Lab 1

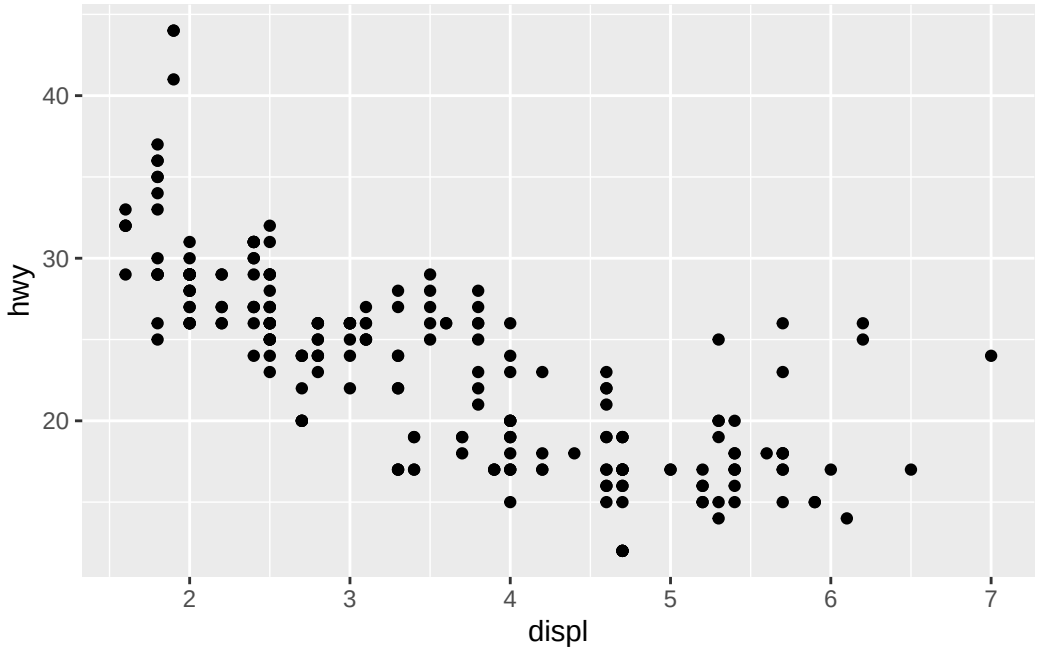
Zahlen Zbinden

2024-04-06

```
library(ggplot2)
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.3     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v lubridate 1.9.2     v tibble    3.2.1
v purrr     1.0.2     v tidyr     1.3.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
mpg |>
  ggplot(aes(x=displ, y=hwy)) +
    geom_point()
```



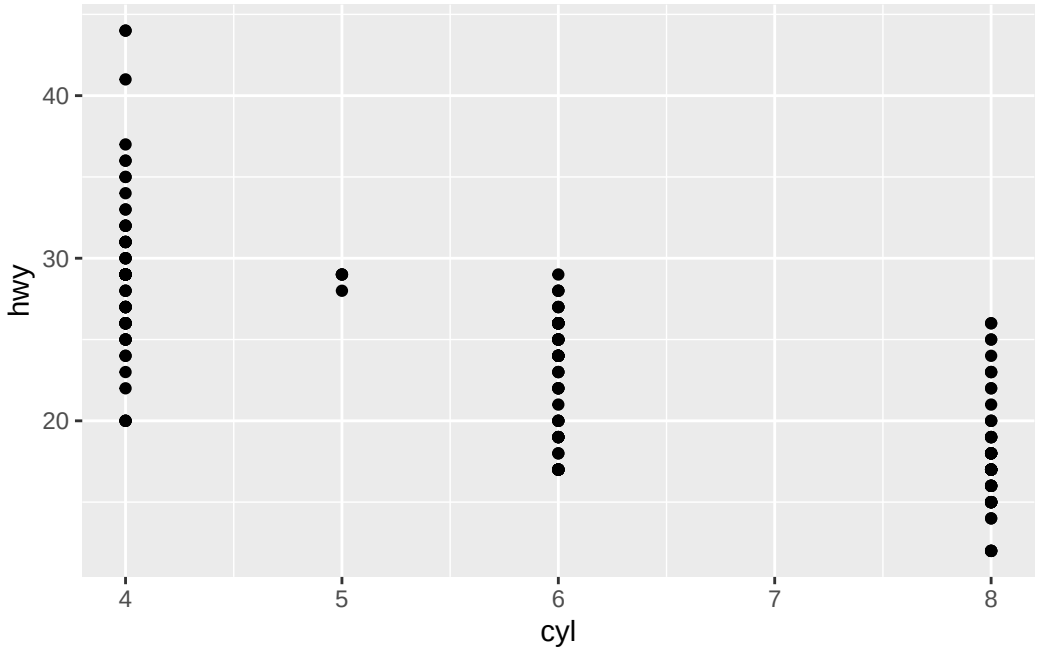

```
ggplot(data=mpg)
```

```r
str(mpg)
```

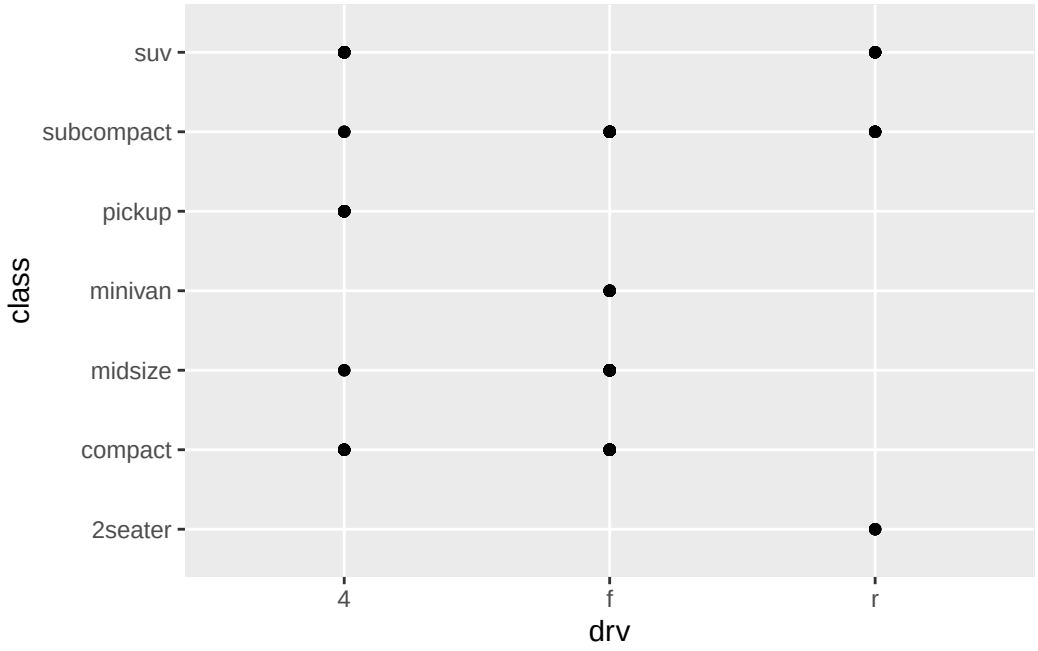
```
tibble [234 x 11] (S3: tbl_df/tbl/data.frame)
 $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
 $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
 $ displ       : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
 $ year        : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 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" ...
```

1. A blank graph
2. 234 x 11
3. Categorical, type of drive train. front-wheel, rear-wheel, 4wd
4. find below
5. It just shows all the unique pairs of class and drive. It doesn't have any numerical values that would lead to a deeper understanding of the data.

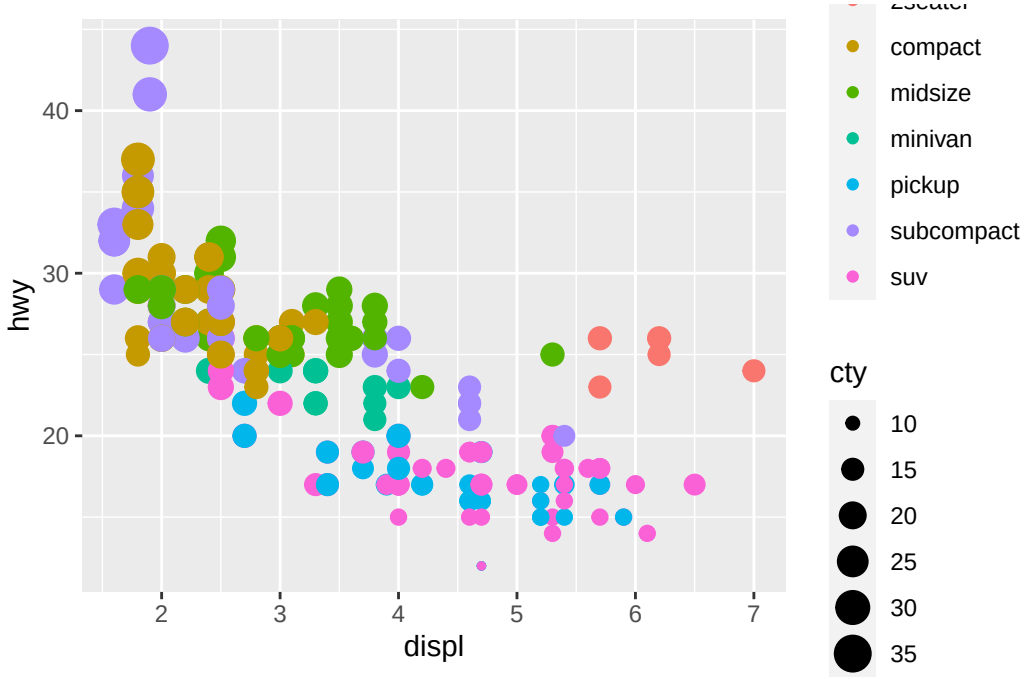
```r
mpg |>
  ggplot() +
    geom_point(aes(x=cyl, y=hwy))
```




```
mpg |>
  ggplot() +
    geom_point(aes(x=drv, y=class))
```

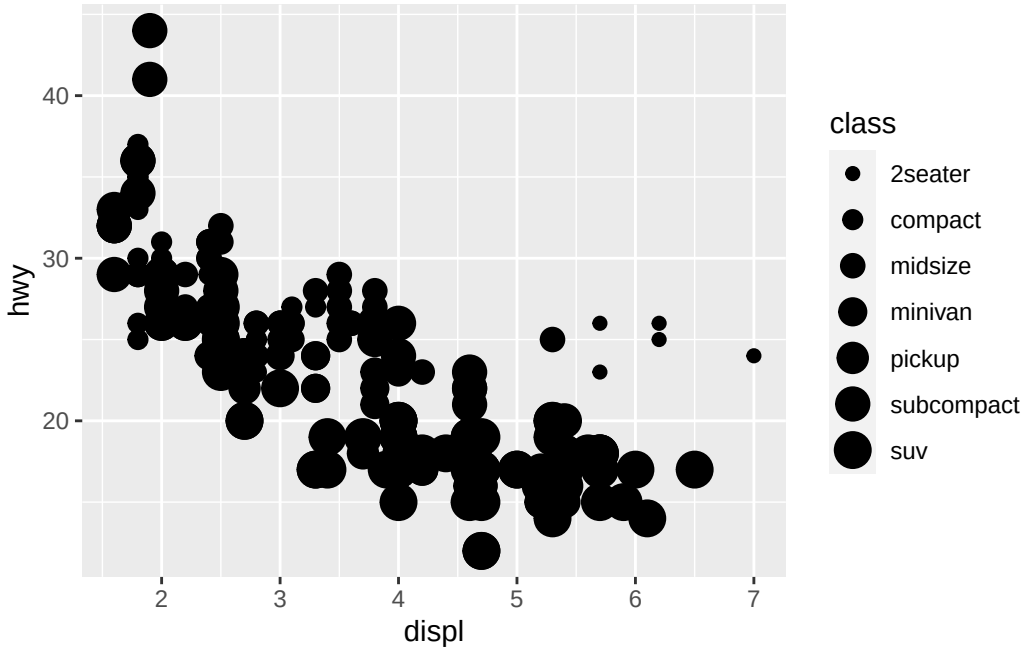



```
mpg |>
  ggplot() +
    geom_point(aes(x=displ, y=hwy, color=class, size=cty))
```




```
mpg |>
  ggplot() +
    geom_point(aes(x=displ, y=hwy, size=class))
```

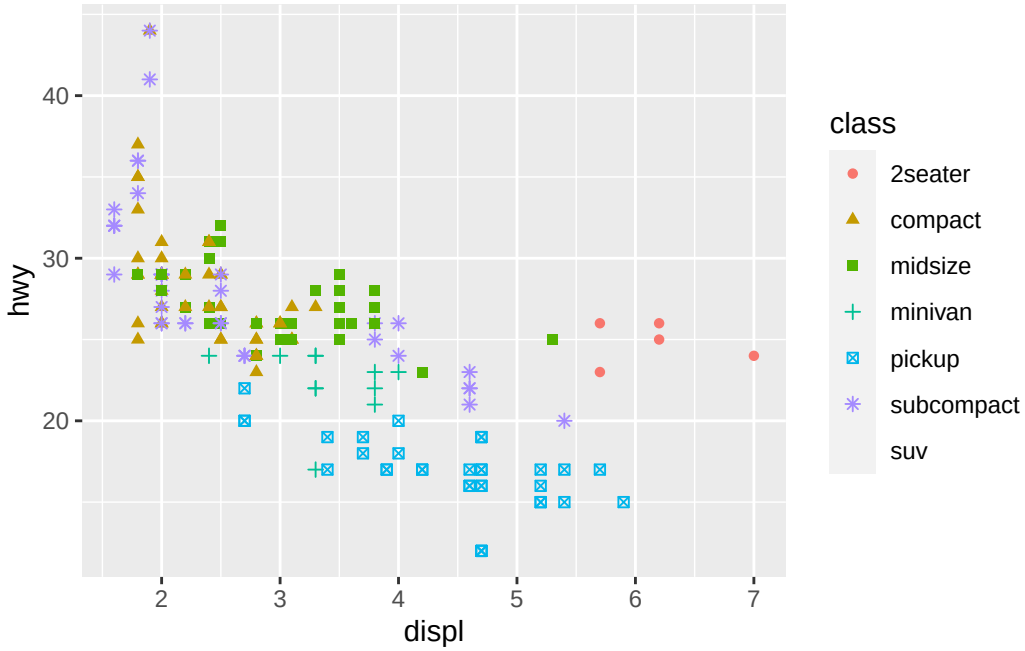
```
Warning: Using size for a discrete variable is not advised.
```




```
mpg |>
  ggplot() +
    geom_point(aes(x=displ, y=hwy, color=class, shape=class))
```

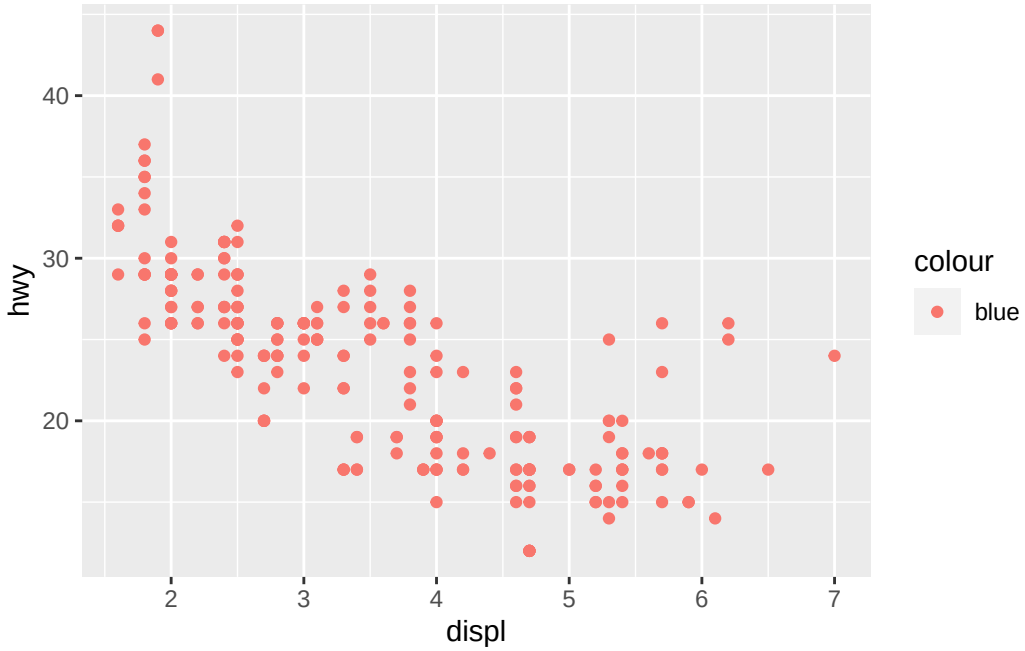
```
Warning: The shape palette can deal with a maximum of 6 discrete values because
more than 6 becomes difficult to discriminate; you have 7. Consider
specifying shapes manually if you must have them.
```

```
Warning: Removed 62 rows containing missing values (`geom_point()`).
```



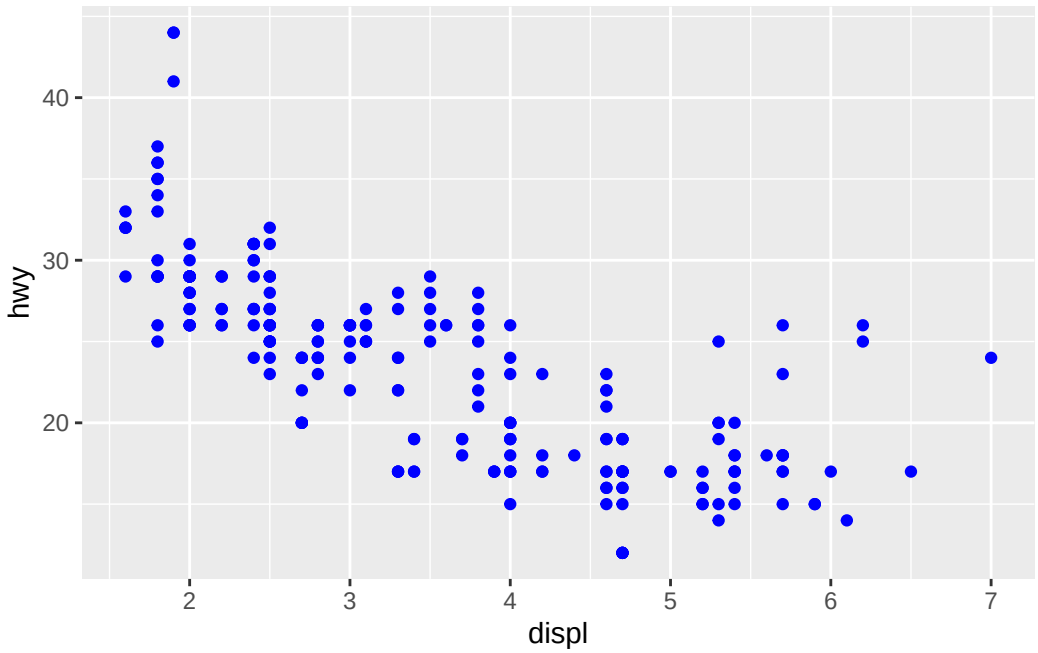
Notice that aes is a mapping, and mapping are displayed in the legend

```
mpg |>
  ggplot() +
    geom_point(aes(x=displ, y=hwy, color="blue"))
```



opposed to

```
mpg |>
  ggplot() +
    geom_point(aes(x=displ, y=hwy), color="blue")
```




```
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" ...
```

```
sapply(mpg, function(x) length(unique(x)))
```

```
manufacturer        model        displ         year          cyl        trans
          15           38           35            2            4           10
         drv          cty          hwy           fl        class
           3           21           27            5            7
```

1. the points are not blue because we mapped color to the constant "blue", we didn't set a color outside of the aes call.
2. manufacturer, model, cyl, trans, drv, fl, class are categorical variables
3.

- color to categorical or to continuous gets a gradient
- size to categorical will give a different size for each category
- size to continuous (displ) will bin the values and give a different size for each bin
- shape to categorical will give a different shape for each category
- shape to continuous is not recomended as shapes are discrete and continuous is not, also there is a limited number of shapes that can be displayed

4. You get a perfect 45 degree line because you are plotting the same variable against itself.
5. This aes is easier to see if you set the shape to something like a circle with no fill, stroke controls the size of the border of the circle.
6. You get different colors based on the logical that you provided.

```
mpg |>
  ggplot() +
    geom_point(aes(x=displ, y=displ, color=displ < 5), shape=21)
```