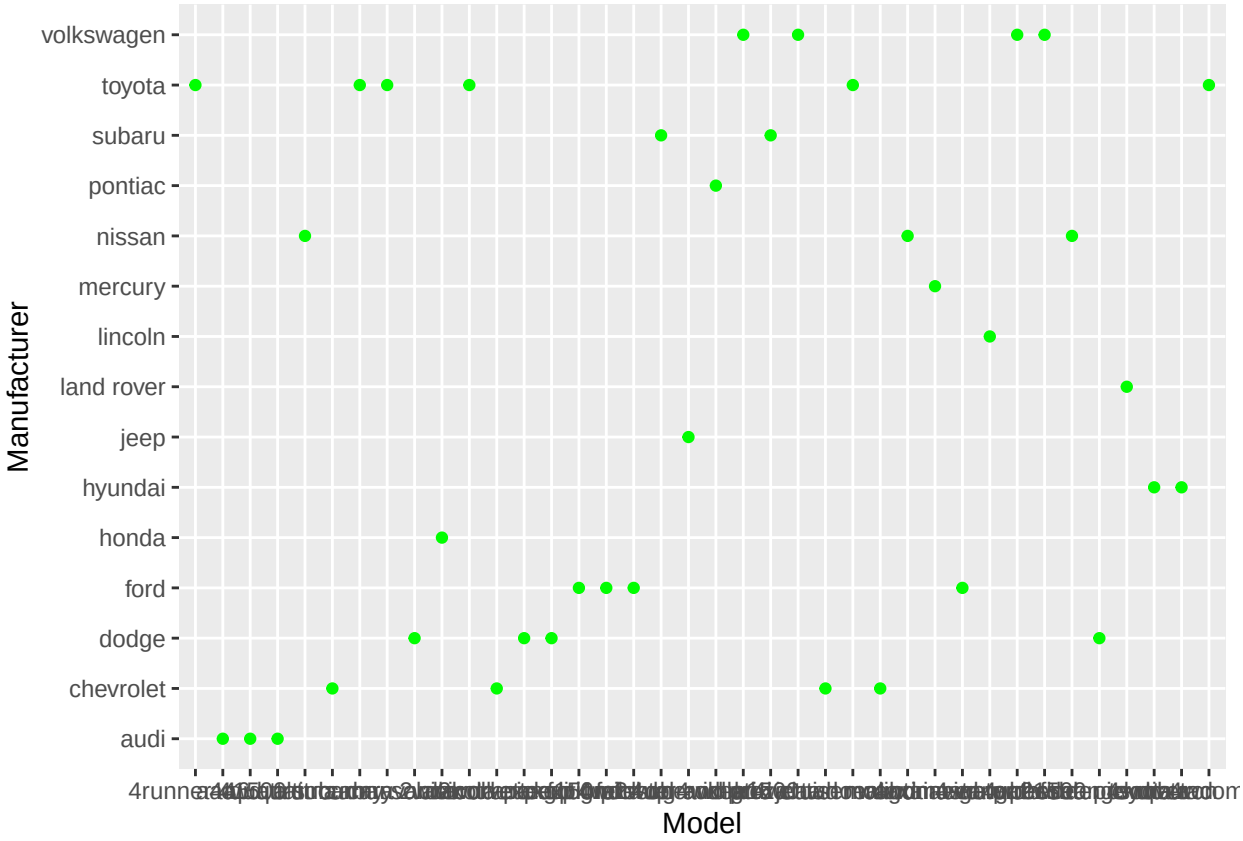
```
plot(UniqueModel)
```



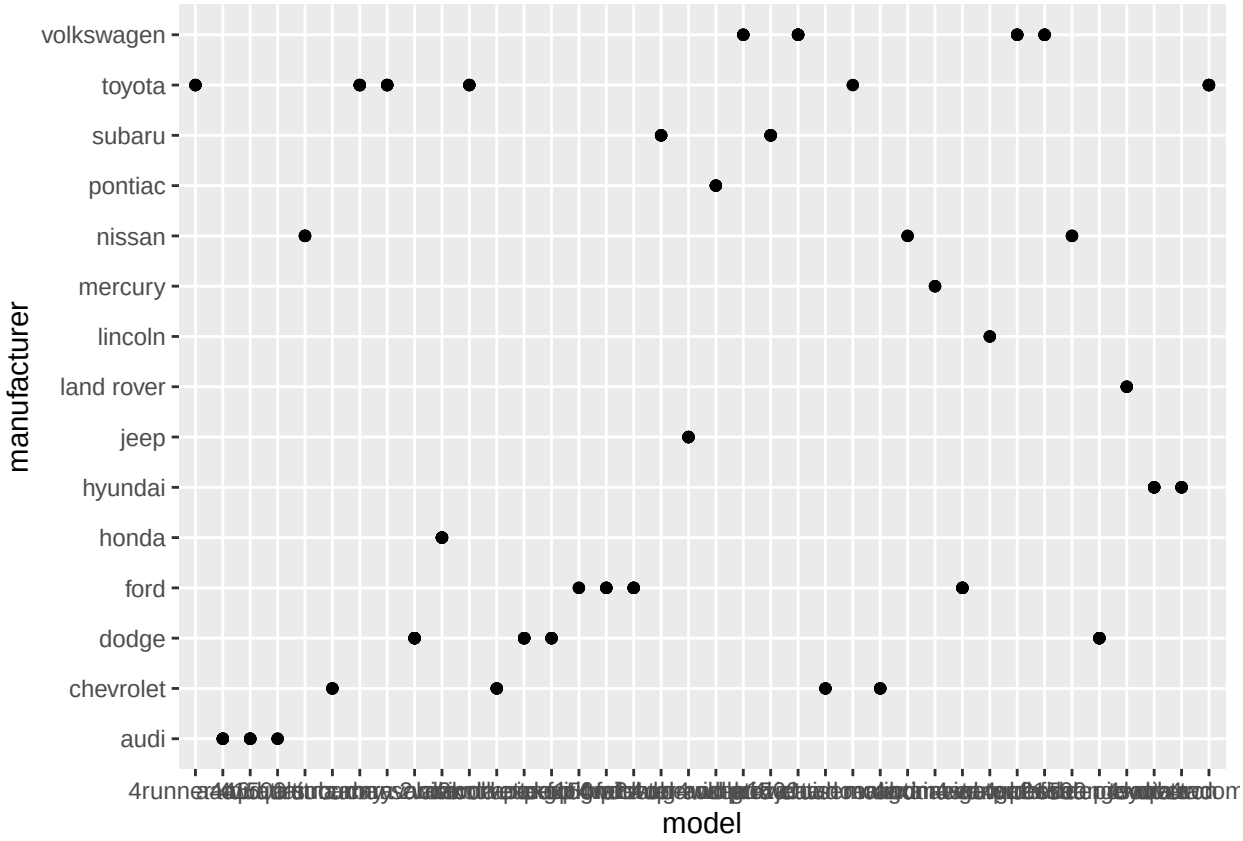

```
ggplot(UniqueModel, aes(x = Model, y = Counts )) + geom_point(color='green')
```




```
ggplot(UniqueModel, aes(x = Model, y = Manufacturer )) + geom_point(color='green')
```




```
ggplot(mpg, aes(model, manufacturer)) + geom_point()
```




```
carsmodel <- data_set %>% group_by(model) %>% count()
carsmodel
```

```
## # A tibble: 38 x 2
## # Groups:   model [38]
##    model                  n
##    <chr>              <int>
##  1 4runner 4wd            6
##  2 a4                     7
##  3 a4 quattro             8
##  4 a6 quattro             3
##  5 altima                 6
##  6 c1500 suburban 2wd     5
##  7 camry                  7
##  8 camry solara           7
##  9 caravan 2wd           11
## 10 civic                  9
## # ... with 28 more rows
```

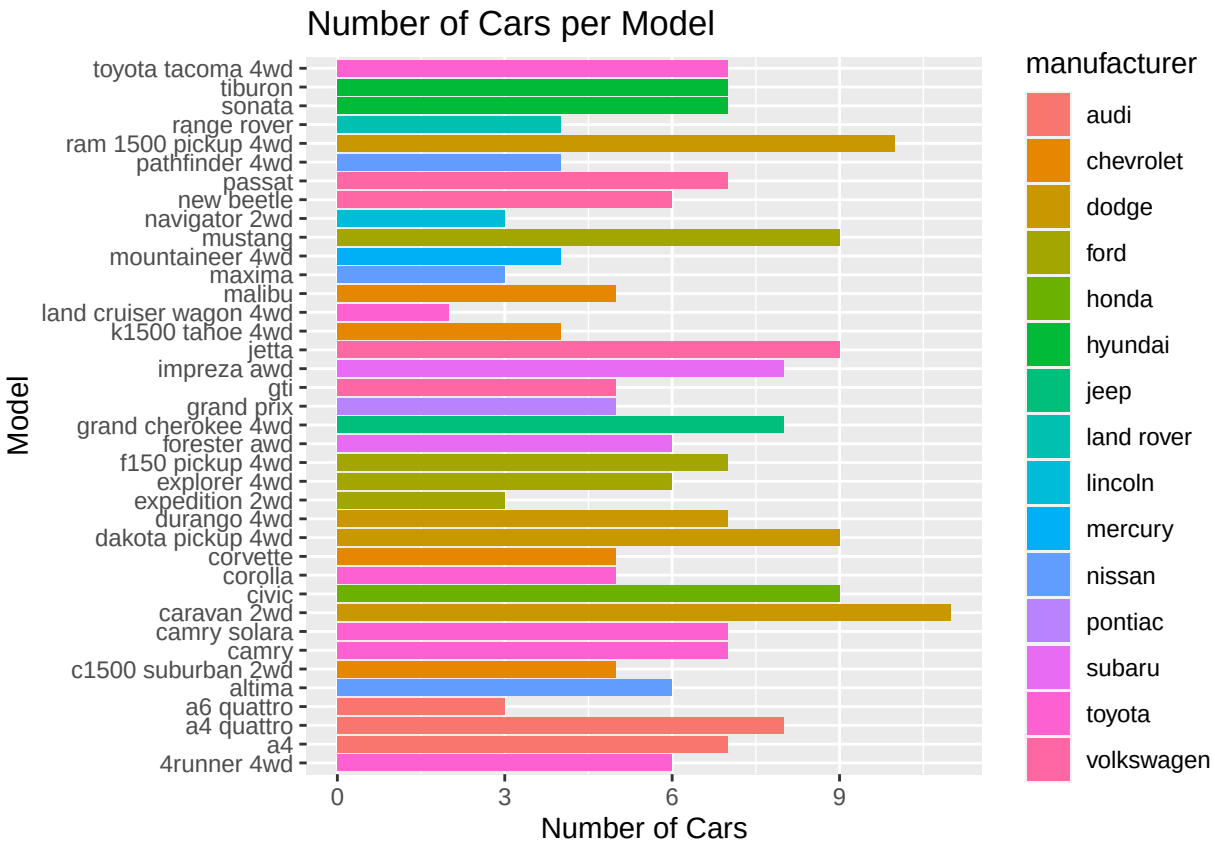
```
colnames(carsmodel) <- c("Model","Counts")
carsmodel
```

```
## # A tibble: 38 x 2
## # Groups:   Model [38]
##    Model              Counts
```

```
##    <chr>               <int>
##  1 4runner 4wd             6
##  2 a4                      7
##  3 a4 quattro              8
##  4 a6 quattro              3
##  5 altima                  6
##  6 c1500 suburban 2wd      5
##  7 camry                   7
##  8 camry solara            7
##  9 caravan 2wd            11
## 10 civic                   9
## # ... with 28 more rows
```

```
qplot(model,data = mpg,main = "Number of Cars per Model", xlab = "Model",
      ylab = "Number of Cars", geom = "bar", fill = manufacturer) +
  coord_flip()
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
```



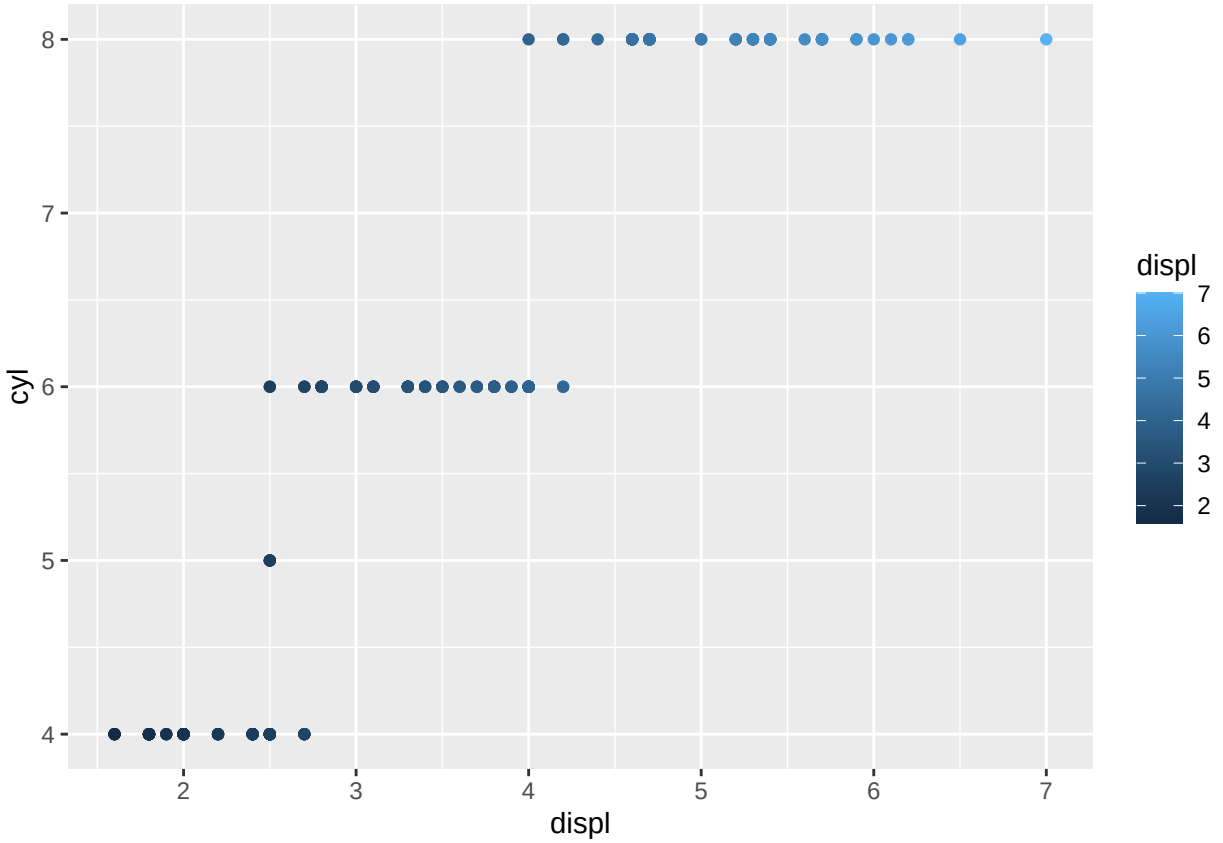

```
del <- carsmodel[1:20,] %>% top_n(2)
```

```
## Selecting by Counts
```

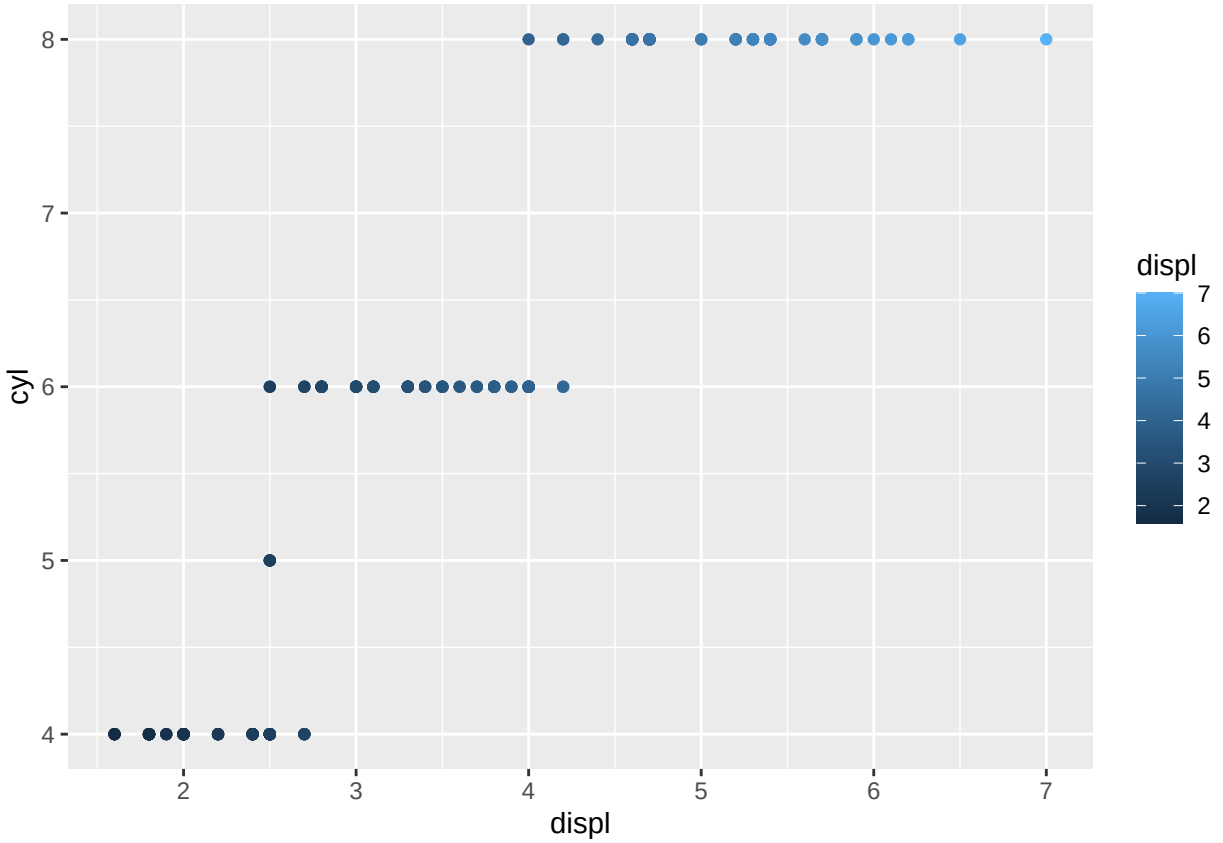
```
del
```

```
## # A tibble: 20 x 2
## # Groups:   Model [20]
##    Model              Counts
##    <chr>               <int>
##  1 4runner 4wd             6
##  2 a4                      7
##  3 a4 quattro              8
##  4 a6 quattro              3
##  5 altima                  6
##  6 c1500 suburban 2wd      5
##  7 camry                   7
##  8 camry solara            7
##  9 caravan 2wd            11
## 10 civic                   9
## 11 corolla                 5
## 12 corvette                5
## 13 dakota pickup 4wd       9
## 14 durango 4wd             7
## 15 expedition 2wd          3
## 16 explorer 4wd            6
## 17 f150 pickup 4wd         7
## 18 forester awd            6
## 19 grand cherokee 4wd      8
## 20 grand prix              5
```

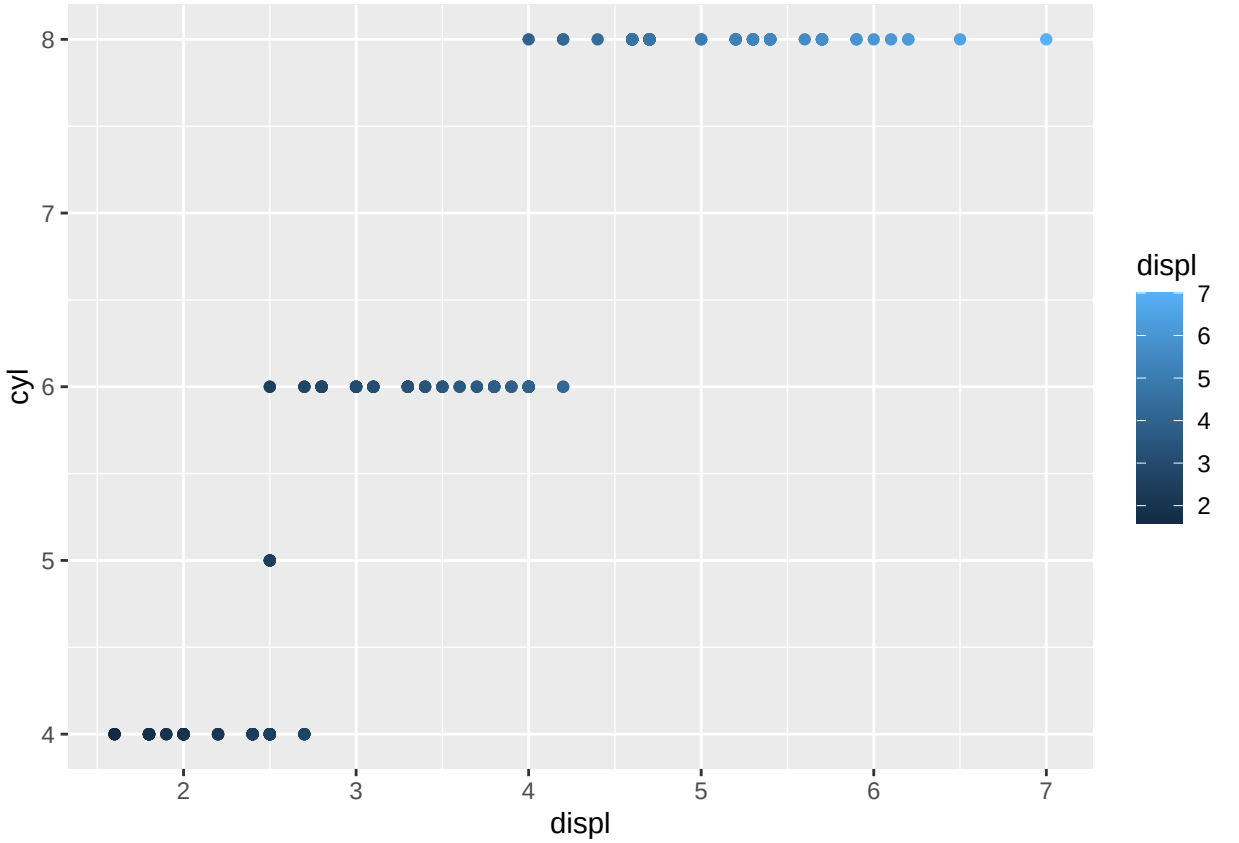
```
ggplot(mpg, aes(x = displ , y = cyl, col = displ )) + geom_point()
```




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




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




```
wheel_drive <- subset(mpg, drv == 'f')
wheel_drive <- nrow(wheel_drive)
wheel_drive
```

```
## [1] 106
```

```
rearwheel <- subset(mpg, drv == 'r')
nrow(rearwheel)
```

```
## [1] 25
```

```
rearwheel
```

```
## # A tibble: 25 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 chevrolet    c1500 sub~   5.3  2008     8 auto~ r        14    20 r     suv
##  2 chevrolet    c1500 sub~   5.3  2008     8 auto~ r        11    15 e     suv
##  3 chevrolet    c1500 sub~   5.3  2008     8 auto~ r        14    20 r     suv
##  4 chevrolet    c1500 sub~   5.7  1999     8 auto~ r        13    17 r     suv
##  5 chevrolet    c1500 sub~   6    2008     8 auto~ r        12    17 r     suv
##  6 chevrolet    corvette     5.7  1999     8 manu~ r        16    26 p     2sea~
##  7 chevrolet    corvette     5.7  1999     8 auto~ r        15    23 p     2sea~
##  8 chevrolet    corvette     6.2  2008     8 manu~ r        16    26 p     2sea~
```

```
##  9 chevrolet    corvette     6.2  2008     8 auto~ r        15    25 p     2sea~
## 10 chevrolet    corvette     7    2008     8 manu~ r        15    24 p     2sea~
## # ... with 15 more rows
```

```
num4 <- subset(mpg, drv == '4')
nrow(num4)
```

```
## [1] 103
```

```
num4
```

```
## # A tibble: 103 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 quattro   1.8  1999     4 manu~ 4        18    26 p     comp~
##  2 audi         a4 quattro   1.8  1999     4 auto~ 4        16    25 p     comp~
##  3 audi         a4 quattro   2    2008     4 manu~ 4        20    28 p     comp~
##  4 audi         a4 quattro   2    2008     4 auto~ 4        19    27 p     comp~
##  5 audi         a4 quattro   2.8  1999     6 auto~ 4        15    25 p     comp~
##  6 audi         a4 quattro   2.8  1999     6 manu~ 4        17    25 p     comp~
##  7 audi         a4 quattro   3.1  2008     6 auto~ 4        17    25 p     comp~
##  8 audi         a4 quattro   3.1  2008     6 manu~ 4        15    25 p     comp~
##  9 audi         a6 quattro   2.8  1999     6 auto~ 4        15    24 p     mids~
## 10 audi         a6 quattro   3.1  2008     6 auto~ 4        17    25 p     mids~
## # ... with 93 more rows
```

```
c <- subset(mpg, class == 'compact')
nrow(c)
```

```
## [1] 47
```

```
c
```

```
## # A tibble: 47 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~ f        18    29 p     comp~
##  2 audi         a4           1.8  1999     4 manu~ f        21    29 p     comp~
##  3 audi         a4           2    2008     4 manu~ f        20    31 p     comp~
##  4 audi         a4           2    2008     4 auto~ f        21    30 p     comp~
##  5 audi         a4           2.8  1999     6 auto~ f        16    26 p     comp~
##  6 audi         a4           2.8  1999     6 manu~ f        18    26 p     comp~
##  7 audi         a4           3.1  2008     6 auto~ f        18    27 p     comp~
##  8 audi         a4 quattro   1.8  1999     4 manu~ 4        18    26 p     comp~
##  9 audi         a4 quattro   1.8  1999     4 auto~ 4        16    25 p     comp~
## 10 audi         a4 quattro   2    2008     4 manu~ 4        20    28 p     comp~
## # ... with 37 more rows
```

```
m_size <- subset(mpg, class == 'midsize')

nrow(m_size)
```

```
## [1] 41
```

```
m_size
```

```
## # A tibble: 41 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         a6 quattro   2.8  1999     6 auto~ 4        15    24 p     mids~
##  2 audi         a6 quattro   3.1  2008     6 auto~ 4        17    25 p     mids~
##  3 audi         a6 quattro   4.2  2008     8 auto~ 4        16    23 p     mids~
##  4 chevrolet    malibu       2.4  1999     4 auto~ f        19    27 r     mids~
##  5 chevrolet    malibu       2.4  2008     4 auto~ f        22    30 r     mids~
##  6 chevrolet    malibu       3.1  1999     6 auto~ f        18    26 r     mids~
##  7 chevrolet    malibu       3.5  2008     6 auto~ f        18    29 r     mids~
##  8 chevrolet    malibu       3.6  2008     6 auto~ f        17    26 r     mids~
##  9 hyundai      sonata       2.4  1999     4 auto~ f        18    26 r     mids~
## 10 hyundai      sonata       2.4  1999     4 manu~ f        18    27 r     mids~
## # ... with 31 more rows
```

```
two_seater <- subset(mpg, class == '2seater')
nrow(two_seater)
```

```
## [1] 5
```

```
two_seater
```

```
## # A tibble: 5 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 chevrolet    corvette   5.7  1999     8 manual(~ r        16    26 p     2sea~
## 2 chevrolet    corvette   5.7  1999     8 auto(l4) r        15    23 p     2sea~
## 3 chevrolet    corvette   6.2  2008     8 manual(~ r        16    26 p     2sea~
## 4 chevrolet    corvette   6.2  2008     8 auto(s6) r        15    25 p     2sea~
## 5 chevrolet    corvette   7    2008     8 manual(~ r        15    24 p     2sea~
```

```
mini_van <- subset(mpg, class == 'minivan')
nrow(mini_van)
```

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

```
mini_van
```

```
## # A tibble: 11 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 dodge        caravan 2~   2.4  1999     4 auto~ f        18    24 r     mini~
##  2 dodge        caravan 2~   3    1999     6 auto~ f        17    24 r     mini~
##  3 dodge        caravan 2~   3.3  1999     6 auto~ f        16    22 r     mini~
##  4 dodge        caravan 2~   3.3  1999     6 auto~ f        16    22 r     mini~
##  5 dodge        caravan 2~   3.3  2008     6 auto~ f        17    24 r     mini~
##  6 dodge        caravan 2~   3.3  2008     6 auto~ f        17    24 r     mini~
```

```
##  7 dodge        caravan 2~   3.3  2008     6 auto~ f        11    17 e     mini~
##  8 dodge        caravan 2~   3.8  1999     6 auto~ f        15    22 r     mini~
##  9 dodge        caravan 2~   3.8  1999     6 auto~ f        15    21 r     mini~
## 10 dodge        caravan 2~   3.8  2008     6 auto~ f        16    23 r     mini~
## 11 dodge        caravan 2~   4    2008     6 auto~ f        16    23 r     mini~
```

```
p_u <- subset(mpg, class == 'pickup')
nrow(p_u)
```

```
## [1] 33
```

```
p_u
```

```
## # A tibble: 33 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 dodge        dakota pi~   3.7  2008     6 manu~ 4        15    19 r     pick~
##  2 dodge        dakota pi~   3.7  2008     6 auto~ 4        14    18 r     pick~
##  3 dodge        dakota pi~   3.9  1999     6 auto~ 4        13    17 r     pick~
##  4 dodge        dakota pi~   3.9  1999     6 manu~ 4        14    17 r     pick~
##  5 dodge        dakota pi~   4.7  2008     8 auto~ 4        14    19 r     pick~
##  6 dodge        dakota pi~   4.7  2008     8 auto~ 4        14    19 r     pick~
##  7 dodge        dakota pi~   4.7  2008     8 auto~ 4         9    12 e     pick~
##  8 dodge        dakota pi~   5.2  1999     8 manu~ 4        11    17 r     pick~
##  9 dodge        dakota pi~   5.2  1999     8 auto~ 4        11    15 r     pick~
## 10 dodge        ram 1500 ~   4.7  2008     8 manu~ 4        12    16 r     pick~
## # ... with 23 more rows
```

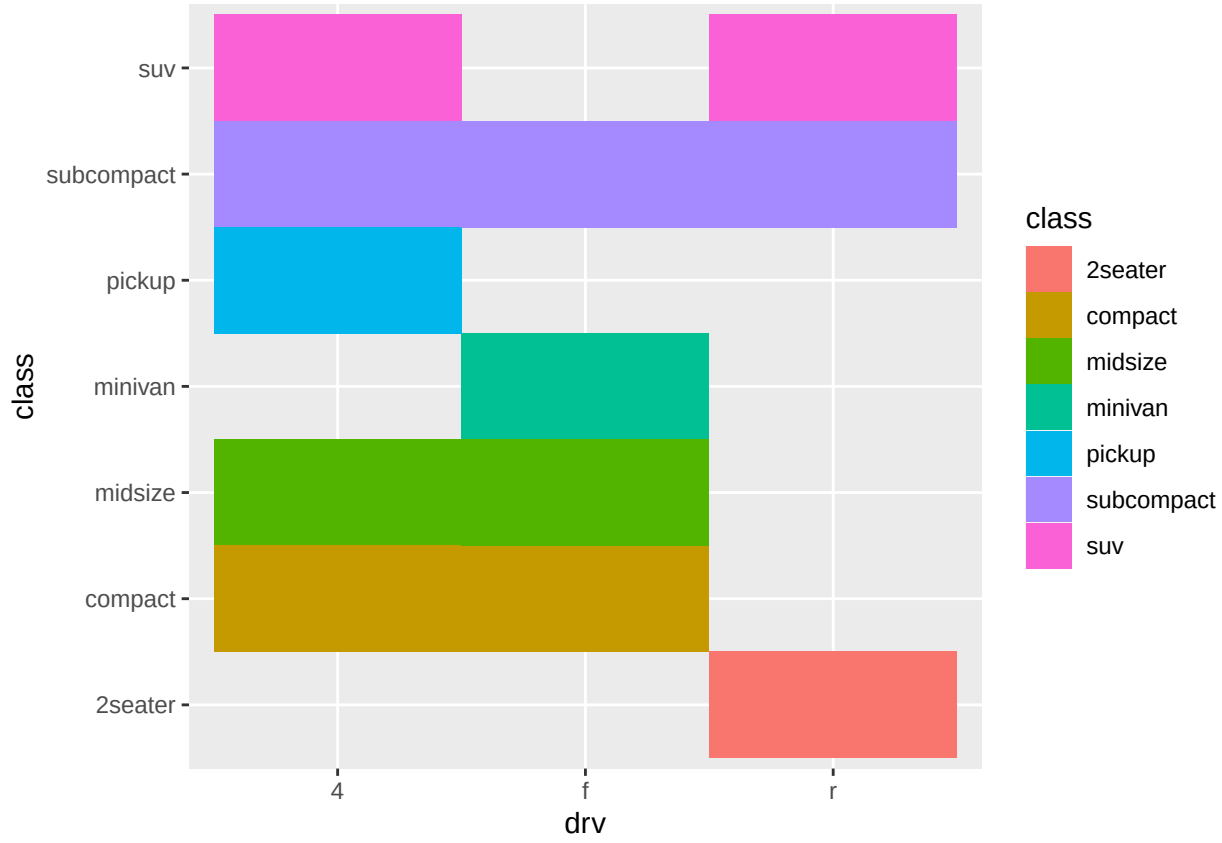
```
sub_comp <- subset(mpg, class == 'subcompact')
nrow(sub_comp)
```

```
## [1] 35
```

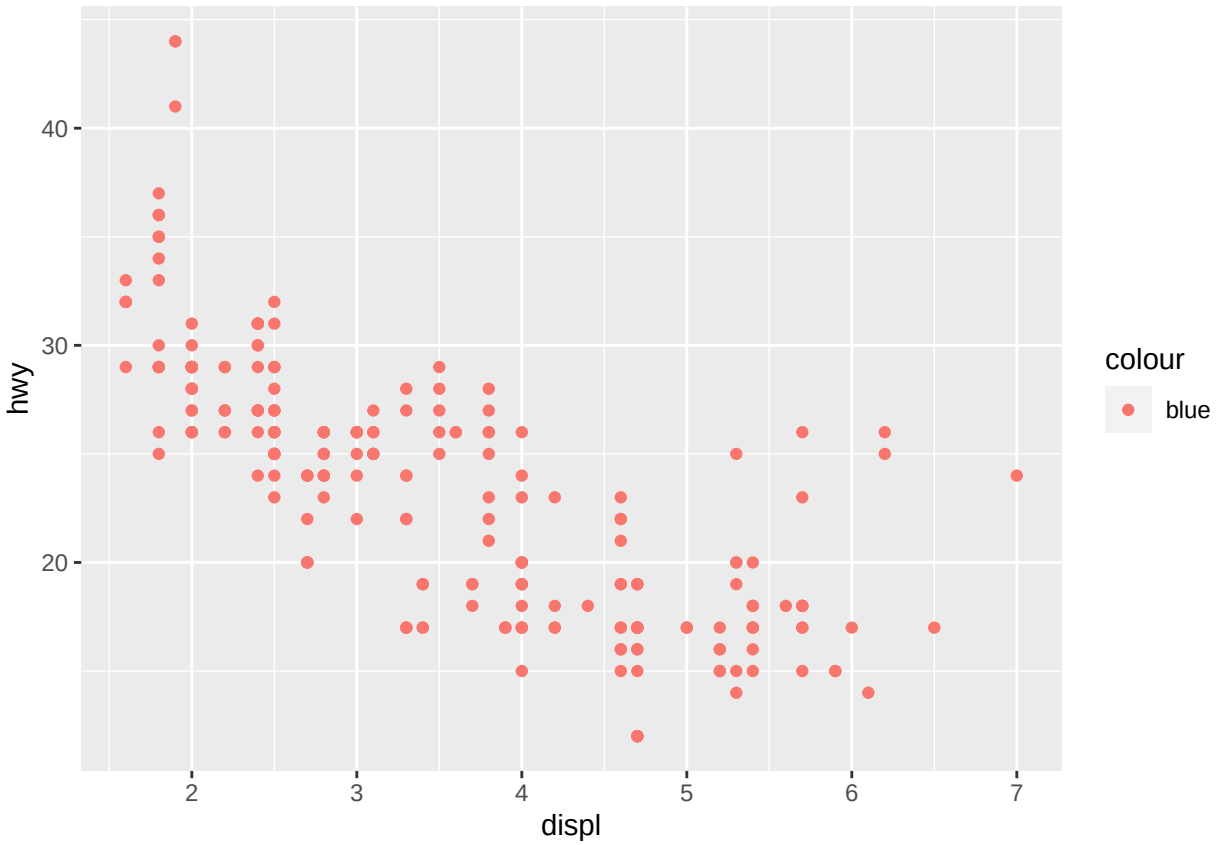
```
sub_comp
```

```
## # A tibble: 35 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 ford         mustang   3.8  1999     6 manual(~ r        18    26 r     subc~
##  2 ford         mustang   3.8  1999     6 auto(l4) r        18    25 r     subc~
##  3 ford         mustang   4    2008     6 manual(~ r        17    26 r     subc~
##  4 ford         mustang   4    2008     6 auto(l5) r        16    24 r     subc~
##  5 ford         mustang   4.6  1999     8 auto(l4) r        15    21 r     subc~
##  6 ford         mustang   4.6  1999     8 manual(~ r        15    22 r     subc~
##  7 ford         mustang   4.6  2008     8 manual(~ r        15    23 r     subc~
##  8 ford         mustang   4.6  2008     8 auto(l5) r        15    22 r     subc~
##  9 ford         mustang   5.4  2008     8 manual(~ r        14    20 p     subc~
## 10 honda        civic     1.6  1999     4 manual(~ f        28    33 r     subc~
## # ... with 25 more rows
```

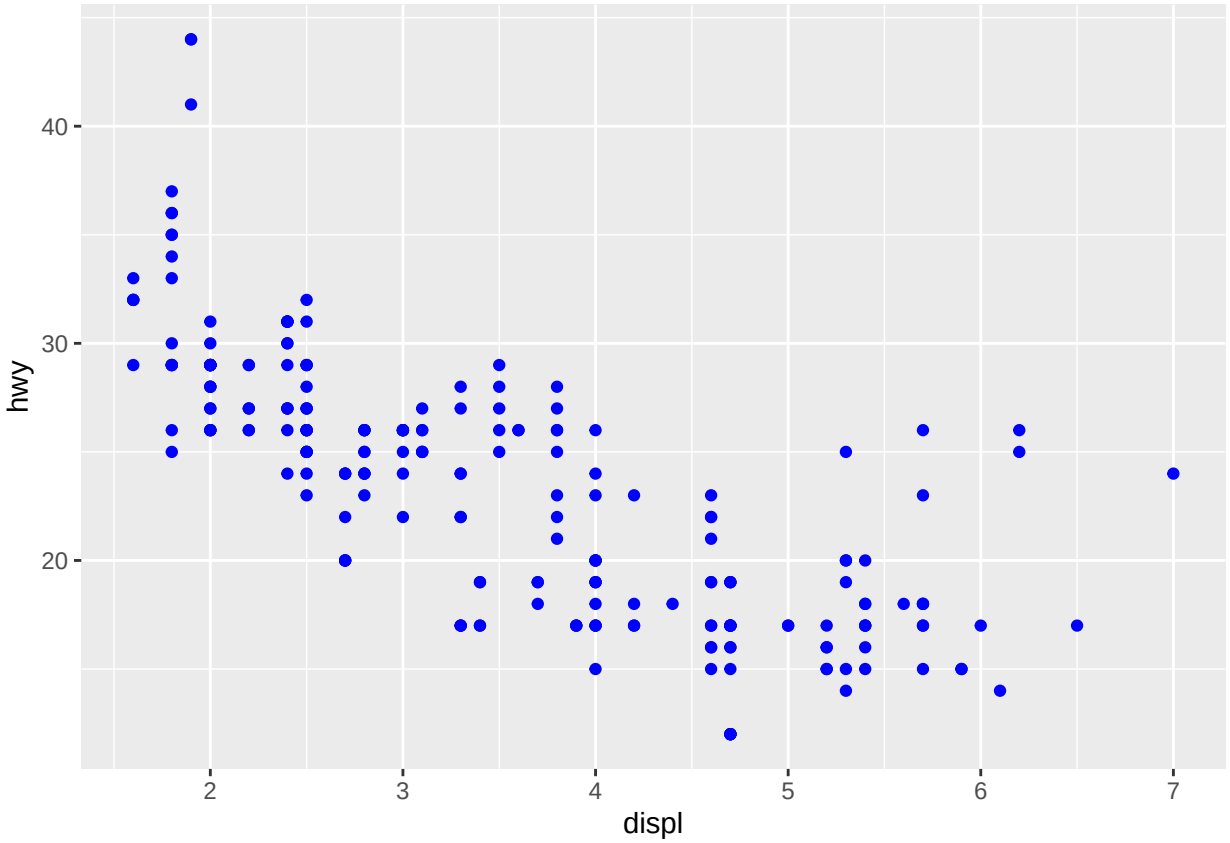
```
ggplot(mpg, aes(drv, class)) +
  geom_tile (aes(fill = class))
```



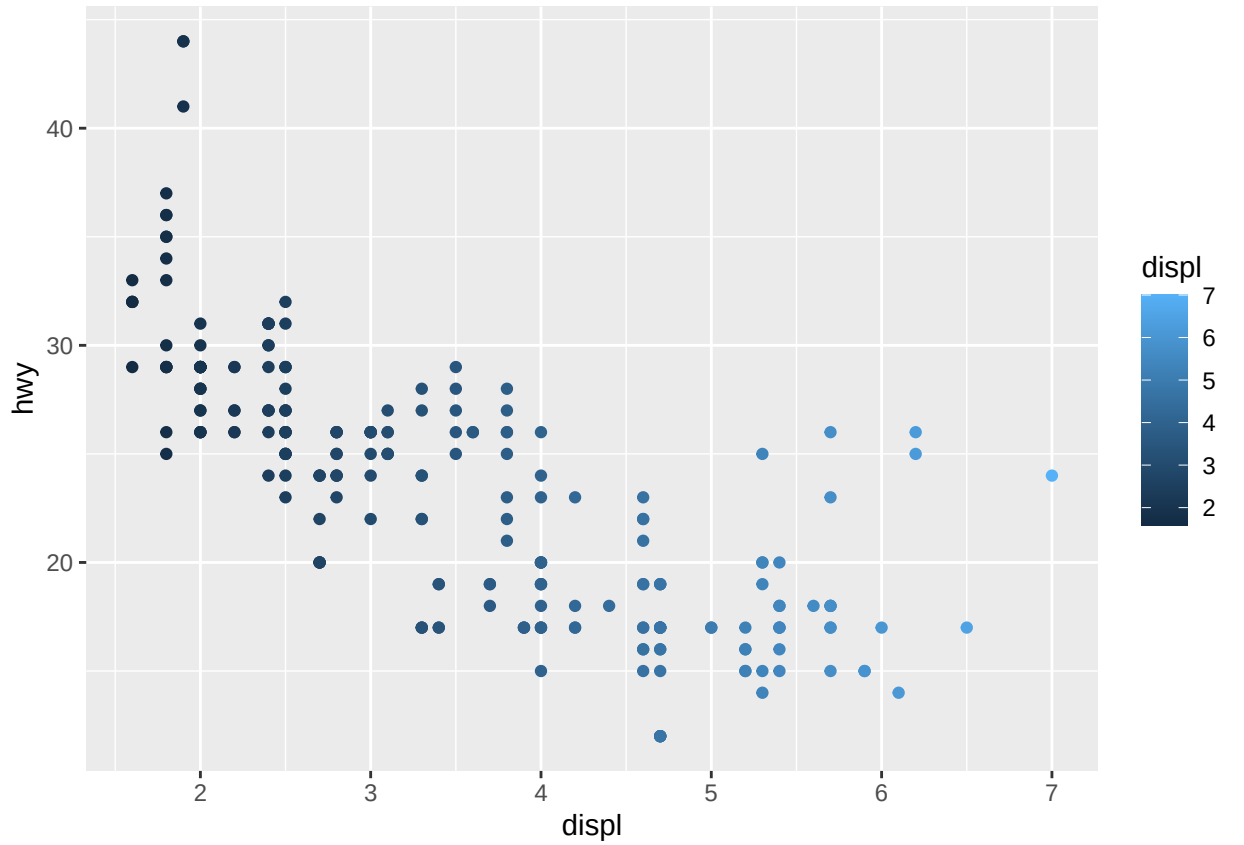

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, colour = "blue"))
```




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

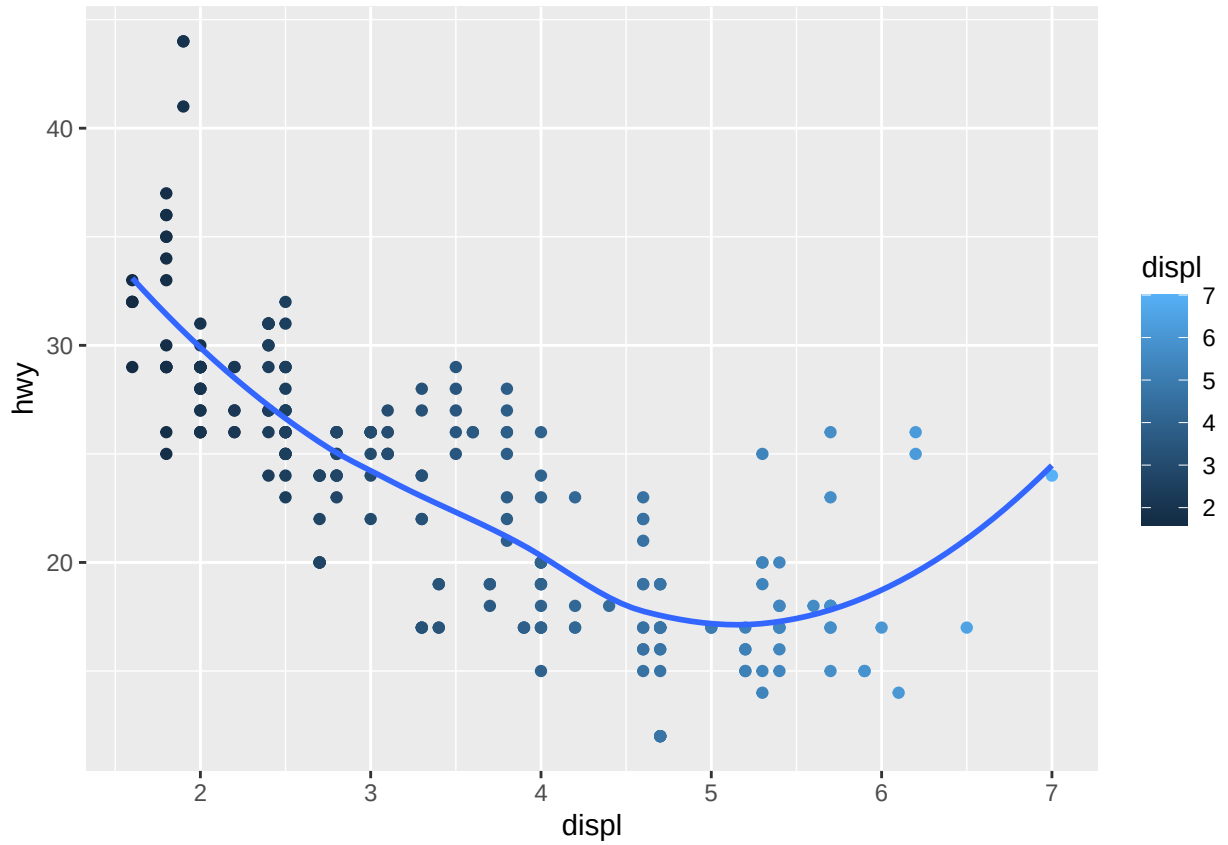



```
ggplot( data = mpg) +
  geom_point(mapping = aes(x = displ , y = hwy, col = displ))
```



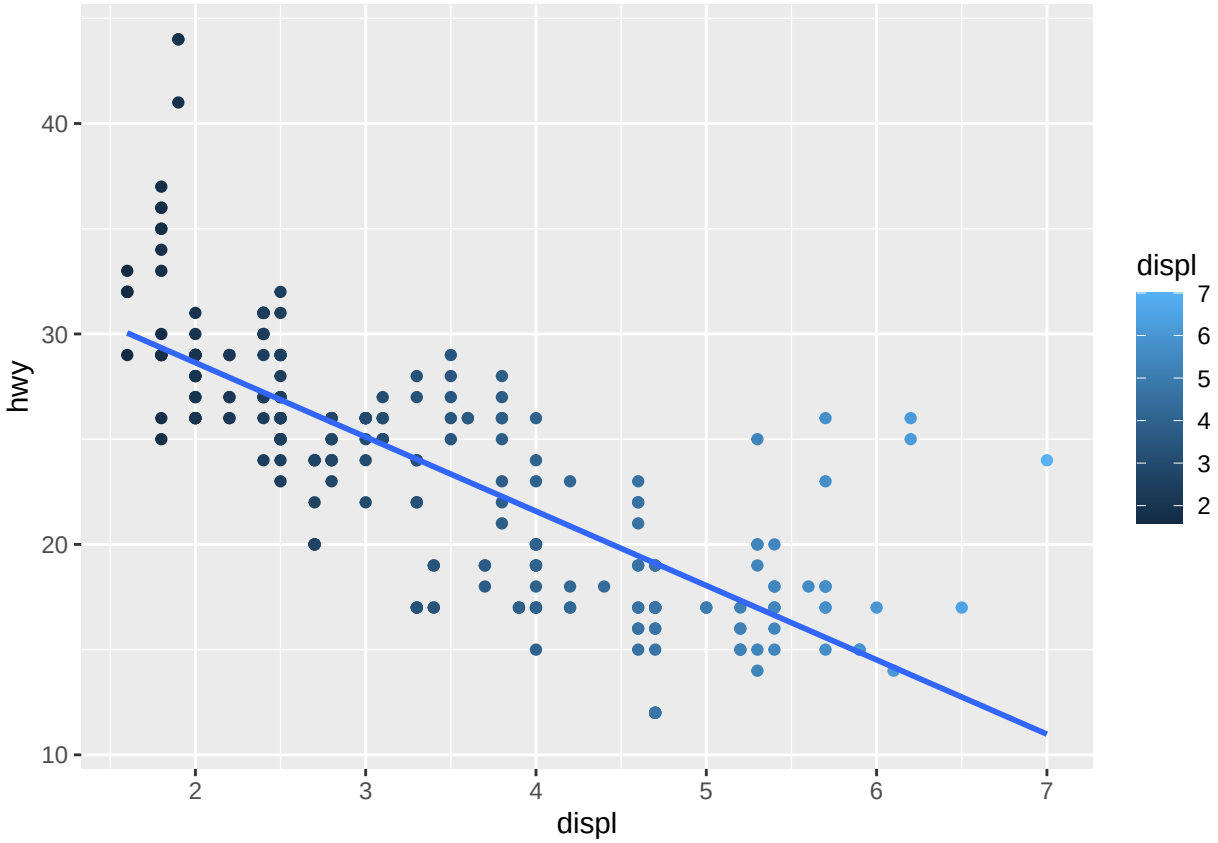

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_point(mapping=aes(color=displ)) +
  geom_smooth(se =FALSE)
```

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




```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_point(mapping=aes(color=displ)) +
  geom_smooth(se =FALSE,method = lm)
```

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


hwy
40
30
20
10
2
3
4
5
6
7
displ
displ
7
6
5
4
3
2