Data -

Categorical data - records categories

E.g.: - Survey in class of who all are left handed or right handed

Numeric data :- number

Eg:- Doctor { weight of a patient, age of a patient

- Continuous
- Discrete

This week :- Data visualisation. (ggplot)

Categorical data :. Visualisation Tool

Bar charts [also Pie charts]

- Compute the frequency / Proportion of each category
- plot it across categories

Numeric data

Histograms :- _____ Box plots :- _____

Scatter plots :- _____

(fill these in as class proceeds)

ggplot2- Data Visualisation

ggplot2 implements grammar of graphics

```
> install.packages("tidyverse")
```

Once installed then to add to current workspace

```
> library("tidyverse")
```

ggplot2- Data Visualisation

Fuel Economy data from 1999 - 2008 of popular model of cars.

Dataset in `tidyverse`

```
> mpg
```

Observations collected by US Environment Protection Agency on 38 models of cars.

ggplot2- Data Visualisation

Lecture Entirely from: Chapter 3 R For Data Science

https://r4ds.had.co.nz/data-visualisation.html

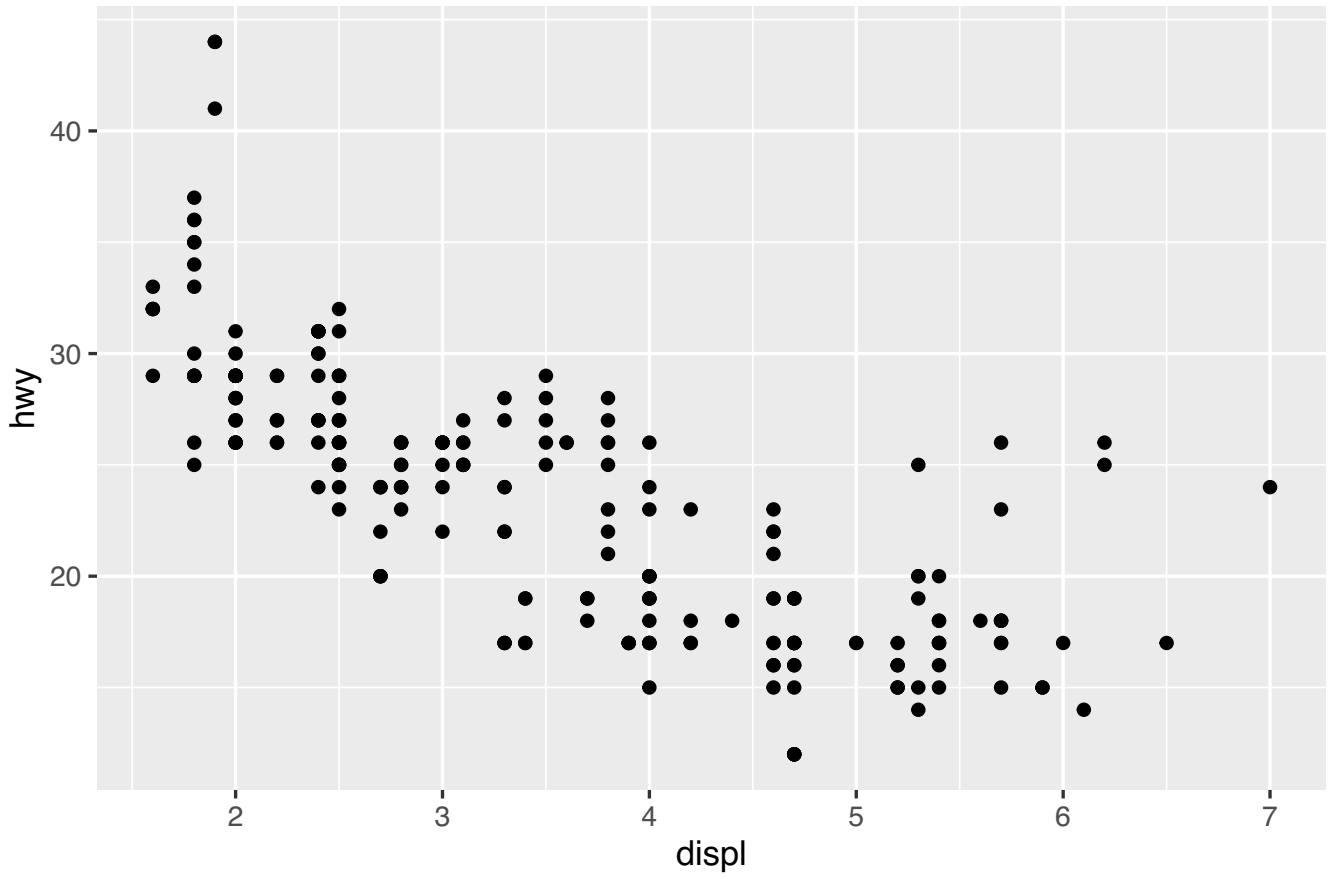
ggplot2- Data Visualisation

Scatter plot :- one variable against another (numeric)

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

displ :-
Engine displacement in litres

hwy :-
highway miles per gallon.



The plots negative relationship between Engine Size and Fuel Effiency.

ggplot()

- Begins with a function `ggplot()`- creates a coordinate system that you can add addlayers to. The first arugment is the data set to use `ggplot(data=mpg)` creates an empty graph.
- Add layers to `ggplot()`- the function `geom_point()` adds a layer of points to your plot
- Each geom function takes a `mapping` argument. The `mapping` argument is always paired with `aes()`
- `ggplot(data= <DATA>)+ GEOM-FUNCTION(mapping=aes(<MAPPINGS>))`
- We will learn how to complete and extend this basic template.

ggplot2- Aesthetics Mappings

```
# A tibble: 6 × 3
  displ   hwy class
  <dbl> <int> <chr>
1   1.8    29 compact
2   1.8    29 compact
3   2      31 compact
4   2      30 compact
5   2.8    26 compact
6   2.8    26 compact
```

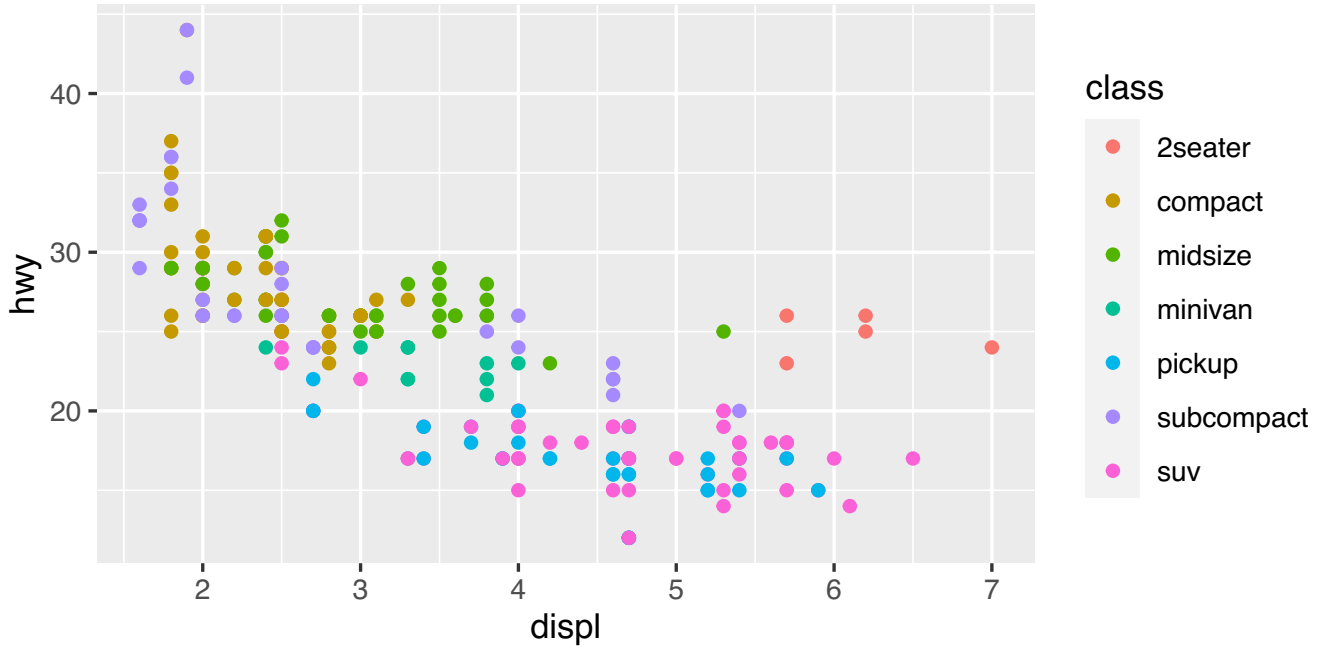
- `disp` a car's engine size, in litres.
- `hwy`, a car's fuel efficiency on the highway, in miles per gallon (mpg).

Question: Is the relationship linear ? Does it depend on `class`?

Graphical Answer: Add a third variable to plot in `class`

ggplot2- Aesthetics Mappings

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



Added a third variable called `class` to a 2−D scatter plot by mapping it to an *a*esthetic.

ggplot2- Scaling

- `aes()`- Associate the `name of the aesthetic` to the `name of the variable.`
  - the function gathers together each of the aesthetic mappings used by a layer.
  - passes them to the layer's mapping argument.
  - selects a reasonable scale to use with the aesthetic, and it constructs a legend (or axis labels) that explains the mapping between levels and values.
- Above example
  - Name=colour and Variable=class.
  - Scaling: assigns a unique level of the aesthetic `colour` to a unique level to the variable `class`.
  - Other aesthetics include : shape and size.

ggplot2- Aesthetics Mappings

```
> ggplot(data=mpg) +
+     geom_point(mapping=aes(x=displ, y=hwy), colour=blue)
> # Left
> ggplot(data = mpg) +
+   geom_point(mapping = aes(x = displ, y = hwy, alpha = class))
> # Right
> ggplot(data = mpg) +
+   geom_point(mapping = aes(x = displ, y = hwy, shape = class))
>
```

Added a third variable called `shape, alpha, blue-colour` to a 2−D scatter plot by mapping it to an *a*esthetic.

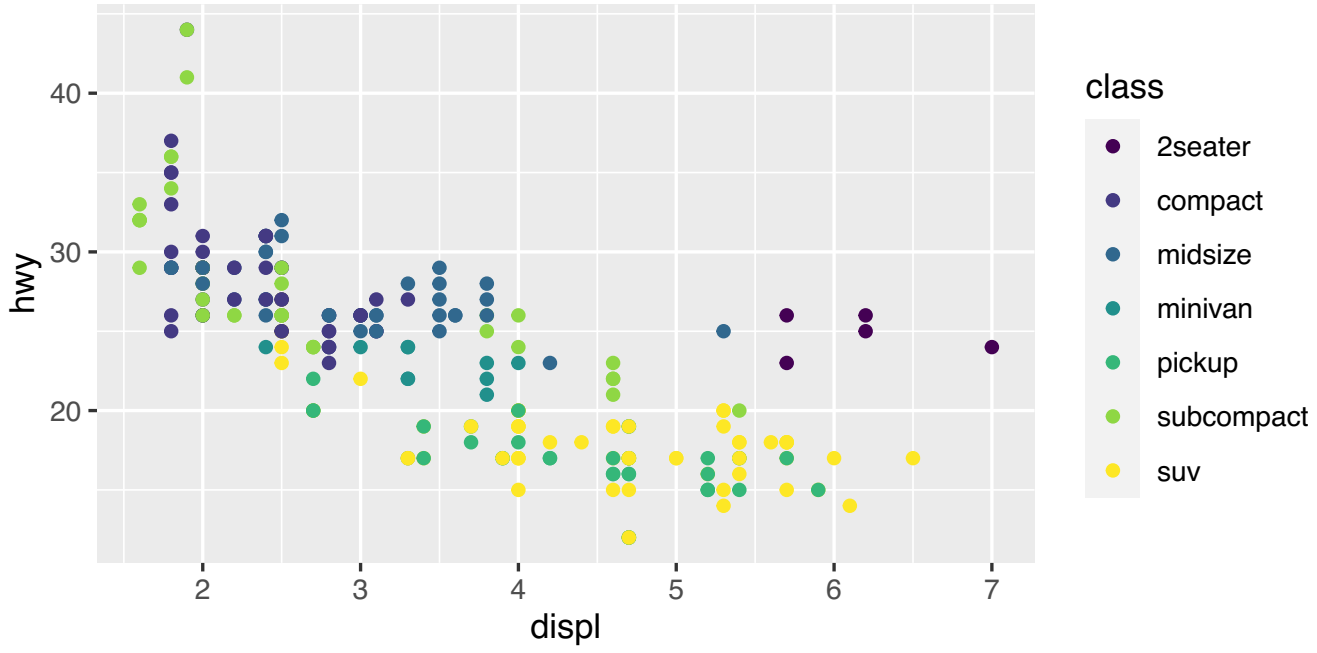
ggplot()-viridis options

- The viridis scales provide colour maps that are perceptually uniform in both colour and black-and-white.
- They are also designed to be perceived by viewers with common forms of colour blindness.
- See also https://bids.github.io/colormap/.

Need : `library(viridis)`

ggplot()-viridis options

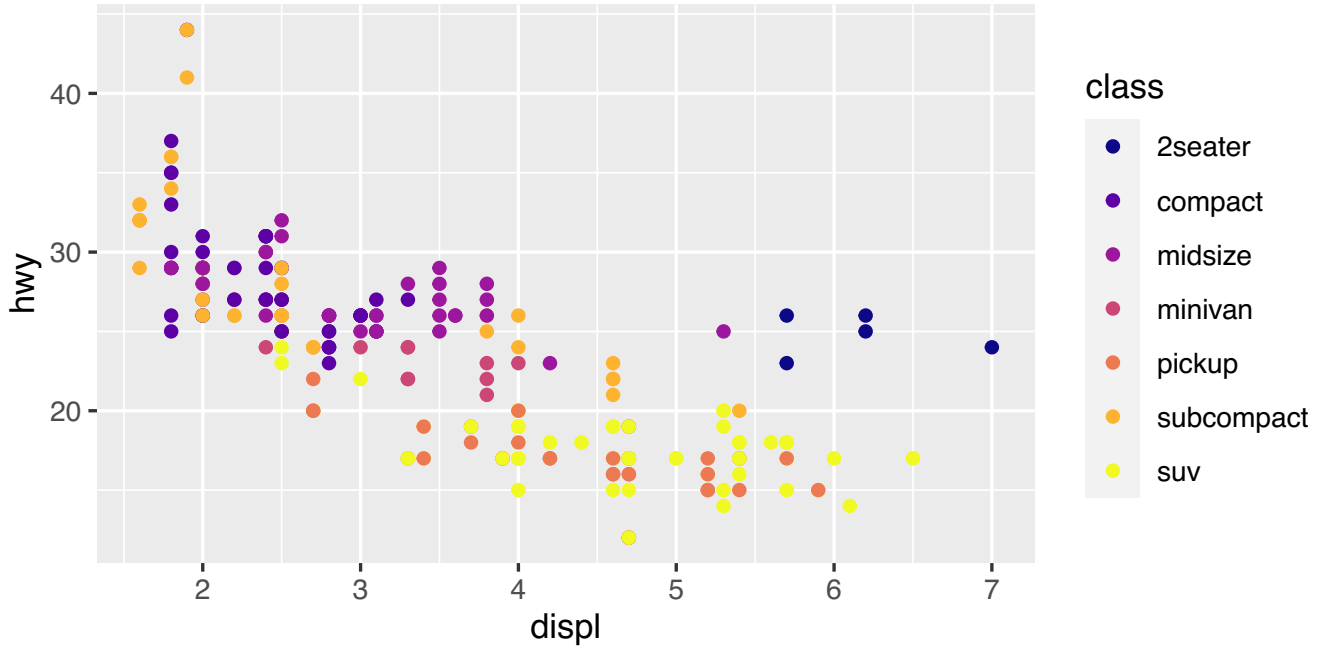
```
> ggplot(data=mpg) +
+     geom_point(mapping=aes(x=displ, y=hwy, colour=class))+
+     scale_colour_viridis_d()
```



Using colour palette from viridis package (colour blind colours).

ggplot()-viridis options

```
> ggplot(data=mpg) +
+     geom_point(mapping=aes(x=displ, y=hwy, colour=class))+
+    scale_colour_viridis_d(option = "plasma")
```



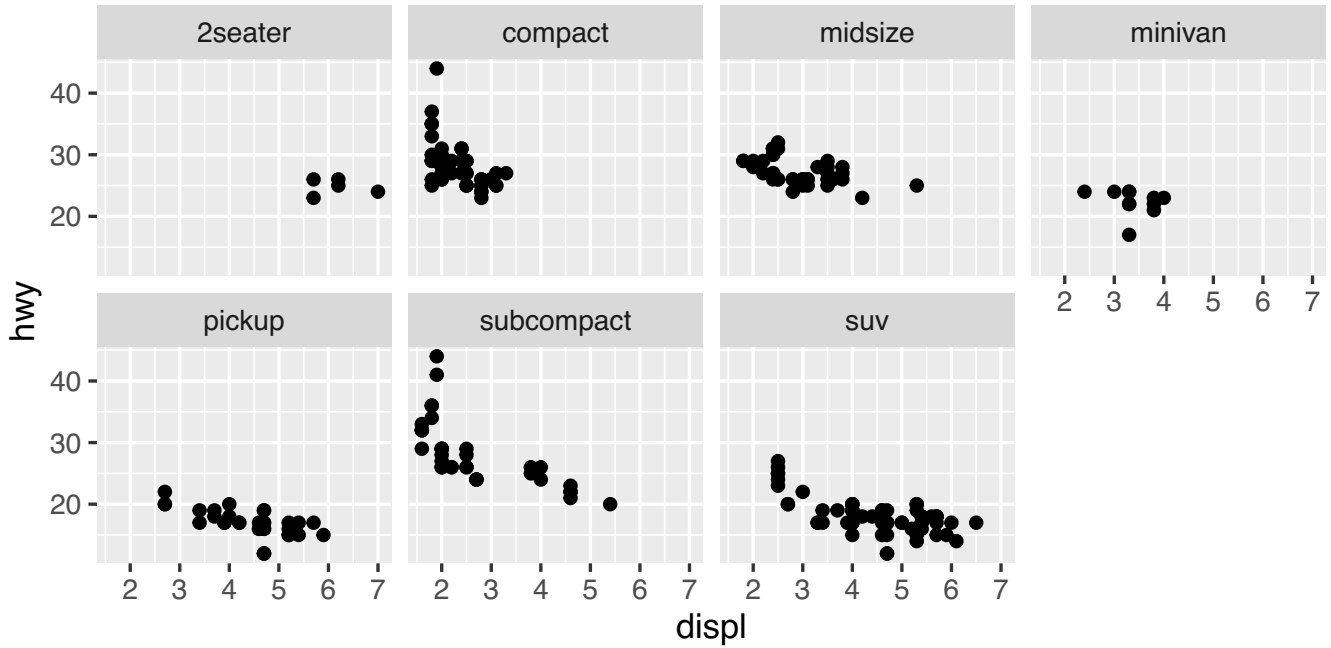
Using colour palette from viridis package (colour blind colours).

ggplot()-facets

- As `aesthetics` was used to add an additional variable to the plot, another way is to add facets (useful for categorical variables).
- `facet_wrap` splits plot by a single variable into subplots that each display one subset of the data.

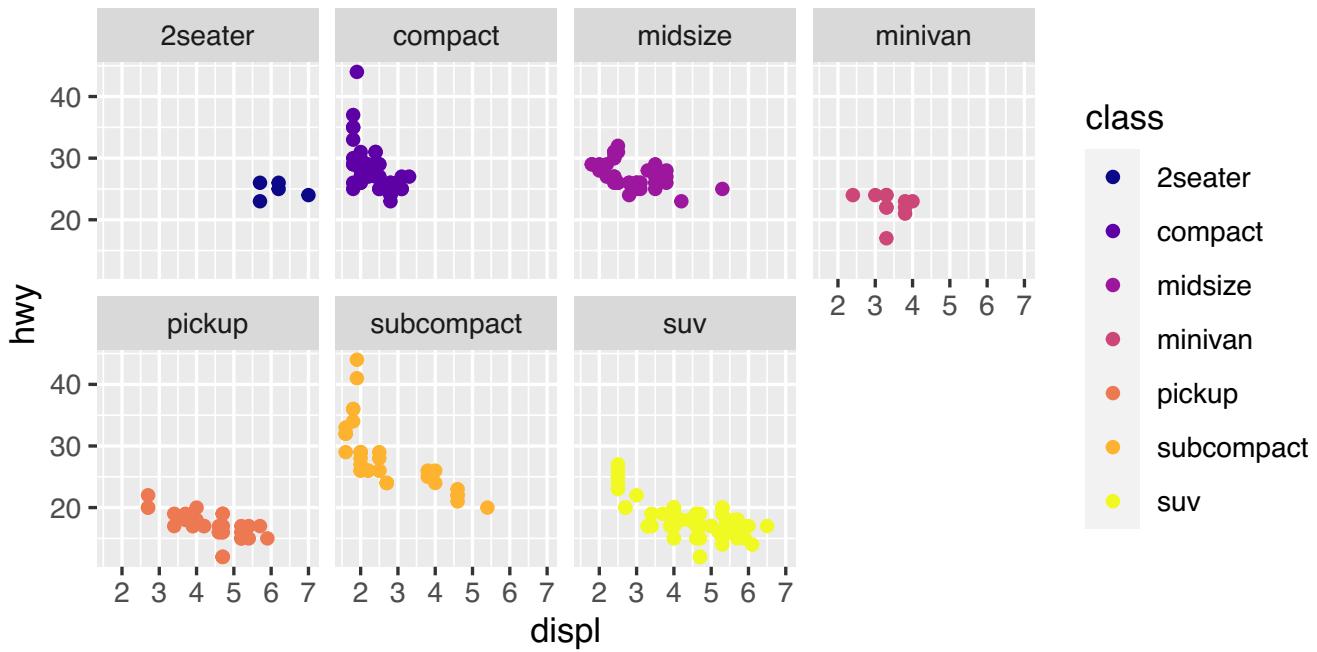
ggplot()-facets

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



ggplot()-facets and Viridis

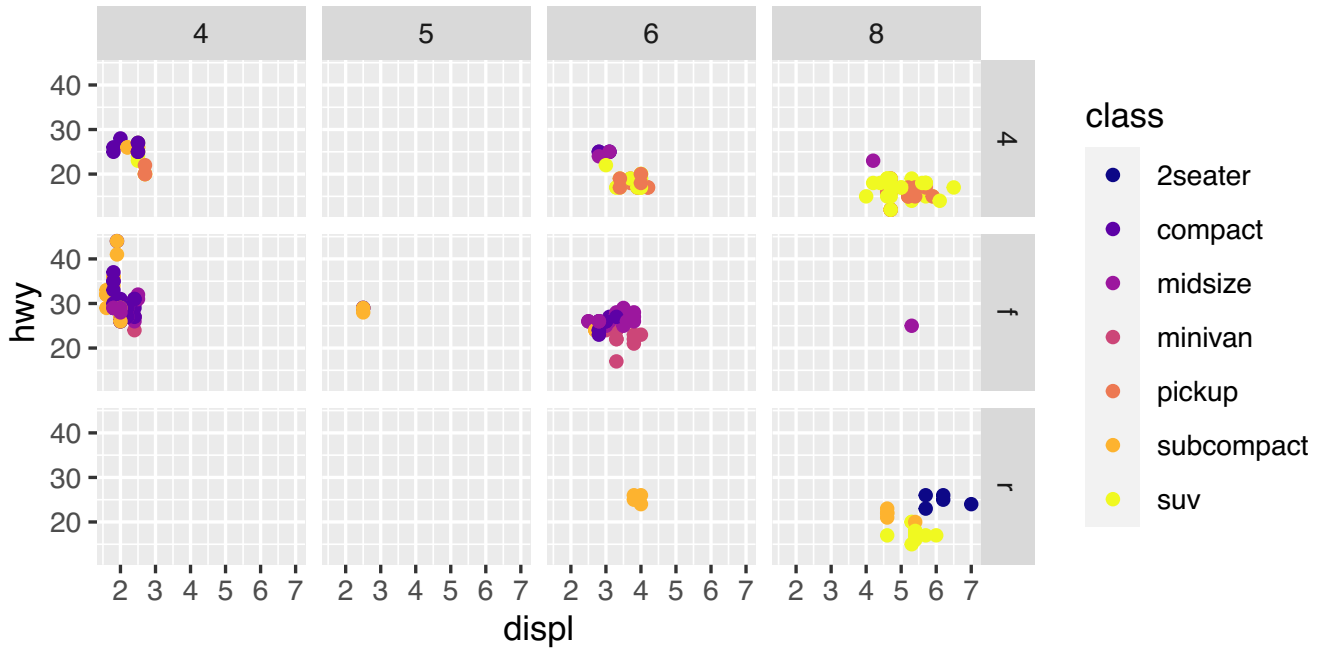
```
> ggplot(data=mpg) +
+     geom_point(mapping=aes(x=displ, y=hwy, colour=class))
+    scale_colour_viridis_d(option = "plasma")+
+     facet_wrap(~class, nrow=2)
```



`facet_wrap` splits plot by a single variable into subplots that each display one subset of the data.

ggplot()-facets

```
> ggplot(data=mpg) +
+     geom_point(mapping=aes(x=displ, y=hwy, colour=class))+
+     scale_colour_viridis_d(option = "plasma")+
+     facet_grid(drv~cyl)
```



drv
- 4 wheel
- front
- rear

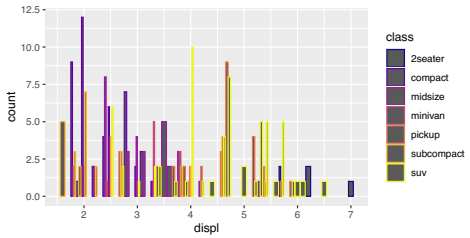
Cyl
- 4
- 5
- 6
- 8

`facet_grid` splits plot by a combination of two variables into subplots that each display one subset of the data.

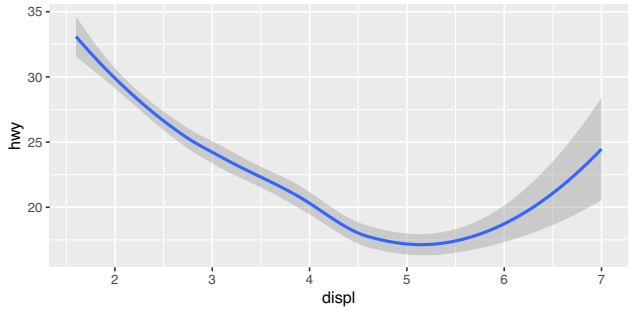
geom-

- geometrical object that a plot uses to represent data.
- can do various plots:

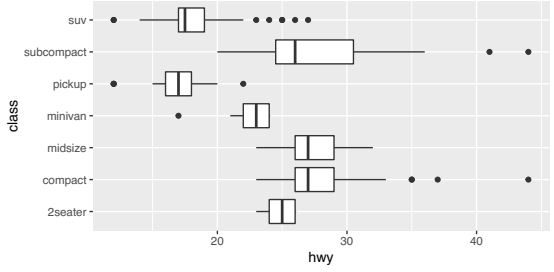
bar charts: `geom_bar`,



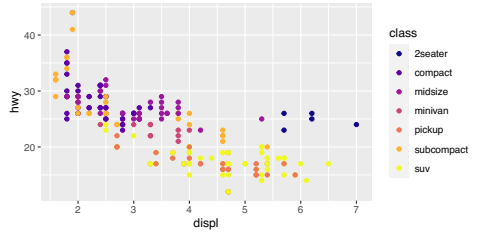
line-charts:`geom_smooth`,



box-plot:`geom_boxplot`,

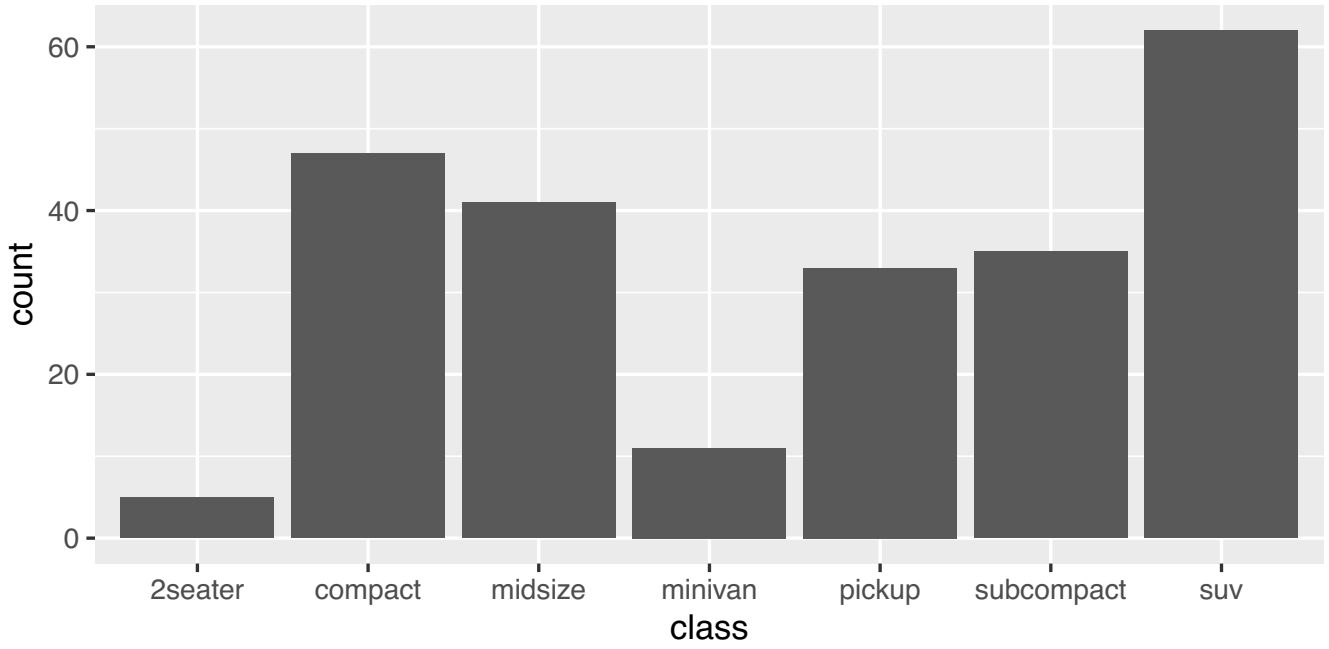


Scatter:`geom_point`,



Stat-count – Bar Charts

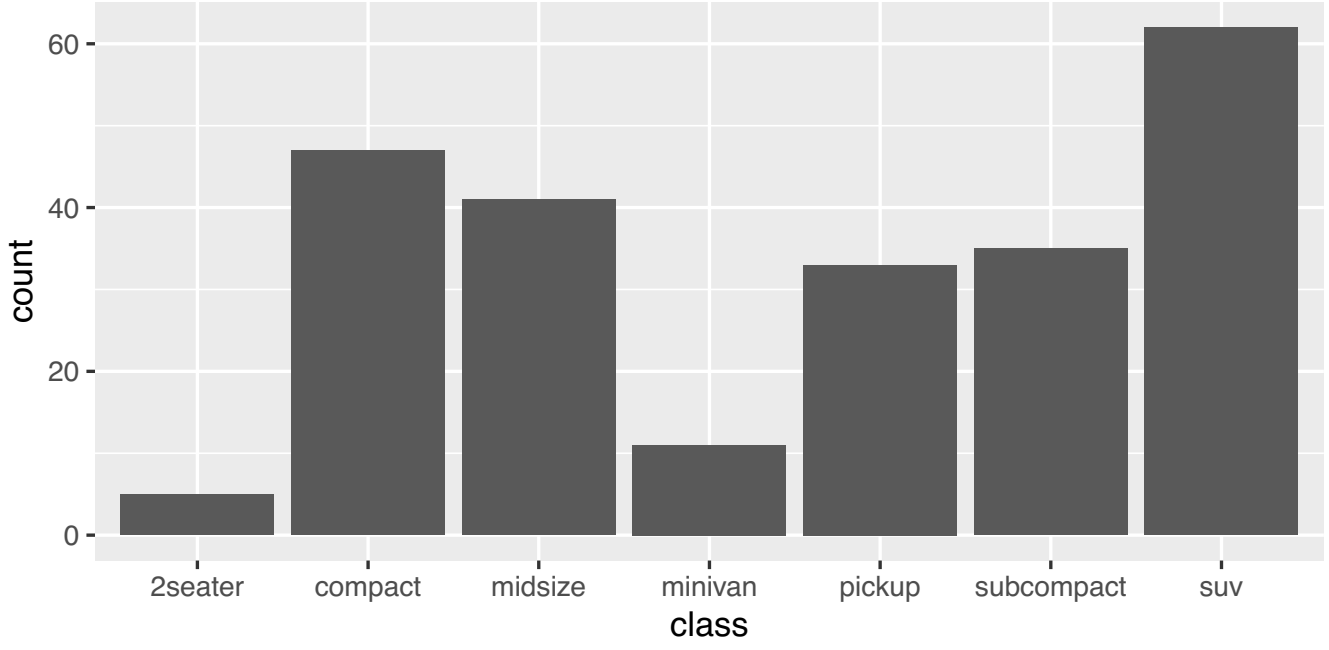
```
> ggplot(data = mpg) +
+   geom_bar(mapping = aes(x = class))
```



Stat-count – Bar Charts

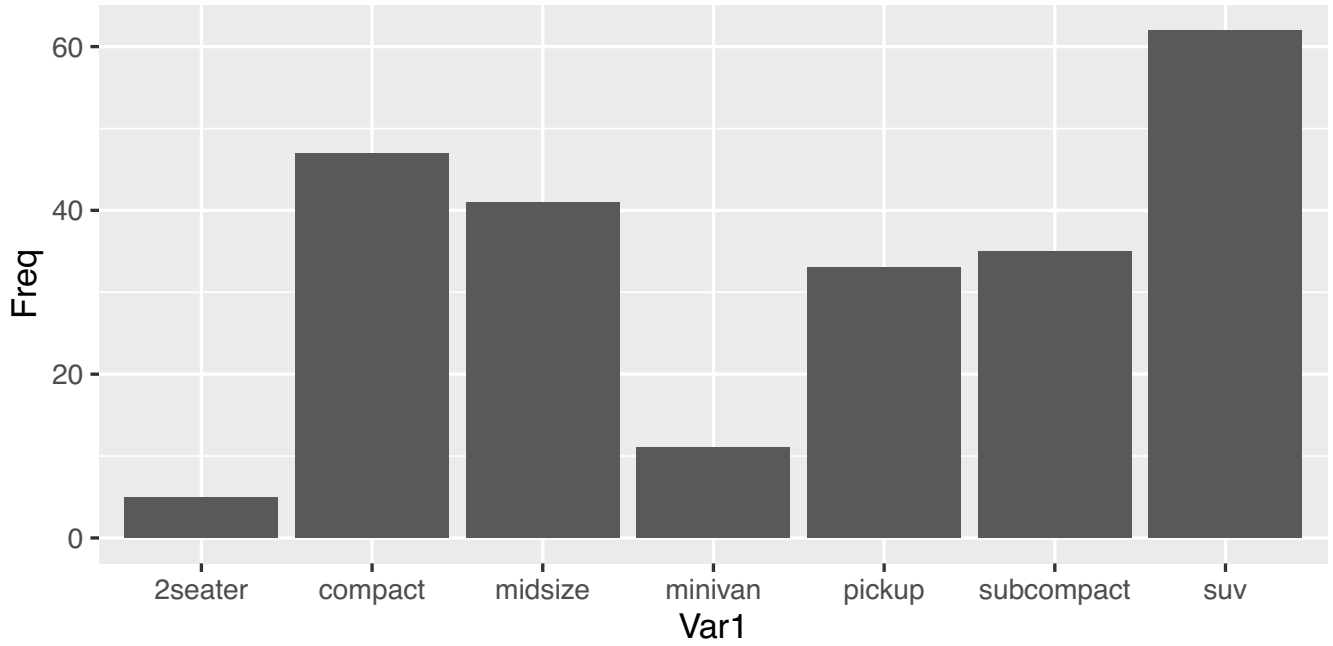
Statistical Transformation

```
> ggplot(data = mpg) +
+   stat_count(mapping = aes(x = class))
```



Stat-count – Bar Charts

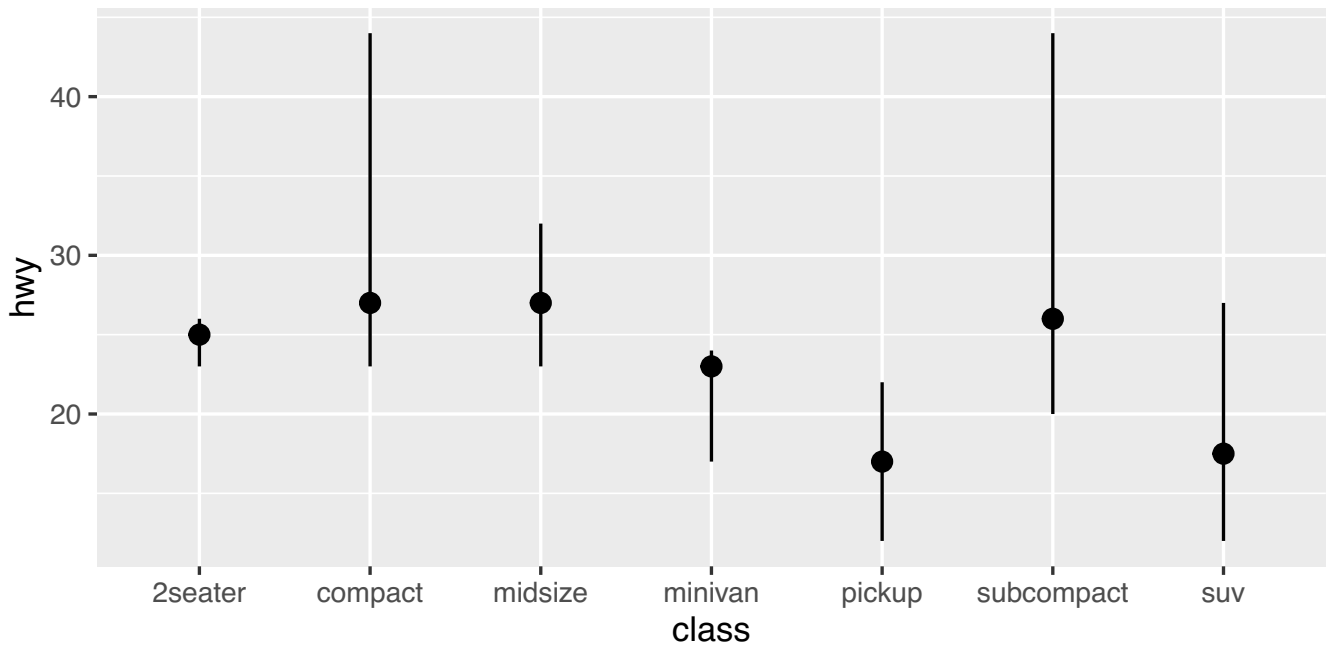
```
> table= as.data.frame(table(mpg$class))
> ggplot(data = table) +   geom_bar(mapping = aes(x = Var1,
+                                    stat="identity")
```



Summary

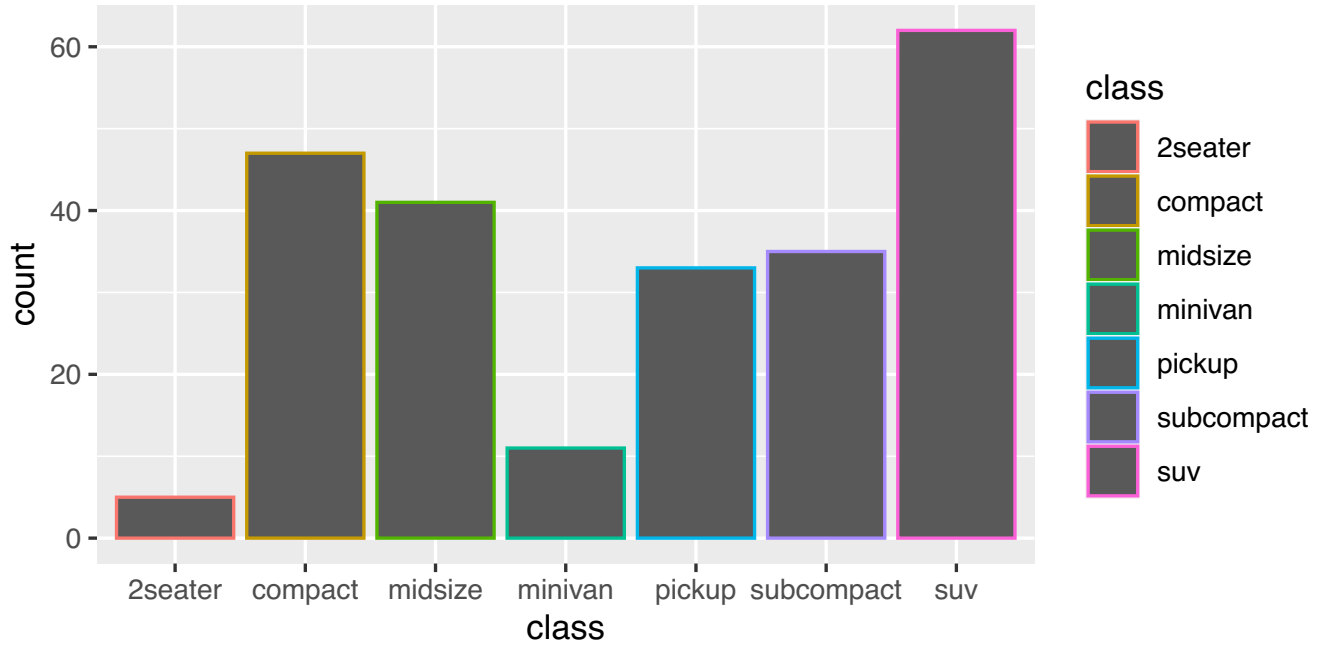
```
> ggplot(data = mpg) +
+   stat_summary(
+     mapping = aes(x = class, y = hwy),
+     fun.min = min,
+     fun.max = max,
+     fun = median
+   )
```

There are around 20 stat functions in ggplot2



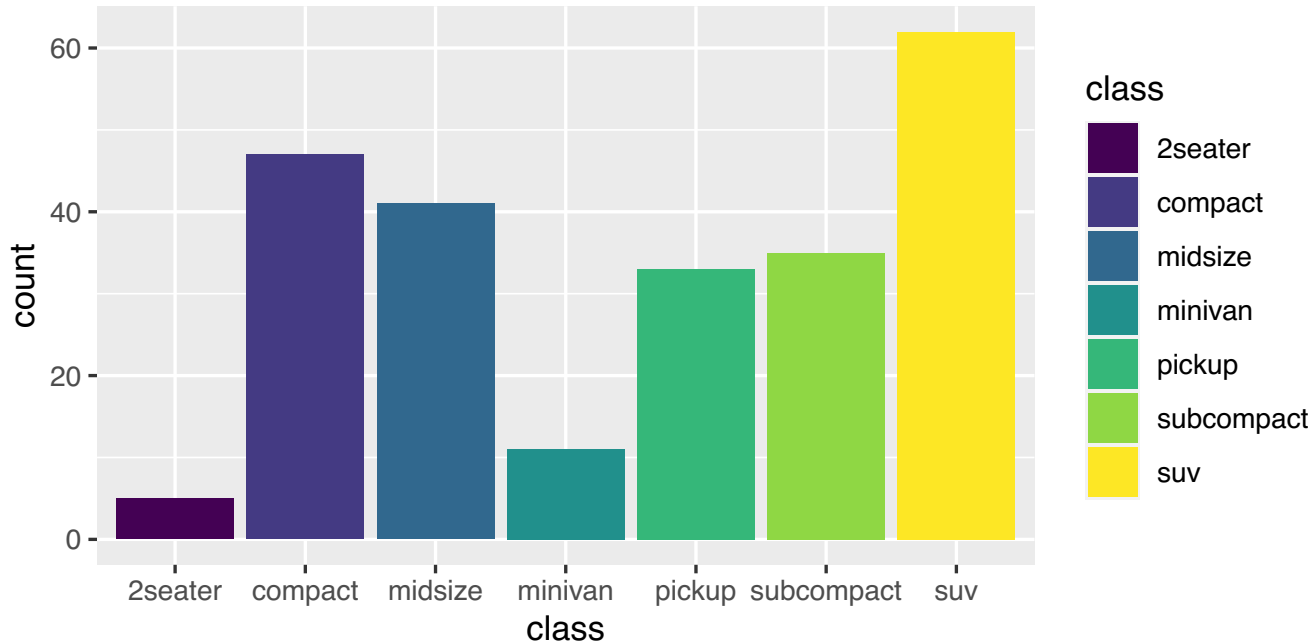
Colours

```
> ggplot(data = mpg) +    geom_bar(mapping = aes(x = class,
+                                                colour = clas
```



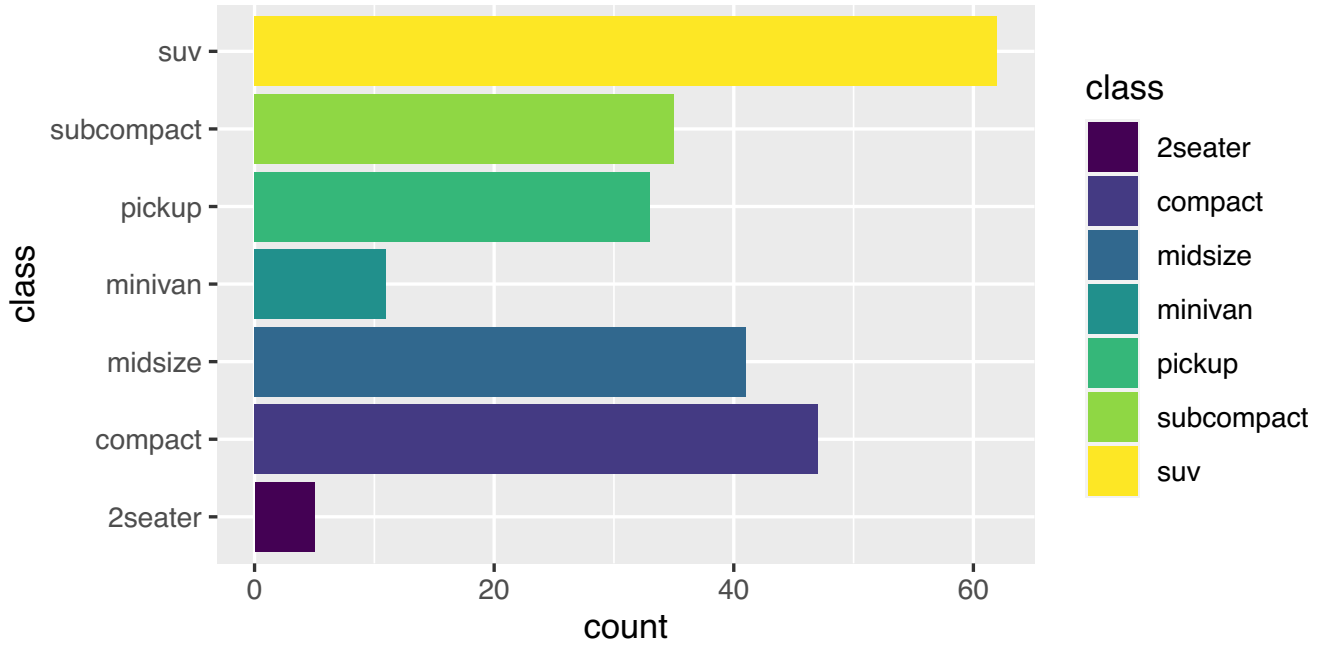
Colours

```
> ggplot(data = mpg) +
+   geom_bar(mapping = aes(x = class, fill = class)) +
+   scale_fill_viridis_d()
```



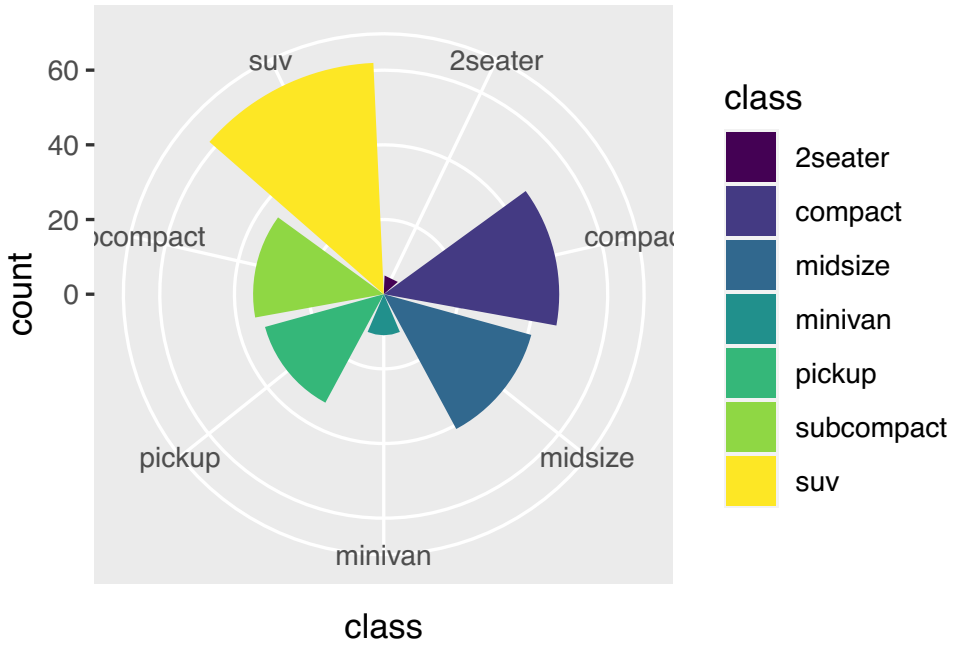
Colours

```
> ggplot(data = mpg) +
+   geom_bar(mapping = aes(x = class, fill = class)) +
+   scale_fill_viridis_d() + coord_flip()
```



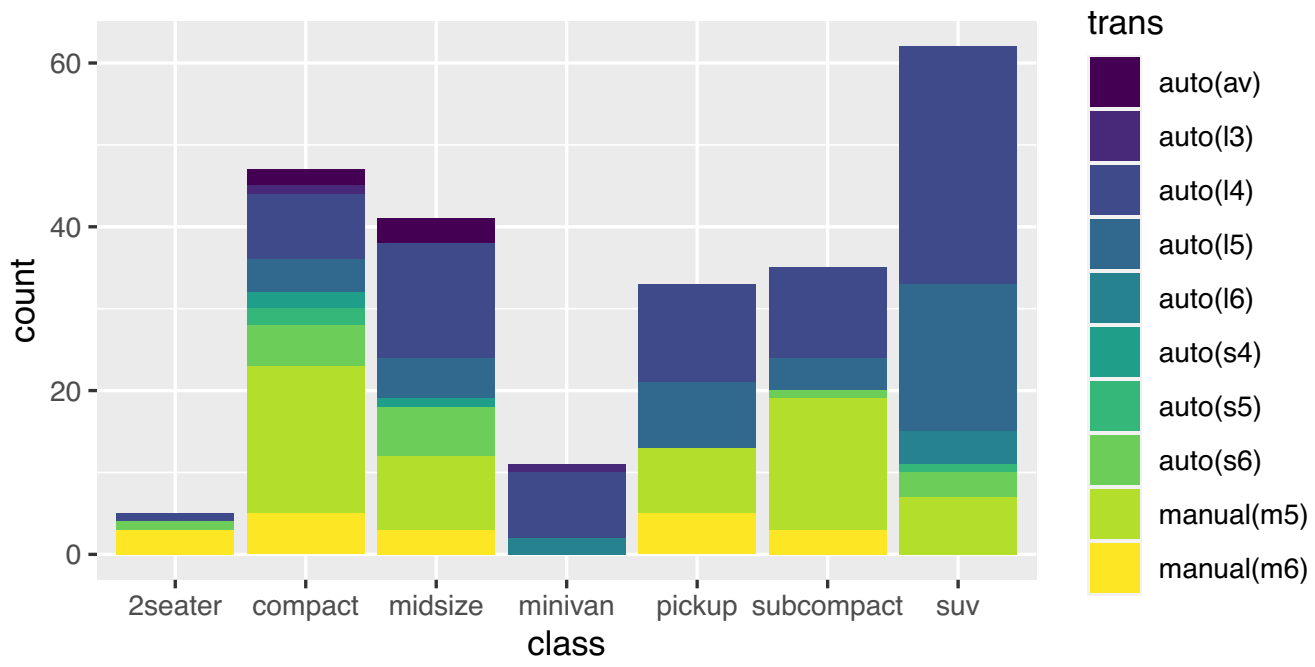
Colours

```
> ggplot(data = mpg) +
+   geom_bar(mapping = aes(x = class, fill = class)) +
+   scale_fill_viridis_d() + coord_polar()
```



Stacked

```
> ggplot(data = mpg) +
+   geom_bar(mapping = aes(x = class, fill = trans)) +
+   scale_fill_viridis_d()
```

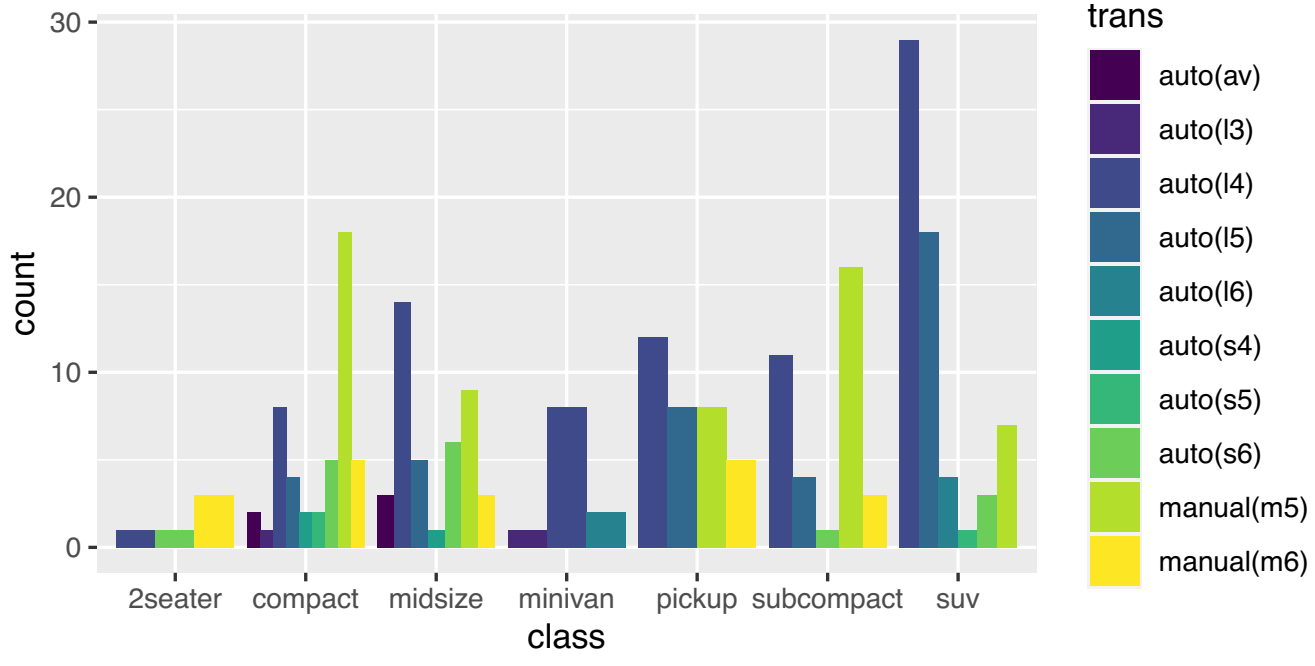


Stacked

position = stack, fill, identity ... check help

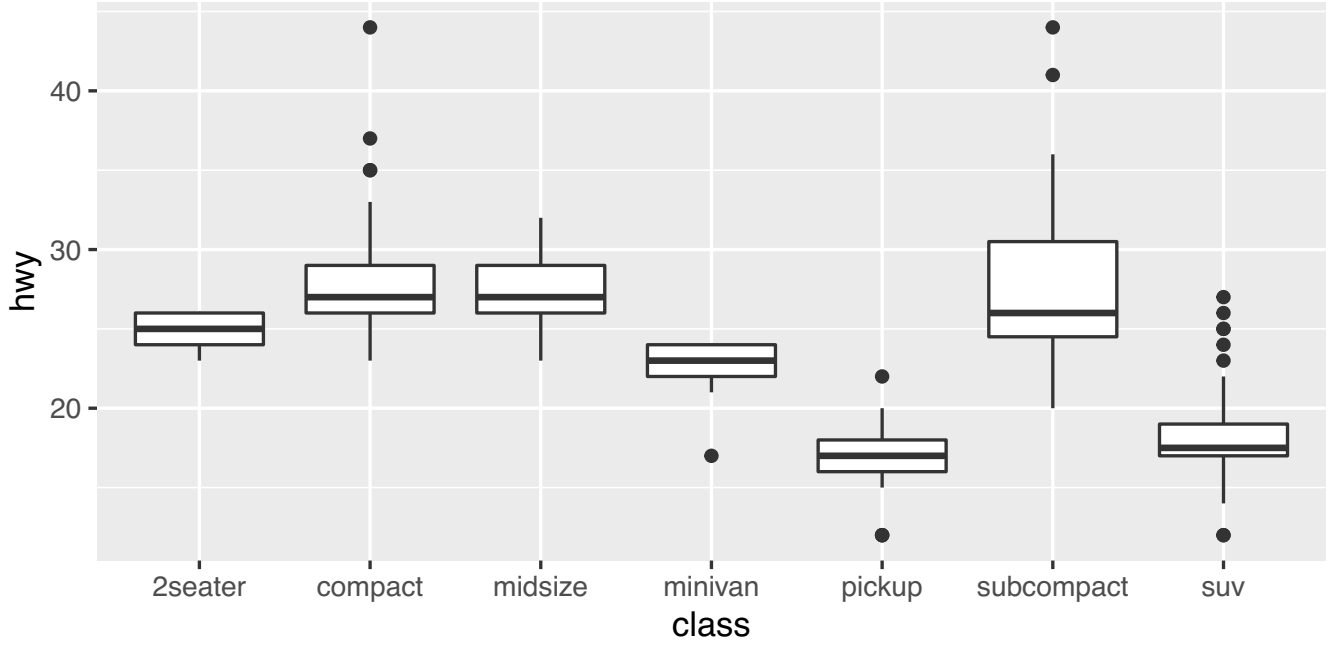
position = "dodge"

```
> ggplot(data = mpg) +
+   geom_bar(mapping = aes(x = class, fill = trans), positi
+   scale_fill_viridis_d()
```



BoxPlot

```
> ggplot(data = mpg, mapping = aes(x = class, y = hwy)) +
+   geom_boxplot()
```



Boxplot

discuss in detail in next class

BoxPlot-Coordinates Flipped

```
> ggplot(data = mpg, mapping = aes(x = class, y = hwy)) +
+   geom_boxplot() +
+   coord_flip()
```