

Data Visualization with ggplot2

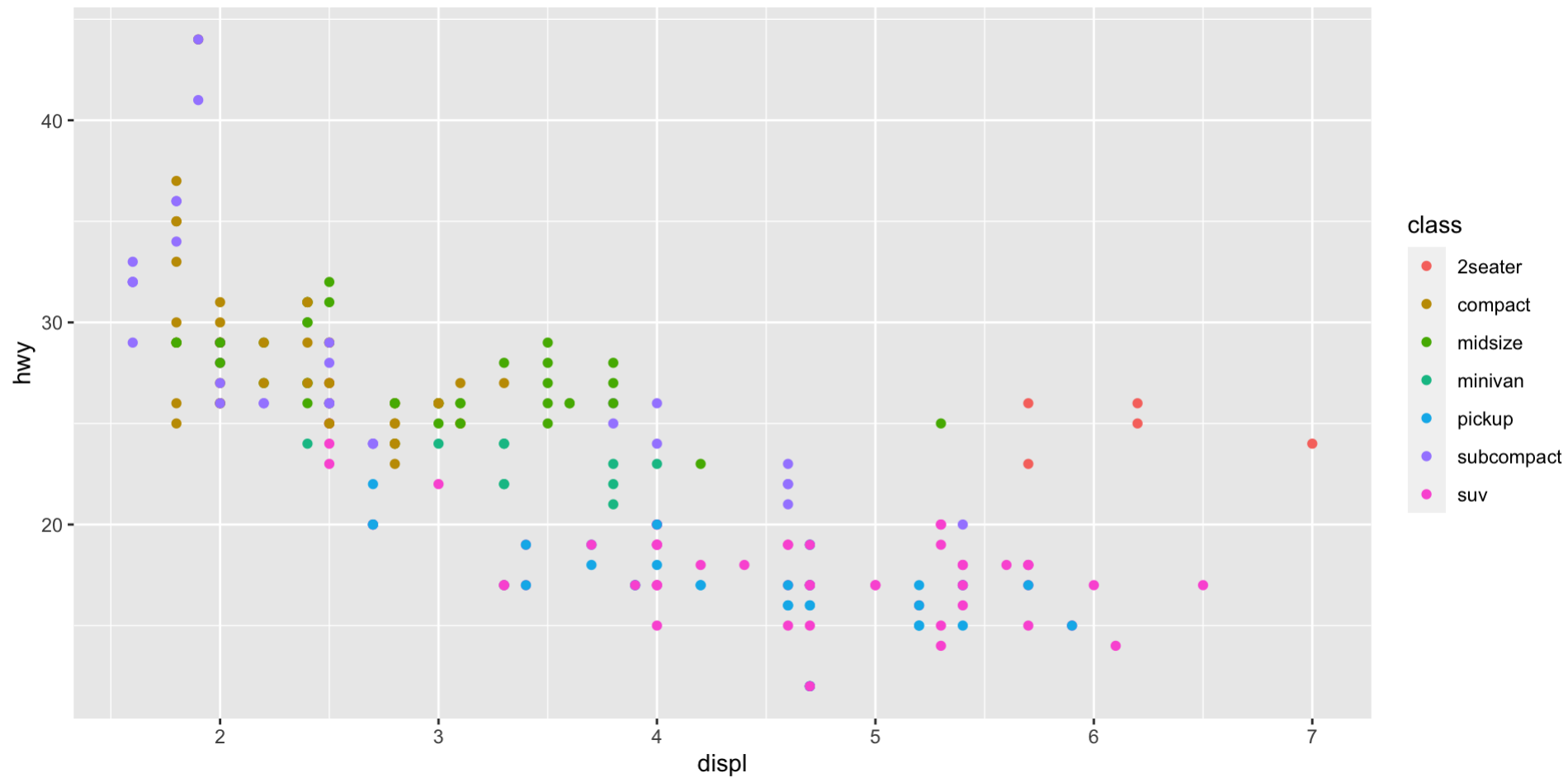
Vahan Aslanyan

GGPlot2

- Popular package to build figures and visualize data
- Website: <https://ggplot2.tidyverse.org/>
- Cheatsheet:
<https://github.com/rstudio/cheatsheets/blob/main/data-visualization-2.1.pdf>

Example

```
1 # to install: install.packages("tidyverse")
2 library("ggplot2")
3 ggplot(mpg, aes(displ, hwy, colour = class)) +
4   geom_point()
```



```
1 #displ=engine displacement, in litres, highway miles per gallon
```

Structure

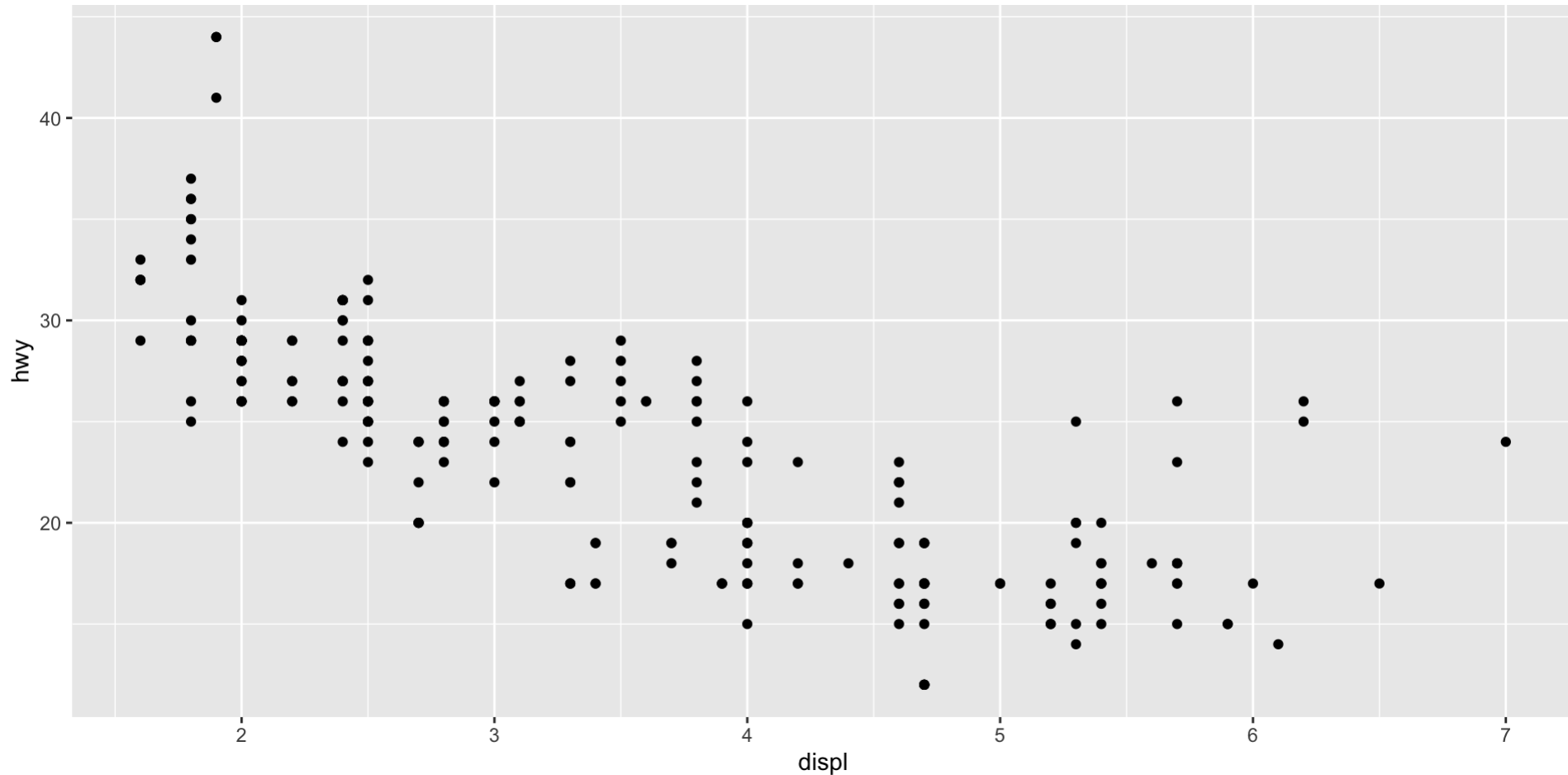
Overall Basic structure

Structure

- ggplot2 is extremely flexible, so mastering takes time
- don't be afraid to google things you are stuck on, chances are someone already encountered and fixed it
- now lets go back to our example

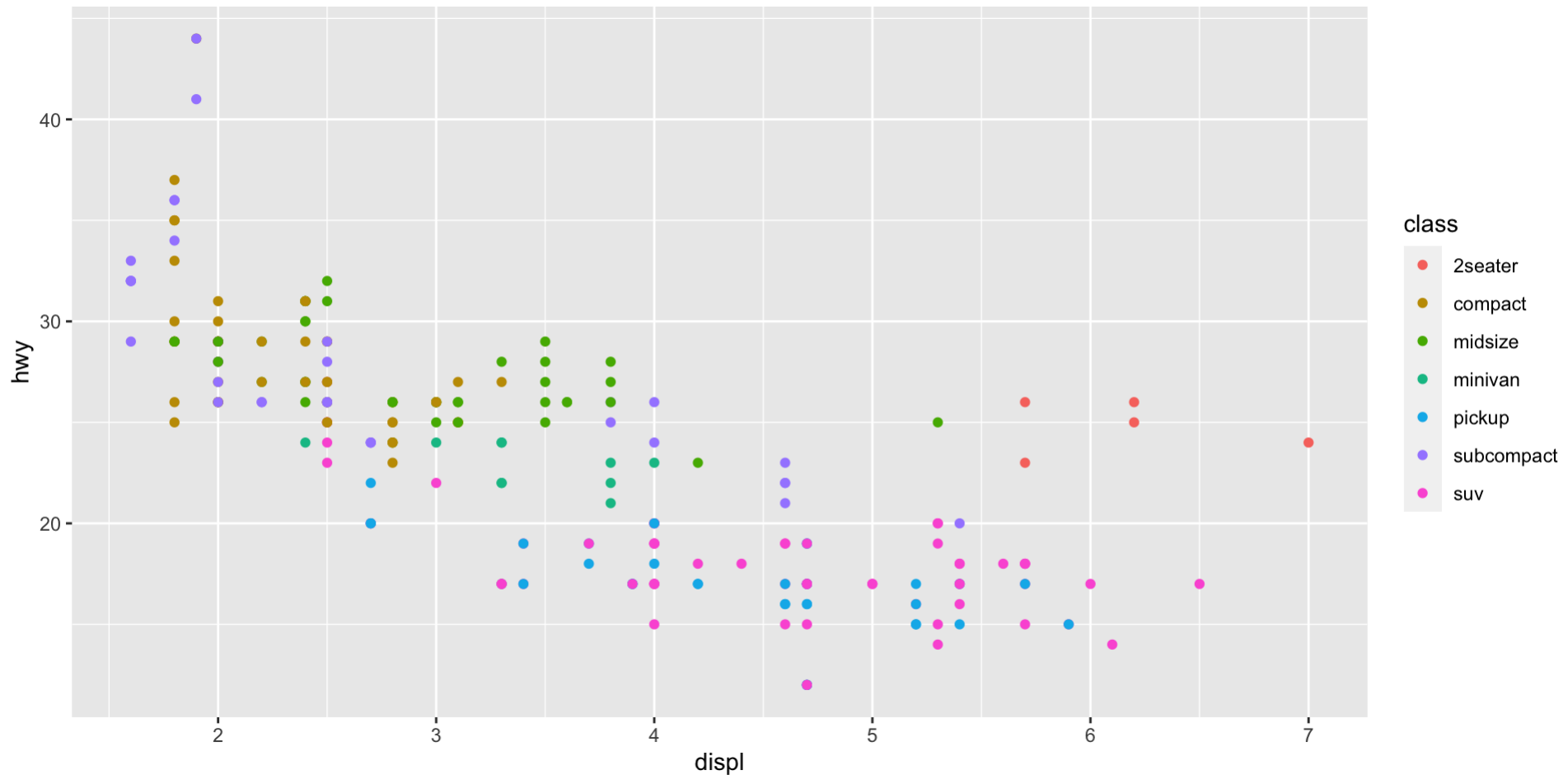
Creating a ggplot

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



Creating a ggplot

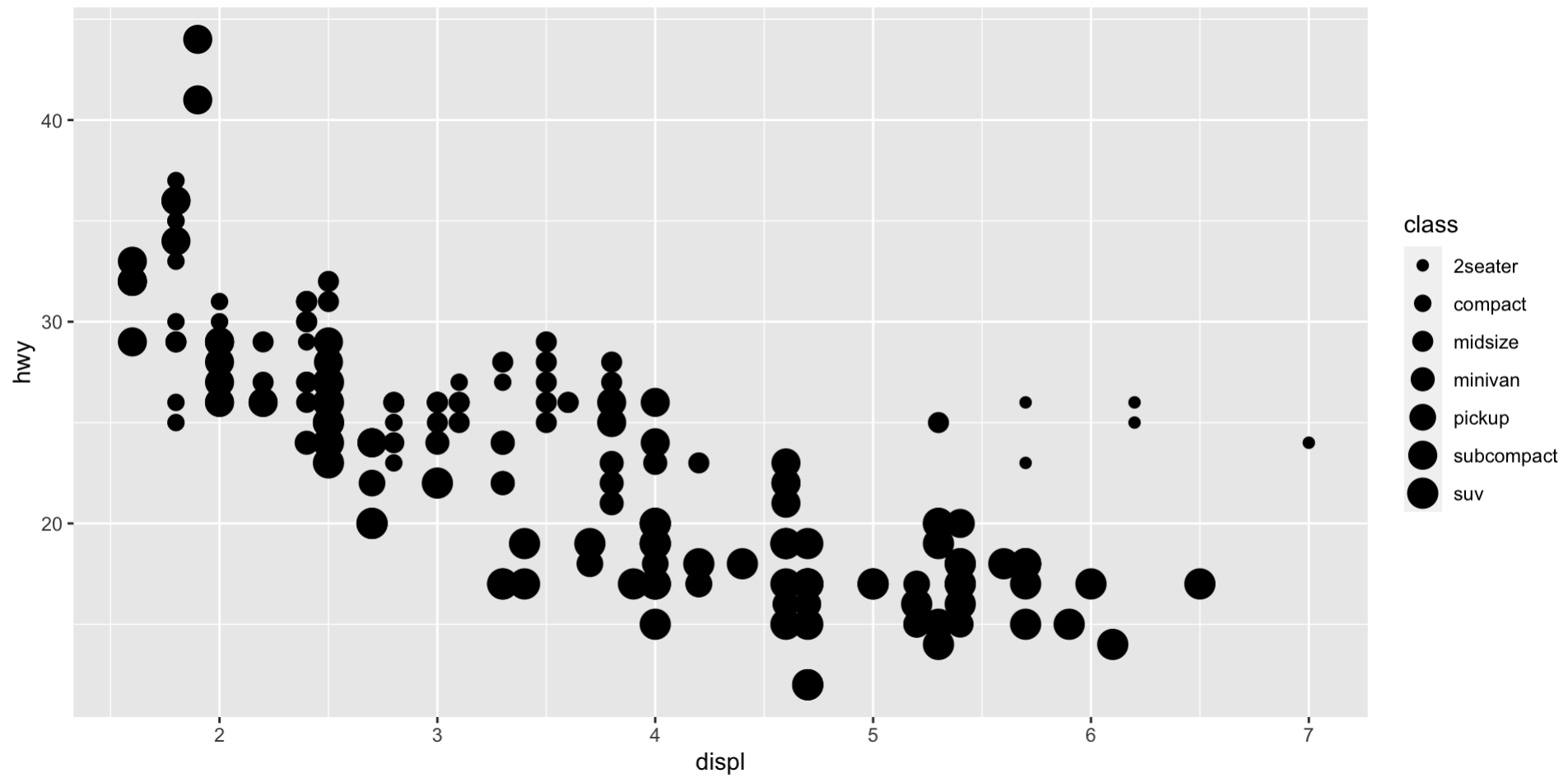
```
1 ggplot(data = mpg) +  
2   geom_point(mapping = aes(x = displ, y = hwy,color=class))
```



Creating a ggplot

Besides color we can pass shape, alpha to aes()

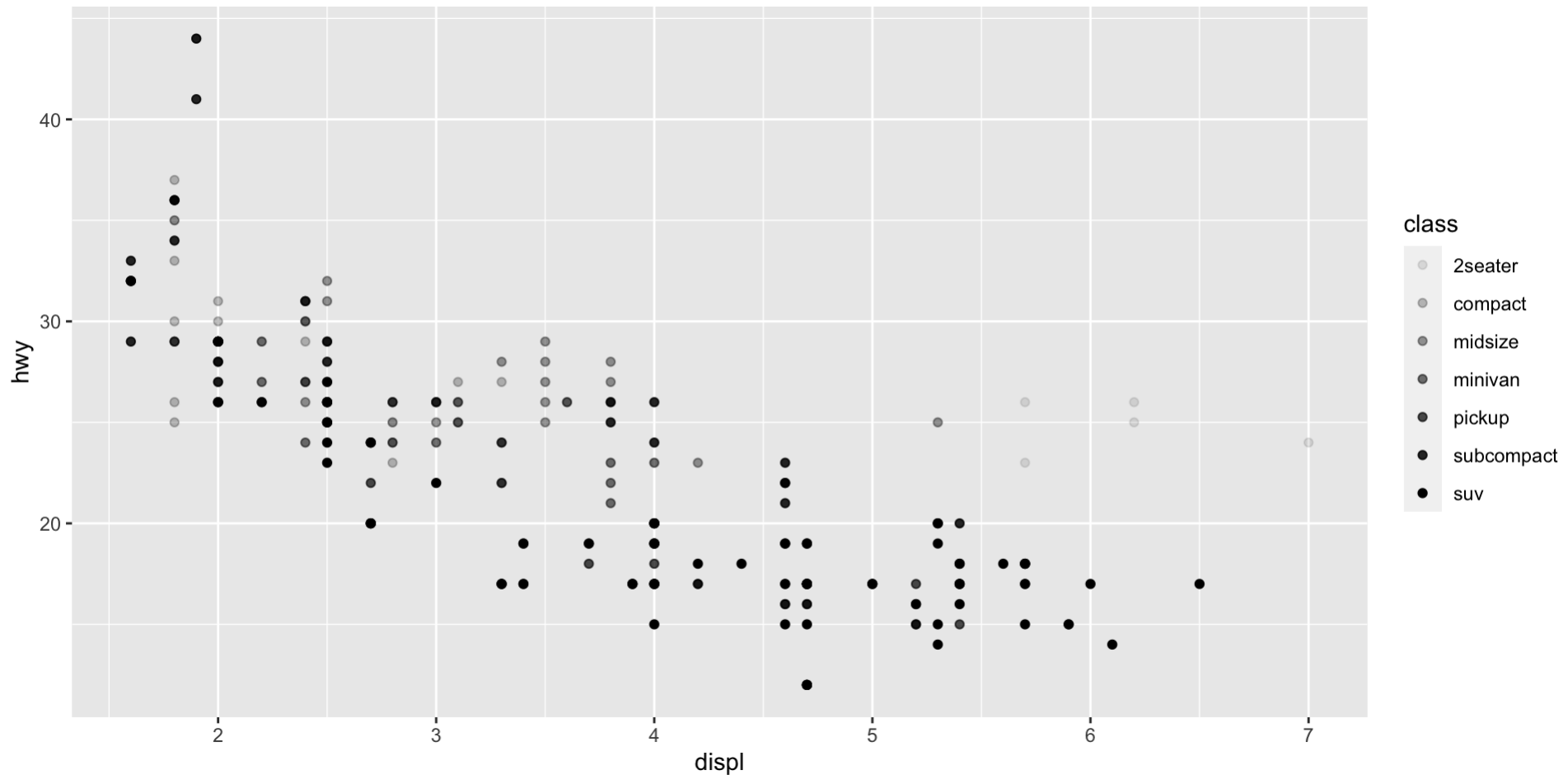
```
1 ggplot(data = mpg) +  
2   geom_point(mapping = aes(x = displ, y = hwy, size=class))
```



```
1 #its going to be ugly
```

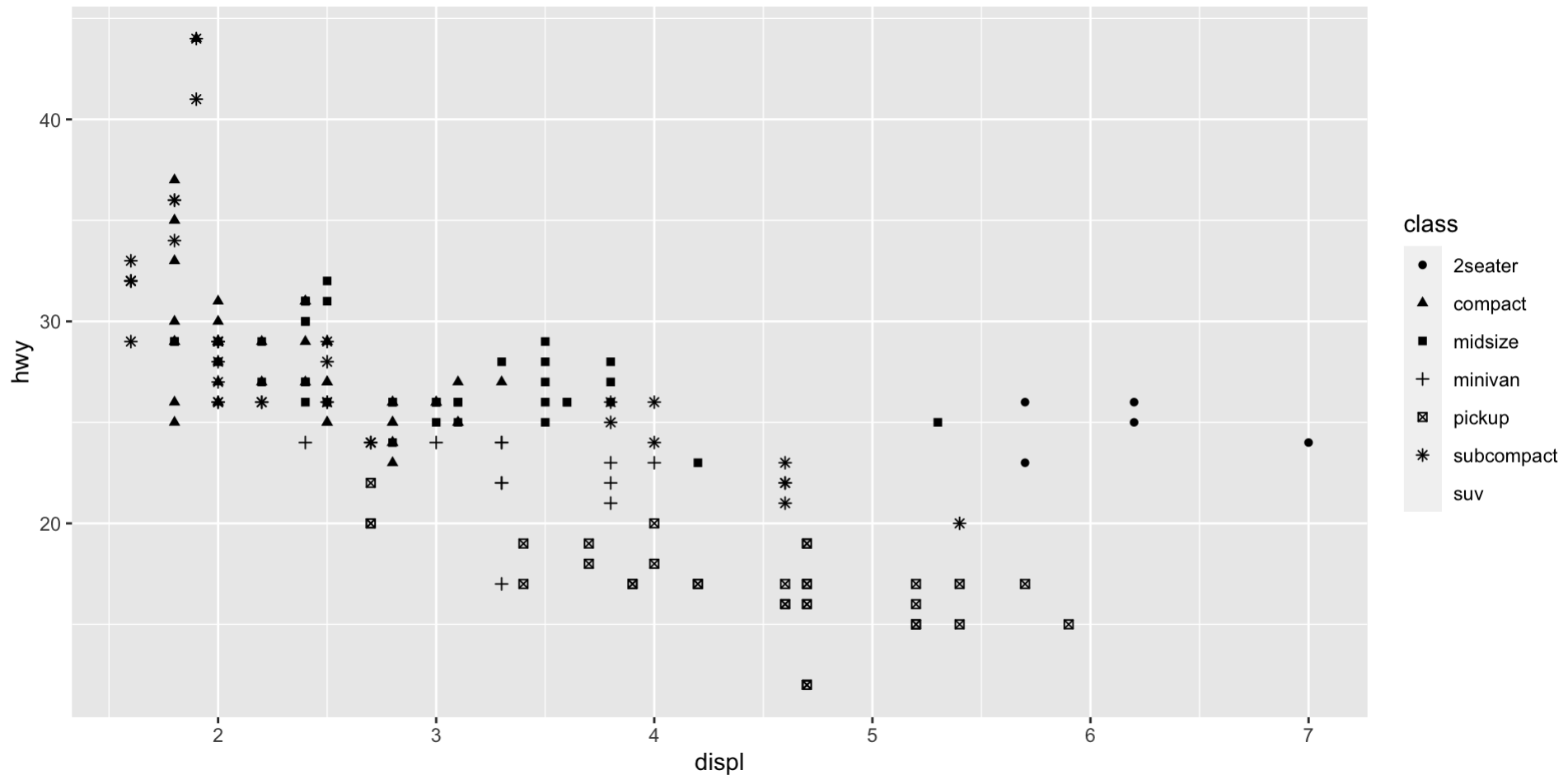
Creating a ggplot

```
1 ggplot(data = mpg) +  
2   geom_point(mapping = aes(x = displ, y = hwy, alpha=class))
```



Creating a ggplot

```
1 ggplot(data = mpg) +  
2   geom_point(mapping = aes(x = displ, y = hwy, shape=class))
```



Creating a ggplot

We can also set the color manually.

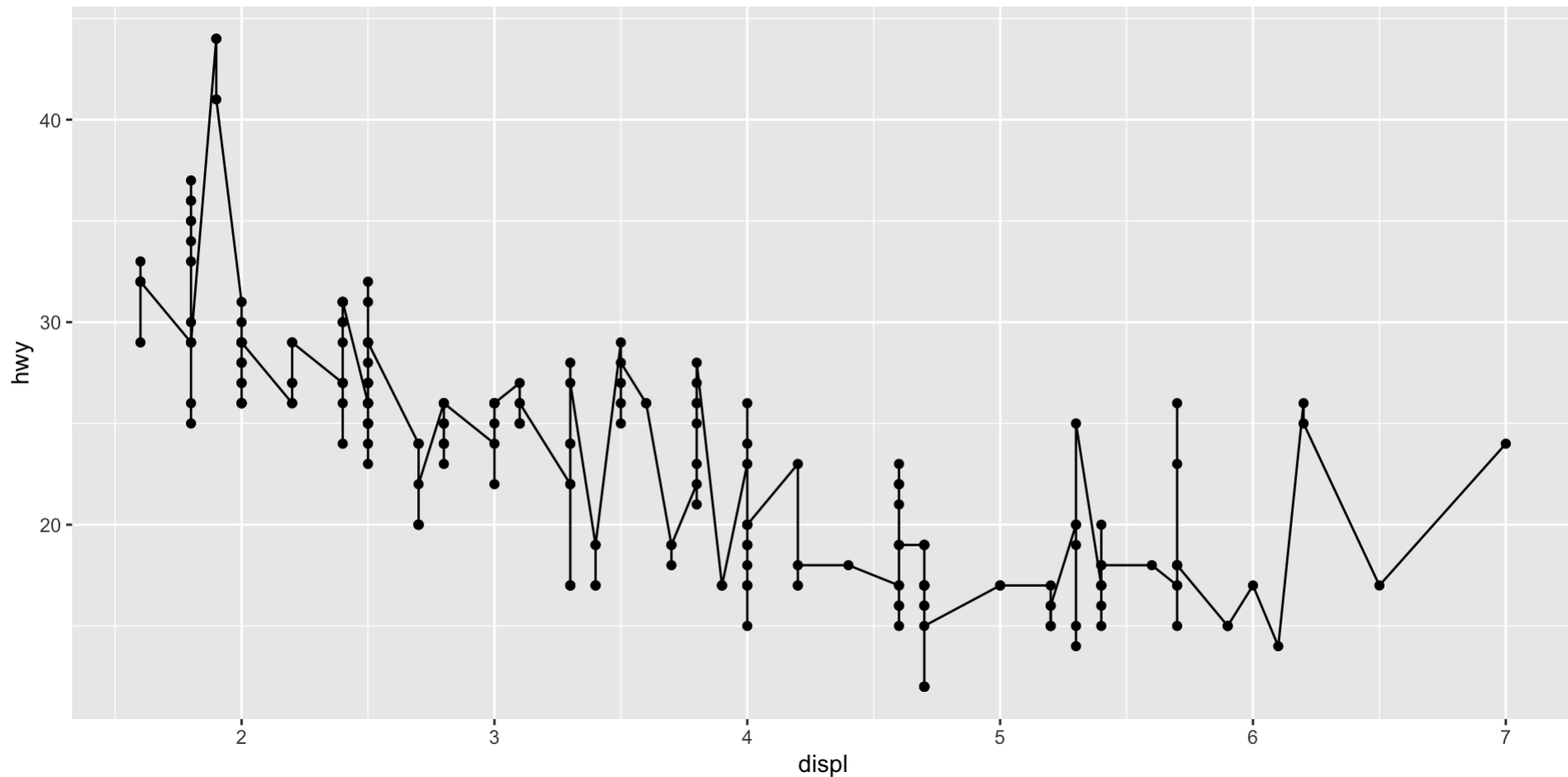
```
1 ggplot(data = mpg) +  
2   geom_point(mapping = aes(x = displ, y = hwy),color="forestgreen")
```

Creating a ggplot

Lets add a regression line here

Creating a ggplot

```
1 ggplot(data = mpg, aes(x = displ, y = hwy)) +  
2   geom_point()+  
3   geom_line()
```

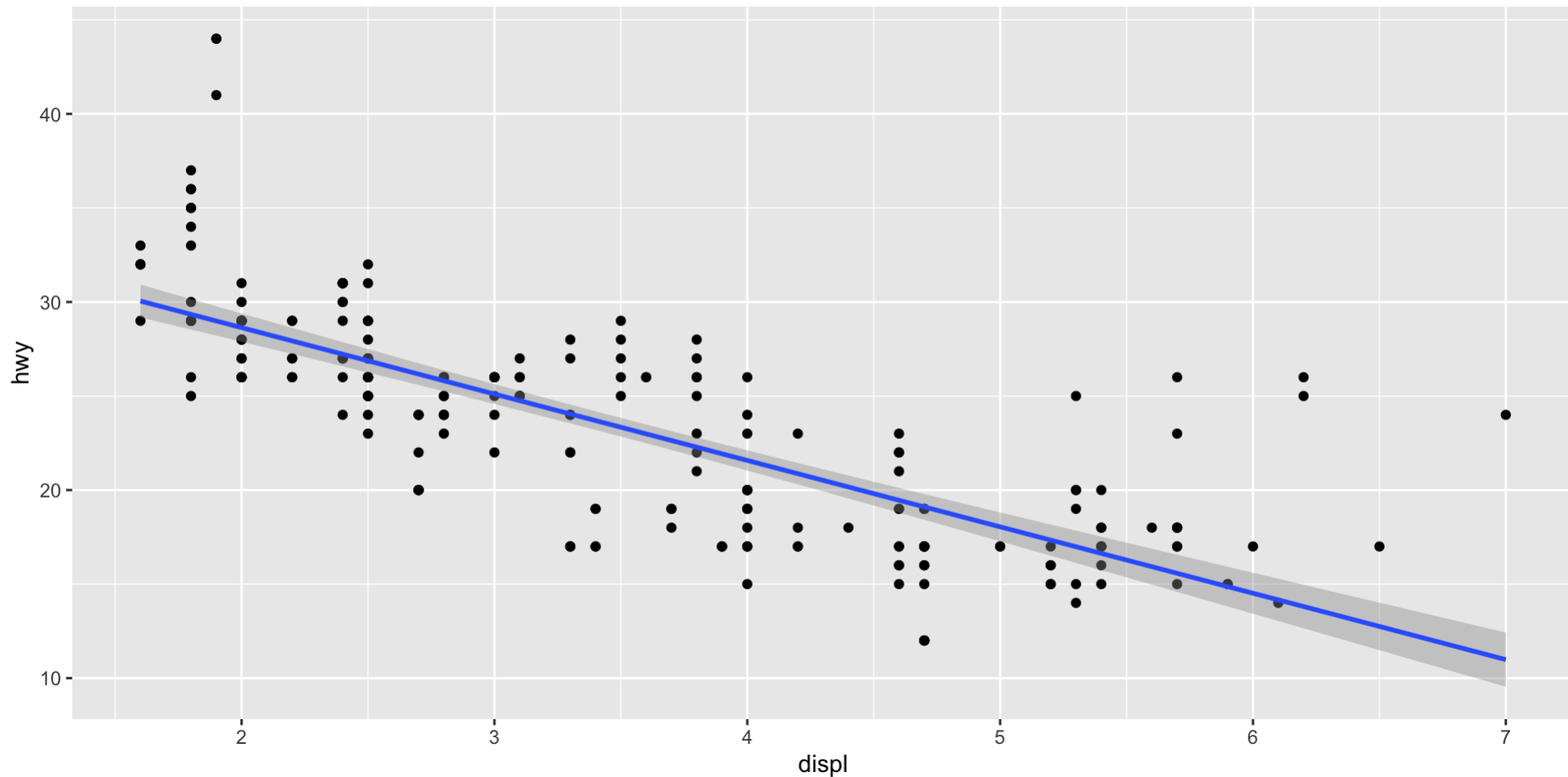


Creating a ggplot

Lets add a regression line here

Creating a ggplot

```
1 ggplot(data = mpg, aes(x = displ, y = hwy)) +  
2   geom_point()+  
3   geom_smooth(method = "lm")
```

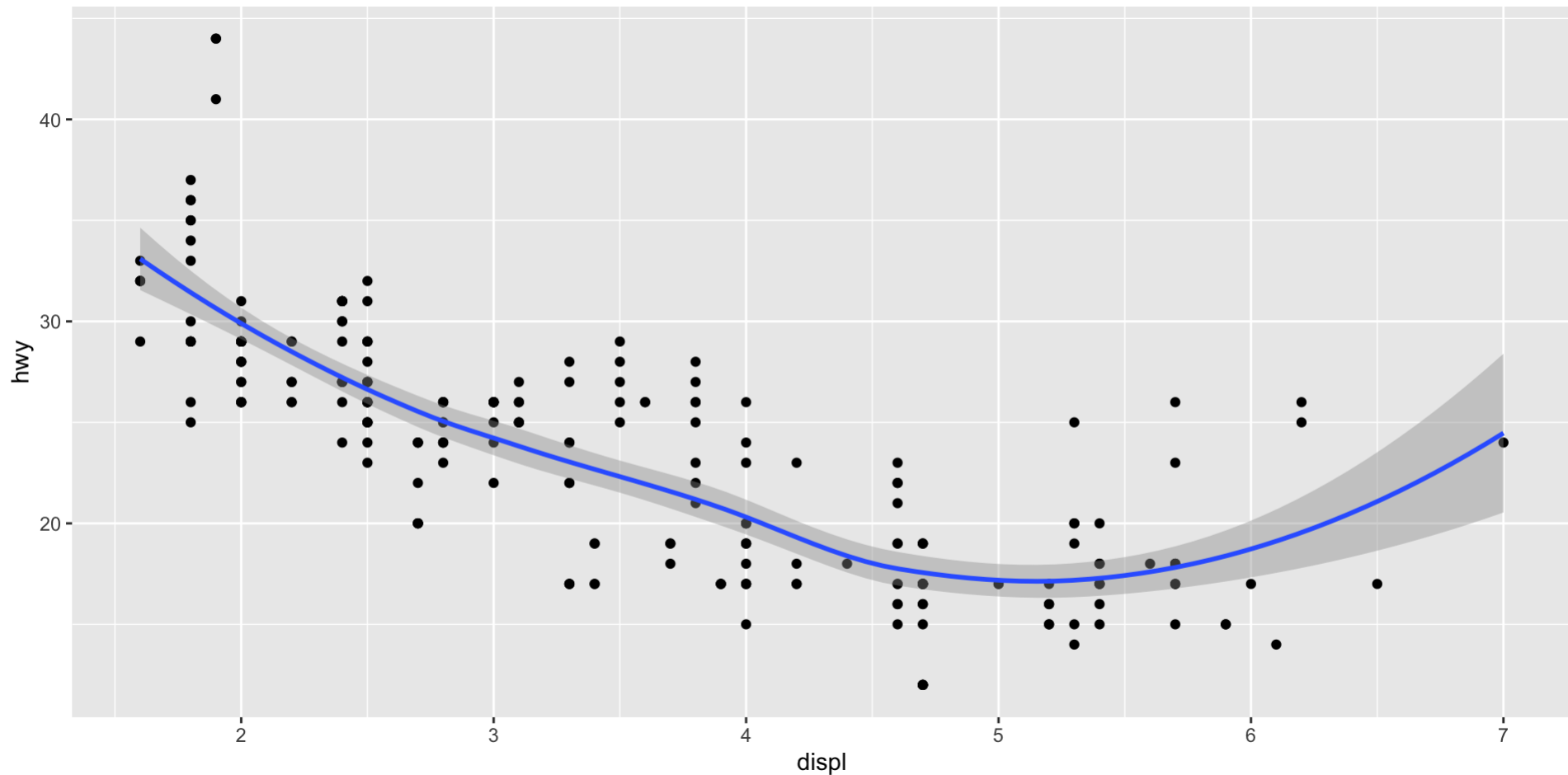


Creating a ggplot

Lets add a regression line here

Creating a ggplot

```
1 ggplot(data = mpg, aes(x = displ, y = hwy)) +  
2   geom_point()+  
3   geom_smooth()
```

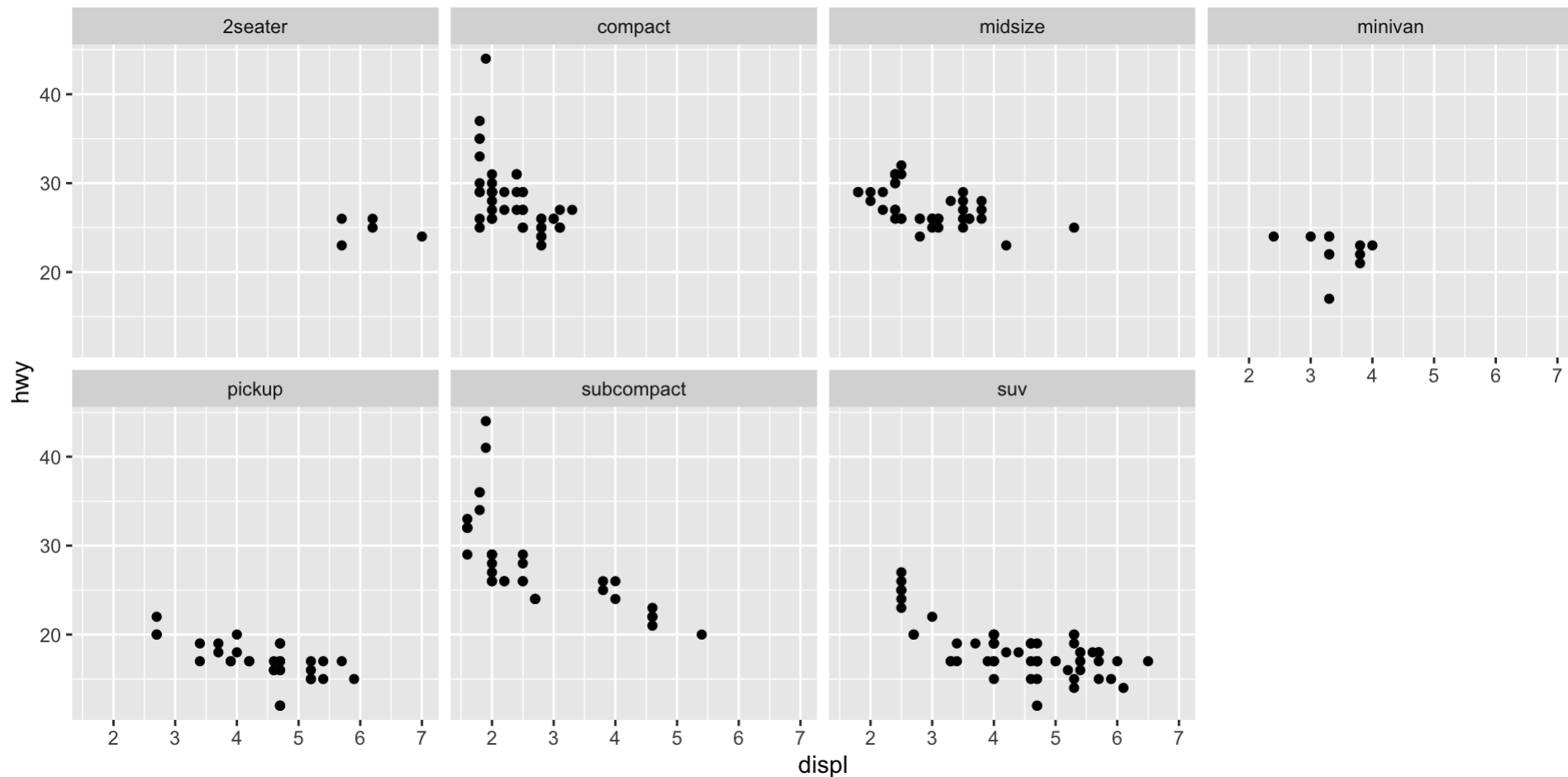


Creating a ggplot-facets

Split the plot into subplots that each display one subset of the data

Creating a ggplot-facets

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

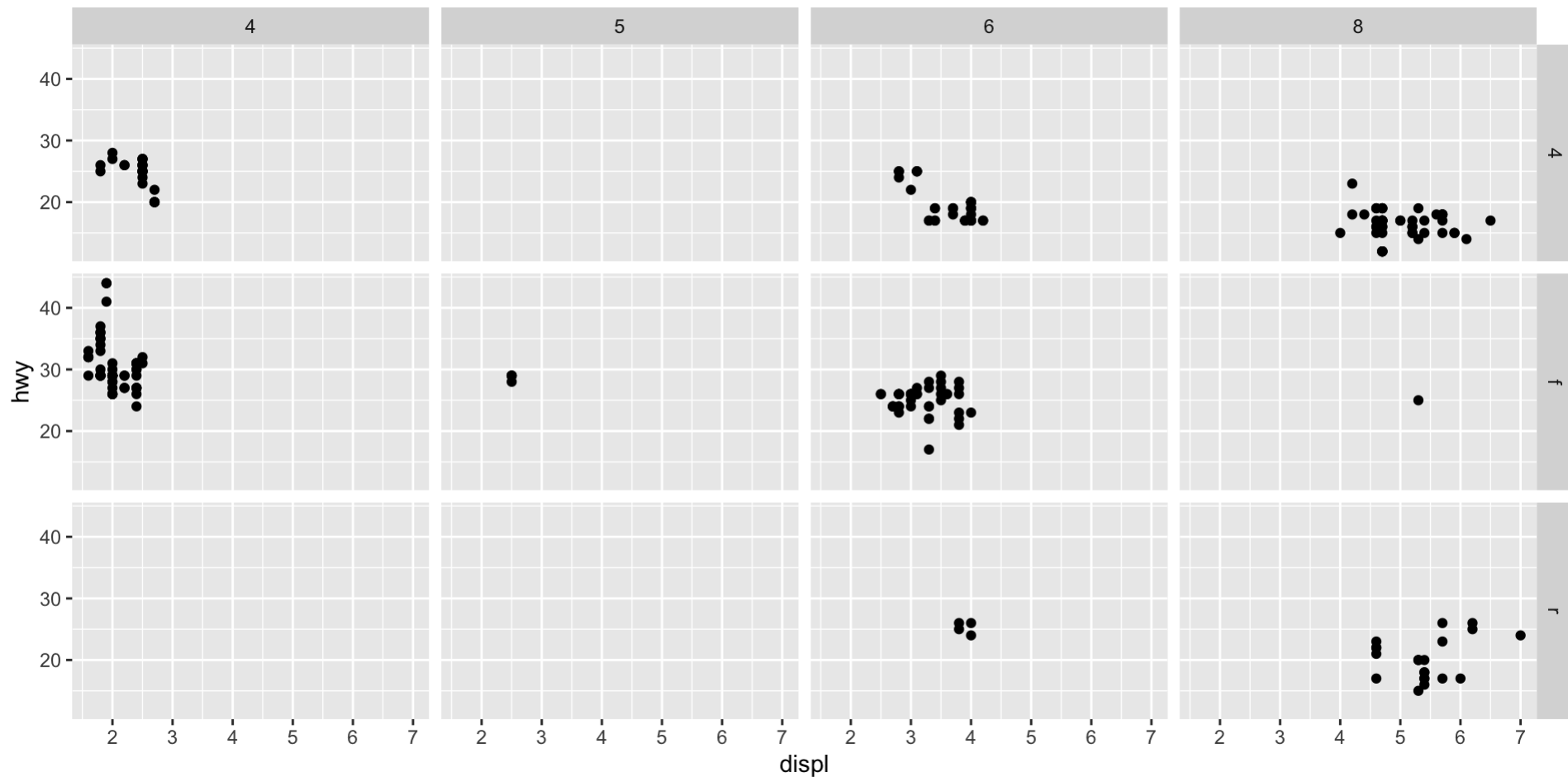


Creating a ggplot-facets

We can create subplots by 2 categorical variables using `facet_grid()`

Creating a ggplot-facets

```
1 ggplot(data = mpg) +  
2   geom_point(mapping = aes(x = displ, y = hwy)) +  
3   facet_grid(drv ~ cyl)
```



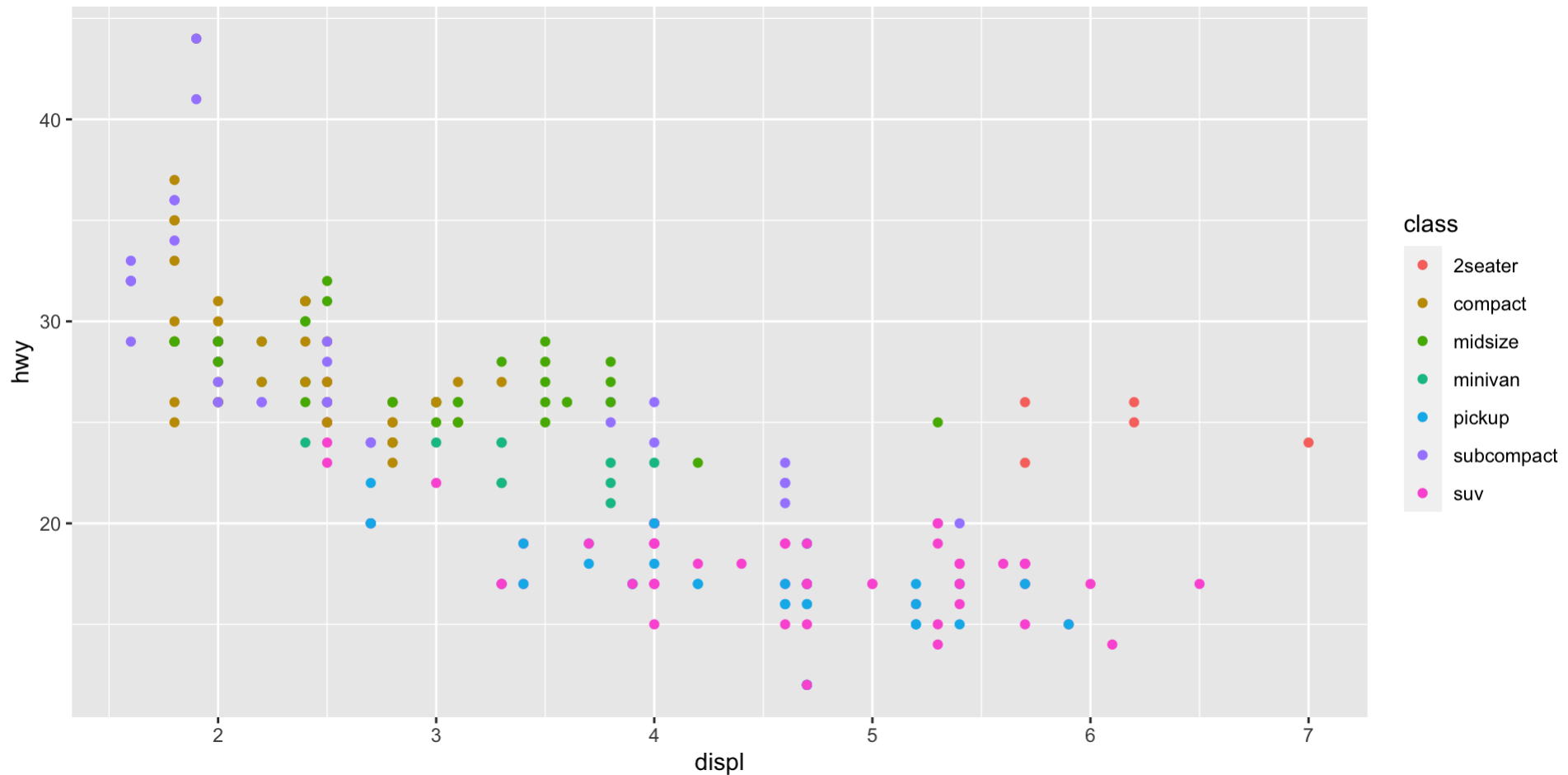
Creating a ggplot

Lets go back to our previous graph

How can we improve this?

Creating a ggplot

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



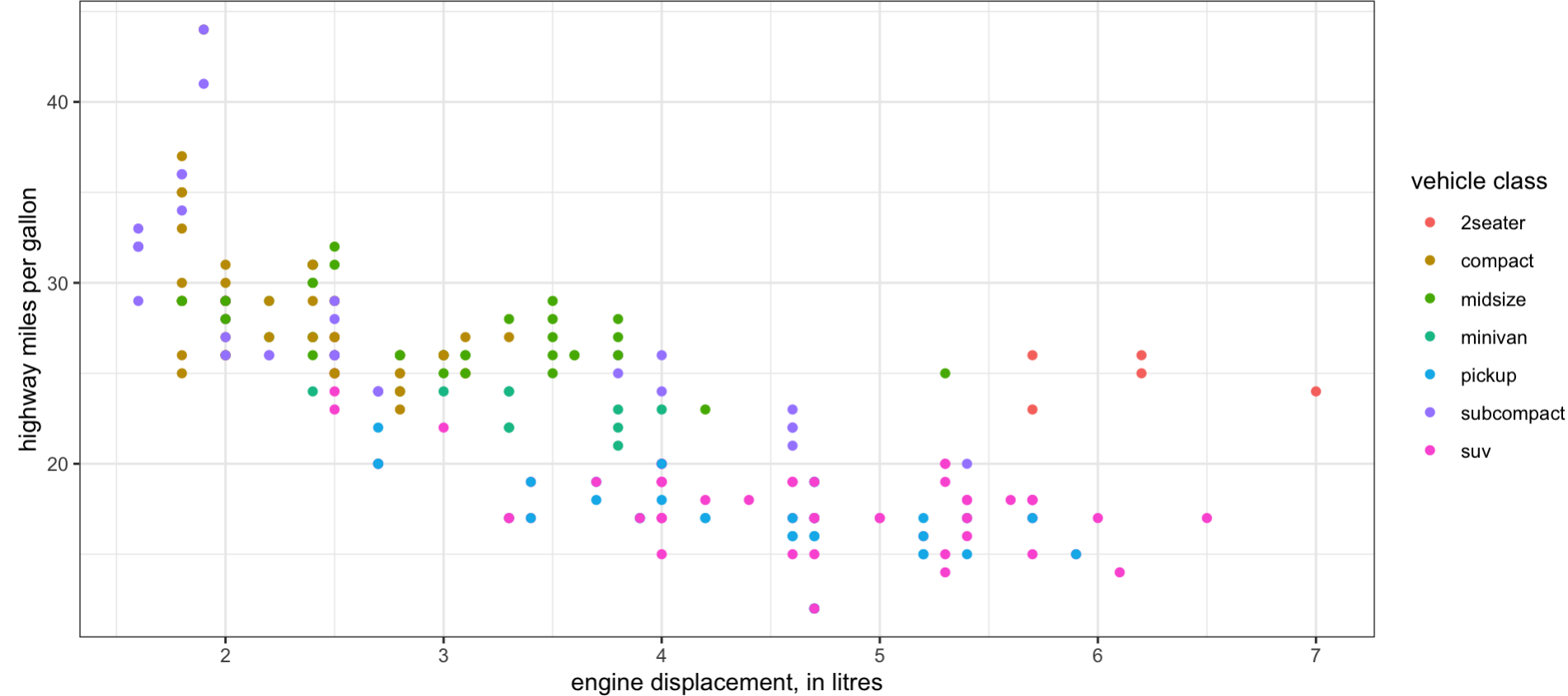
Creating a ggplot

- We can change the background theme (my favorite is theme black and white (`theme_bw()`))
- We need a title
- x and y axis labels can be more descriptive
- legend label can be more descriptive

Creating a ggplot

```
1 ggplot(data = mpg, aes(x = displ, y = hwy,color=class)) +  
2   geom_point()+theme_bw()+  
3   labs(title= "Highway miles per gallon (per vehicle class)",  
4         subtitle= "created by ggplot2",x= "engine displacement, in litres",  
5         y="highway miles per gallon",color="vehicle class")
```

Highway miles per gallon (per vehicle class)
created by ggplot2



Extending ggplot

- There are other packages that extend ggplot2
- <https://exts.ggplot2.tidyverse.org/gallery/>

References

- For both sections, I have used [R for data science]<https://r4ds.had.co.nz/>
- A source I usually use for colors:
<http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf>