Recall : Data visualisation - ggplot2

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

- ggplot(data = mpg) → coordinate system
- geom_point → geom function
- mapping argument is paired with an aesthetic.

- add layers = colour, fill

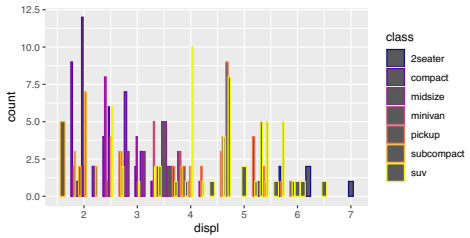
- Scatter plot ≡ geom_point
- bar plot ≡ geom_bar

Today
- histogram
- line chart (worksheet)
- box plot

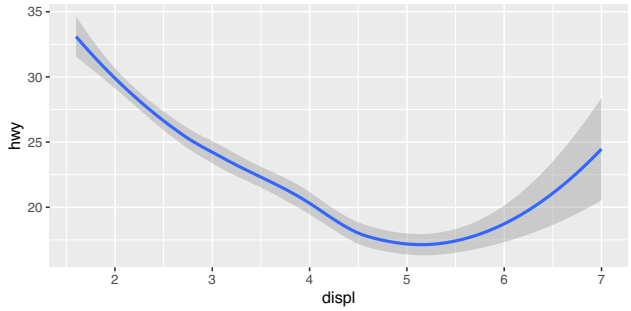
geom-

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

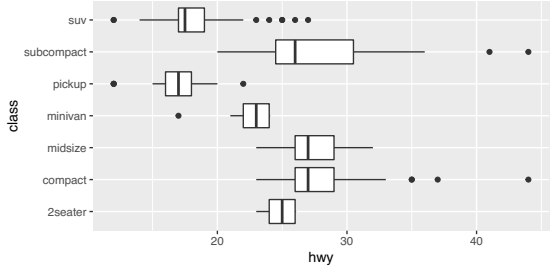
bar charts: `geom_bar`,



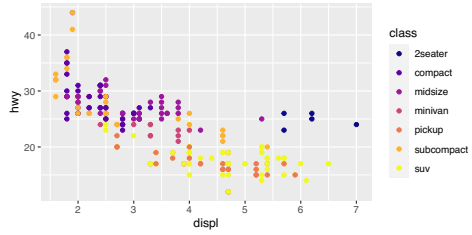
line-charts:`geom_smooth`,



box-plot:`geom_boxplot`,



Scatter:`geom_point`,



Bar Charts for Categorical Data – Recall

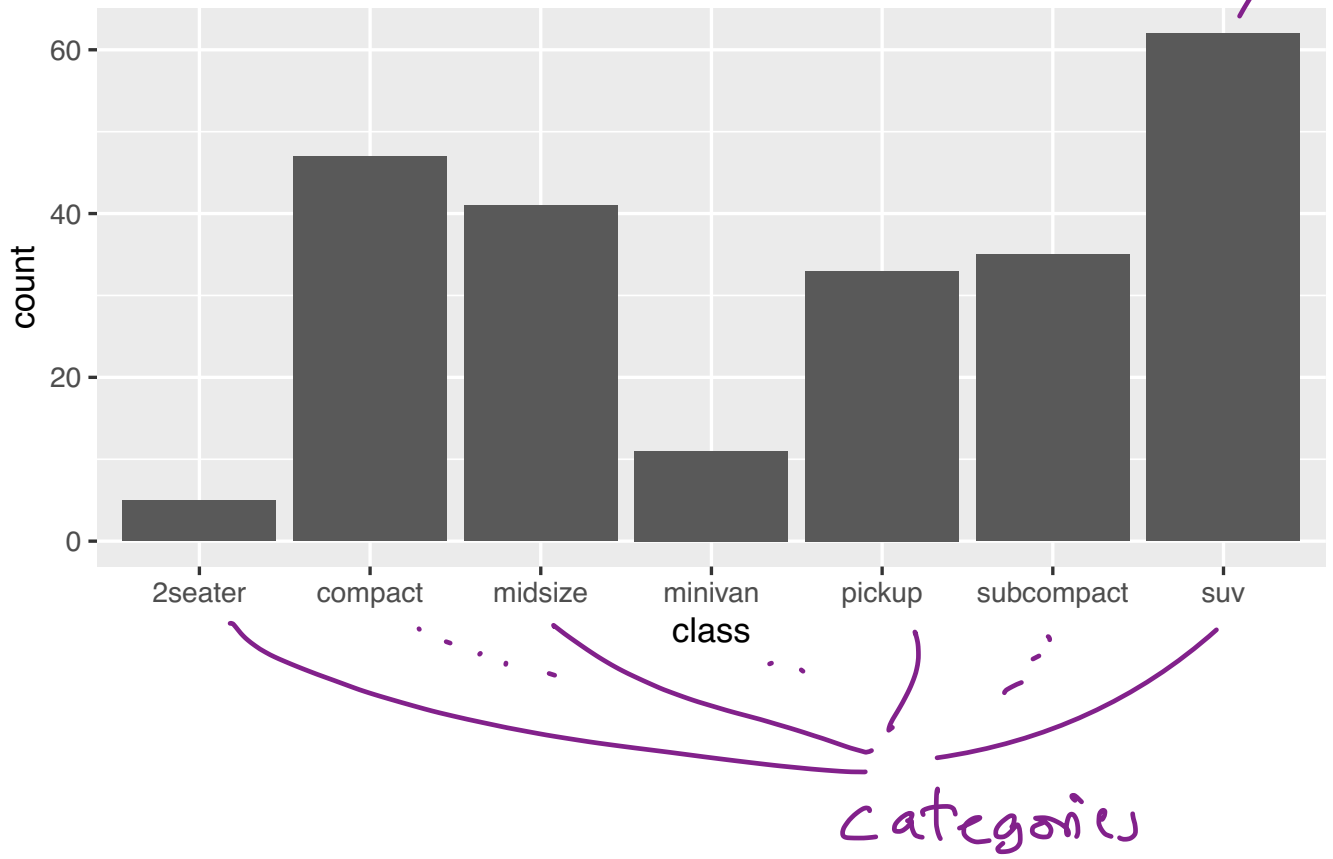
- A bar chart is a graph where for each category a bar with a height proportional to the count in the respective category is drawn.

- Along x-axis the categories (or levels) are displayed.

Stat-count – Bar Charts

Recall

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



numeric value

- count
- proportion
- value

Using

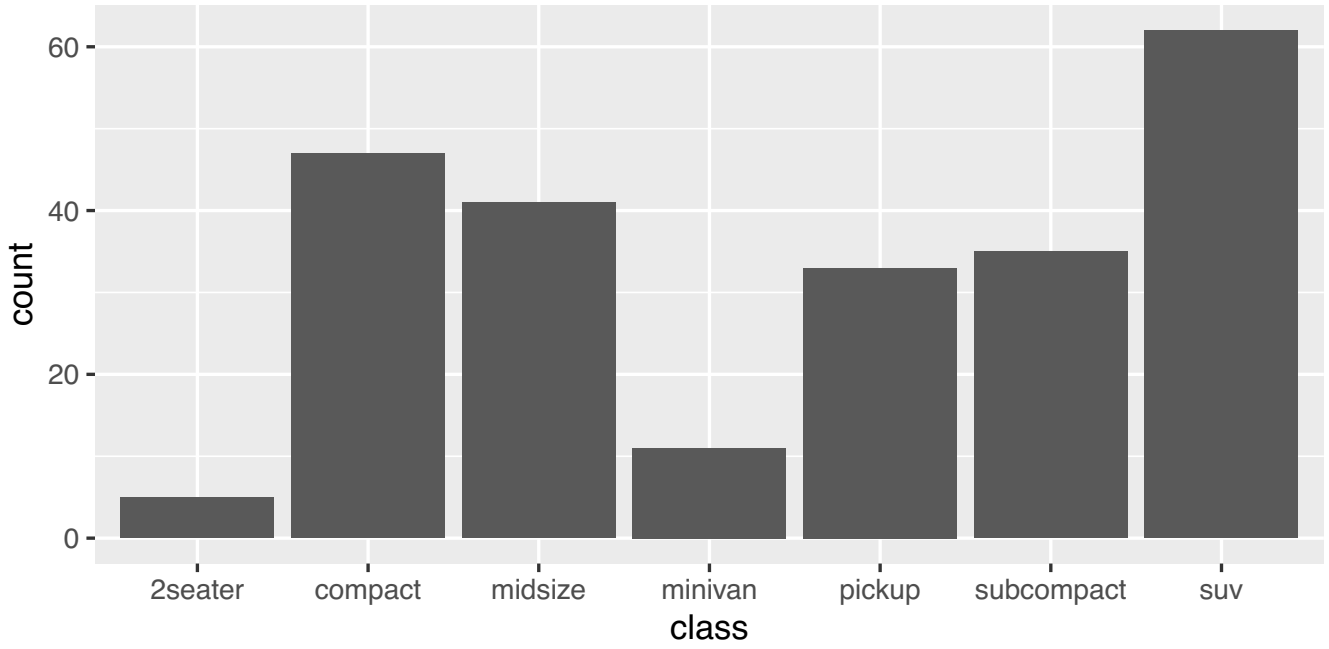
- fill..
- group..(*)

categories

Stat-count – Bar Charts

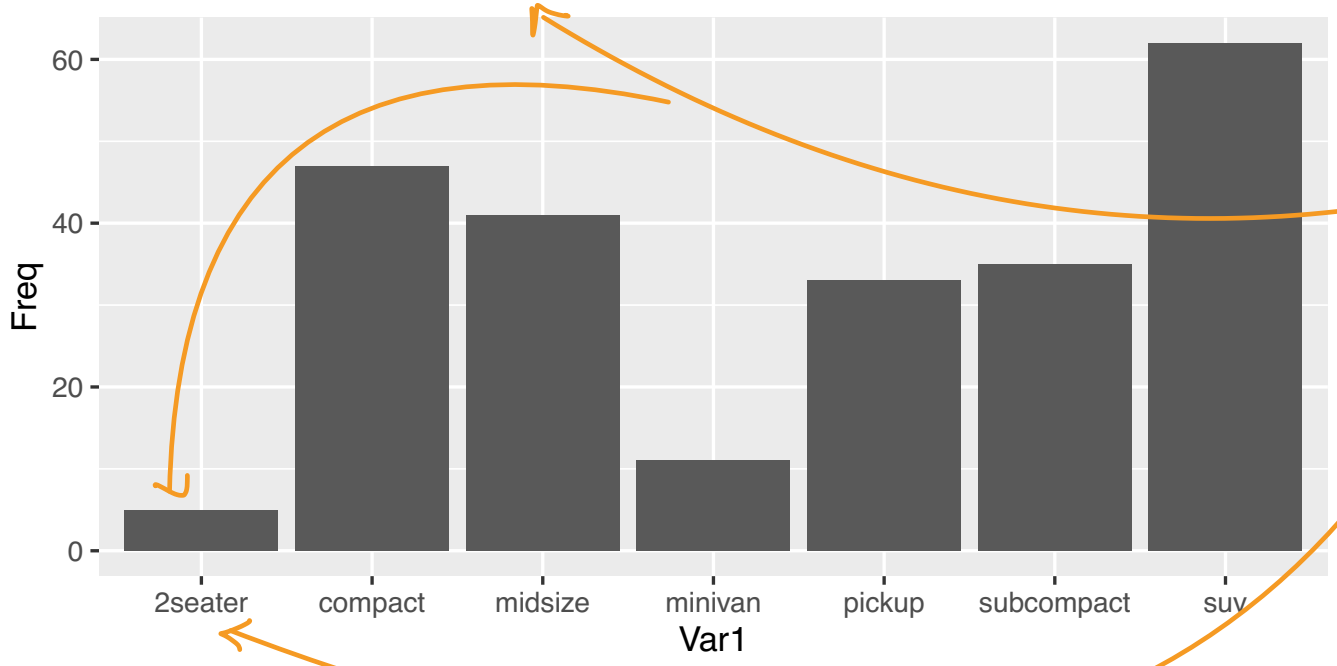
statistical transformation.

```
> 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, y=Freq),
+   stat="identity")
```



frequency table

data frame

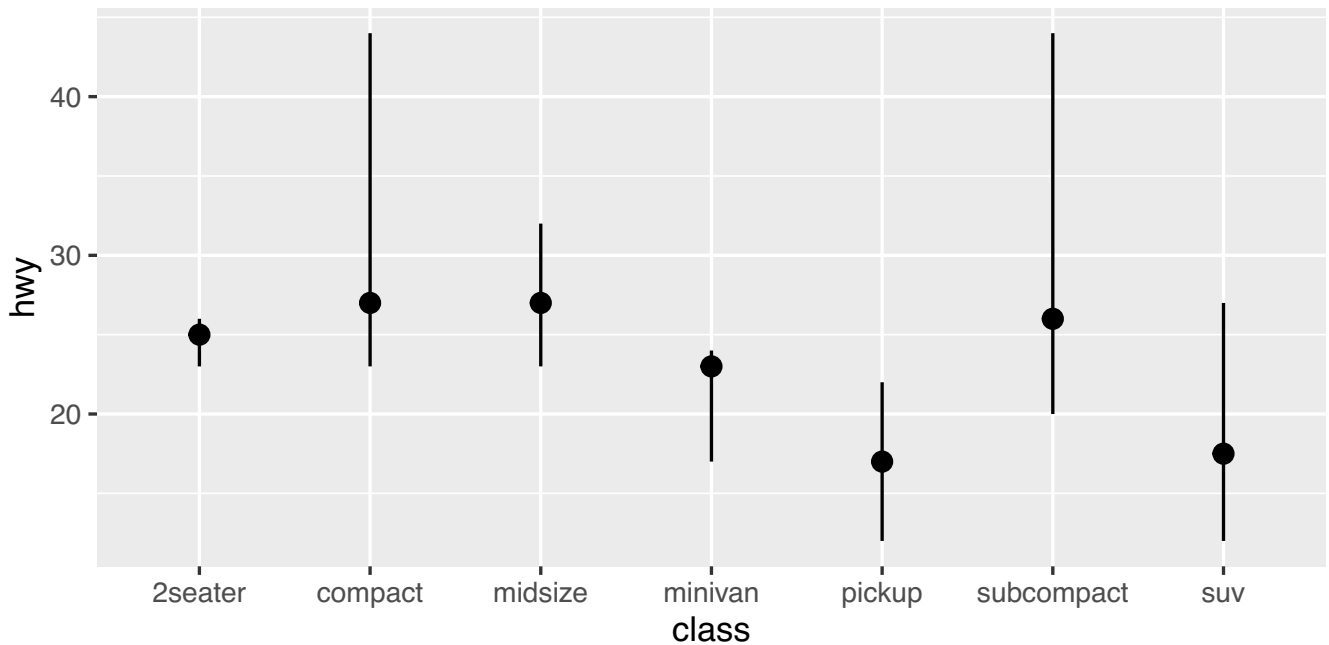
Var1 Freq

2 seater 5

Summary

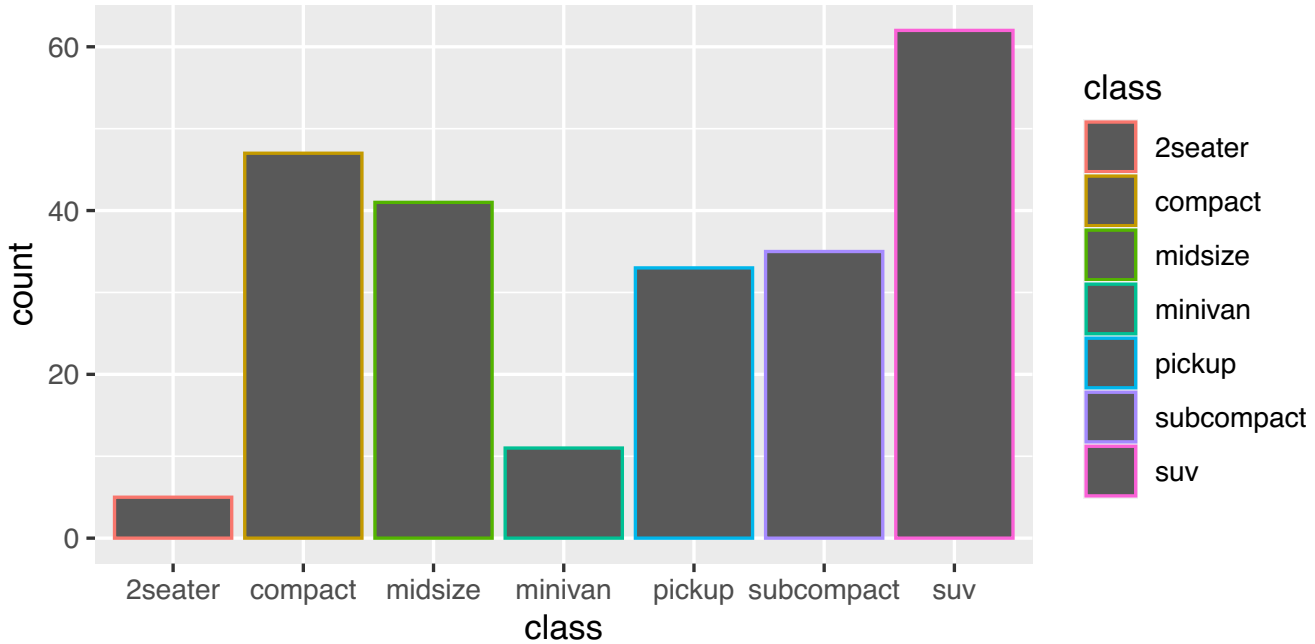
— Recall — stat function

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



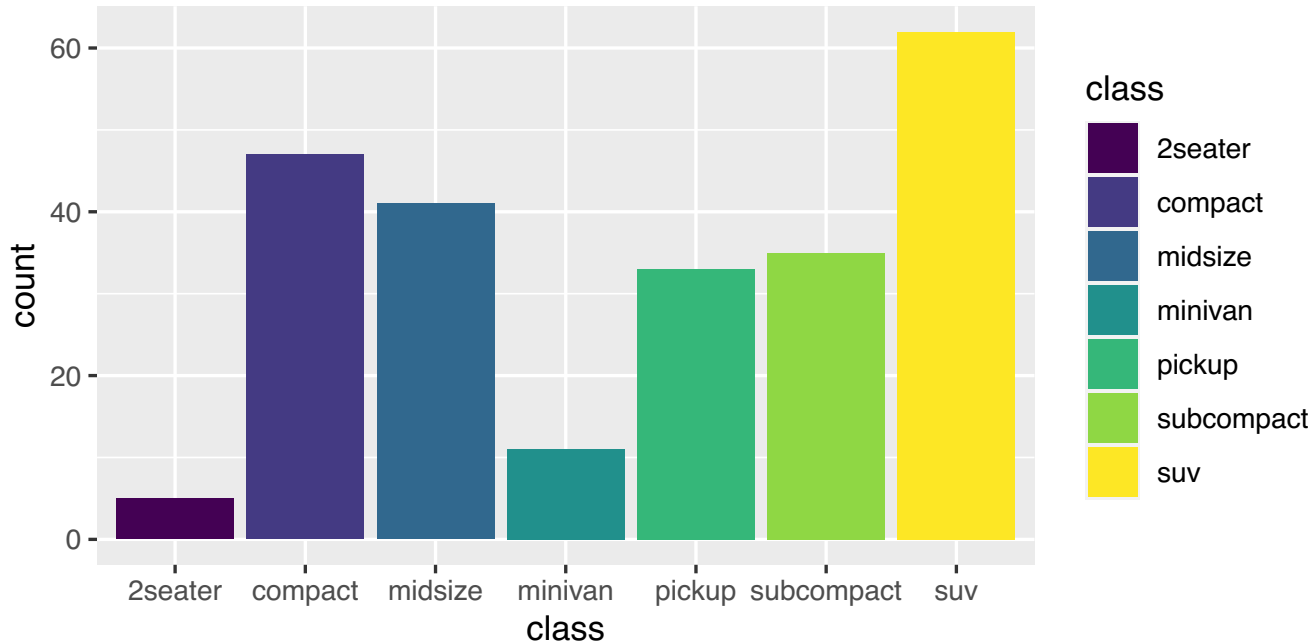
Colours

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



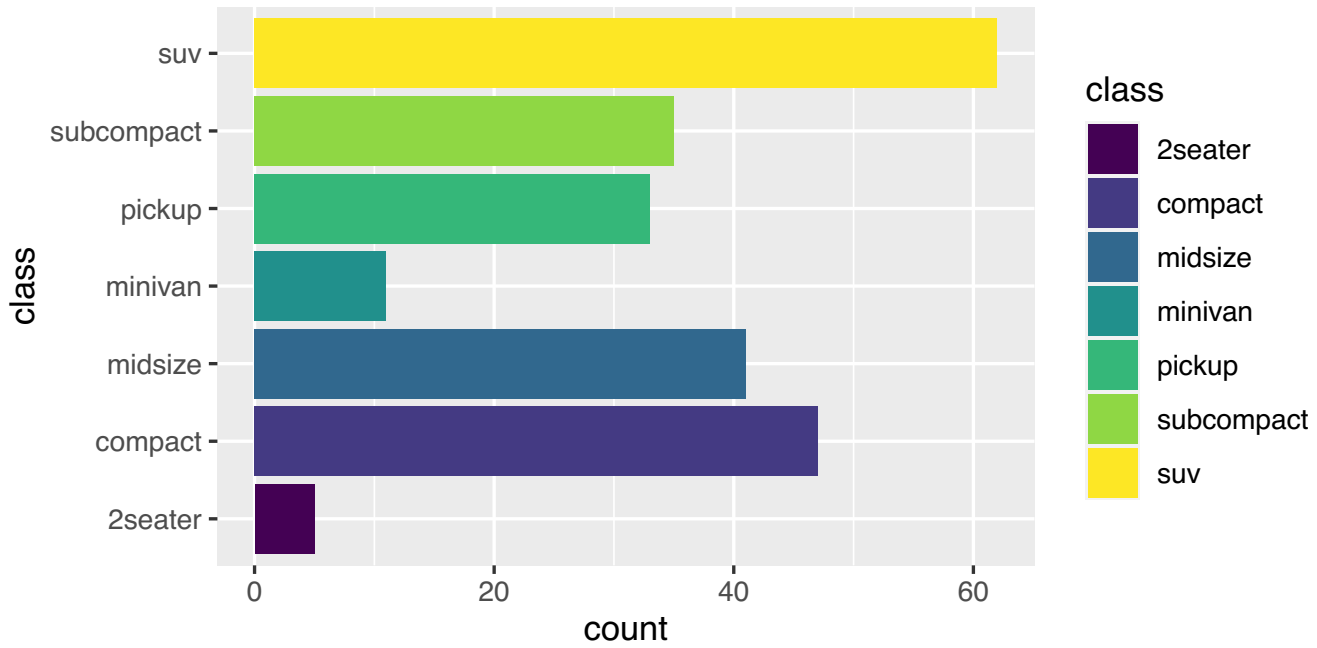
Colours

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



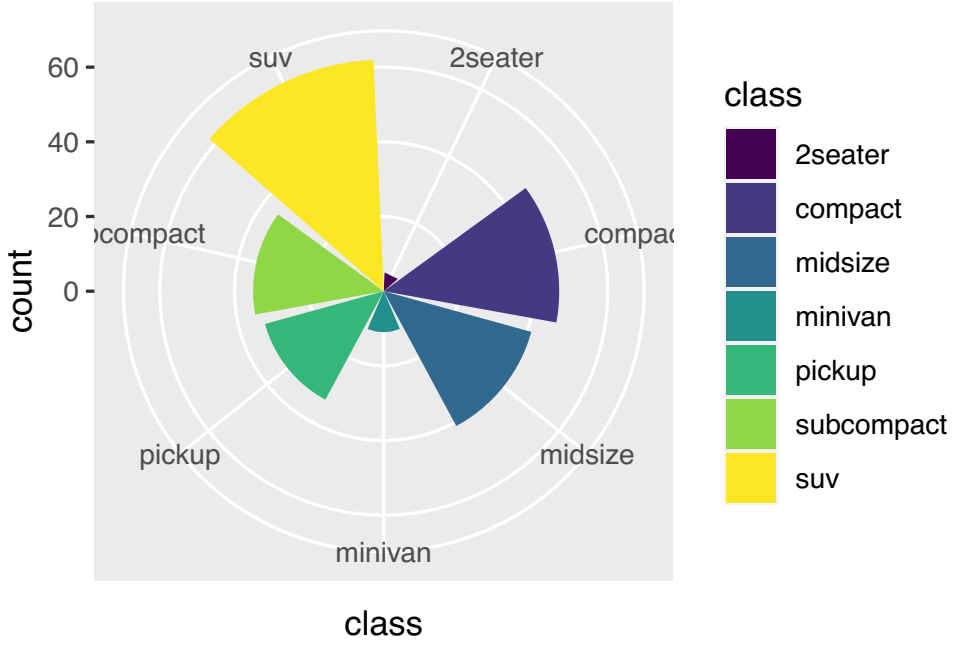
Colours

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



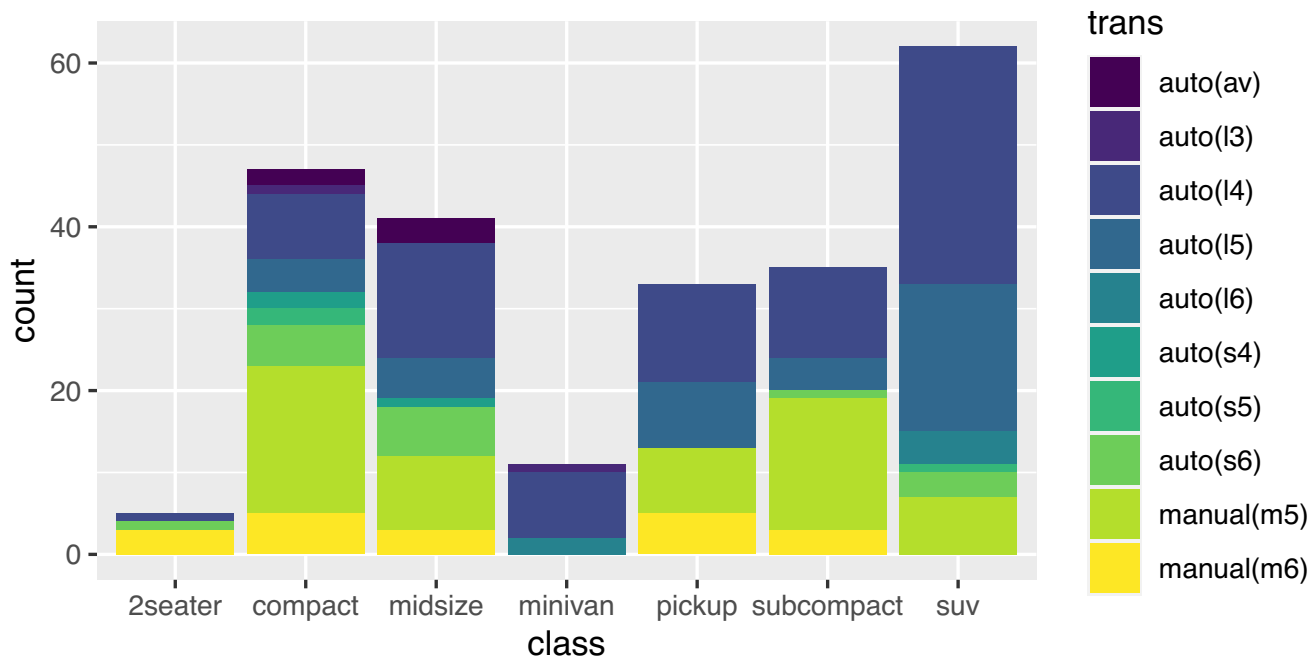
Colours

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



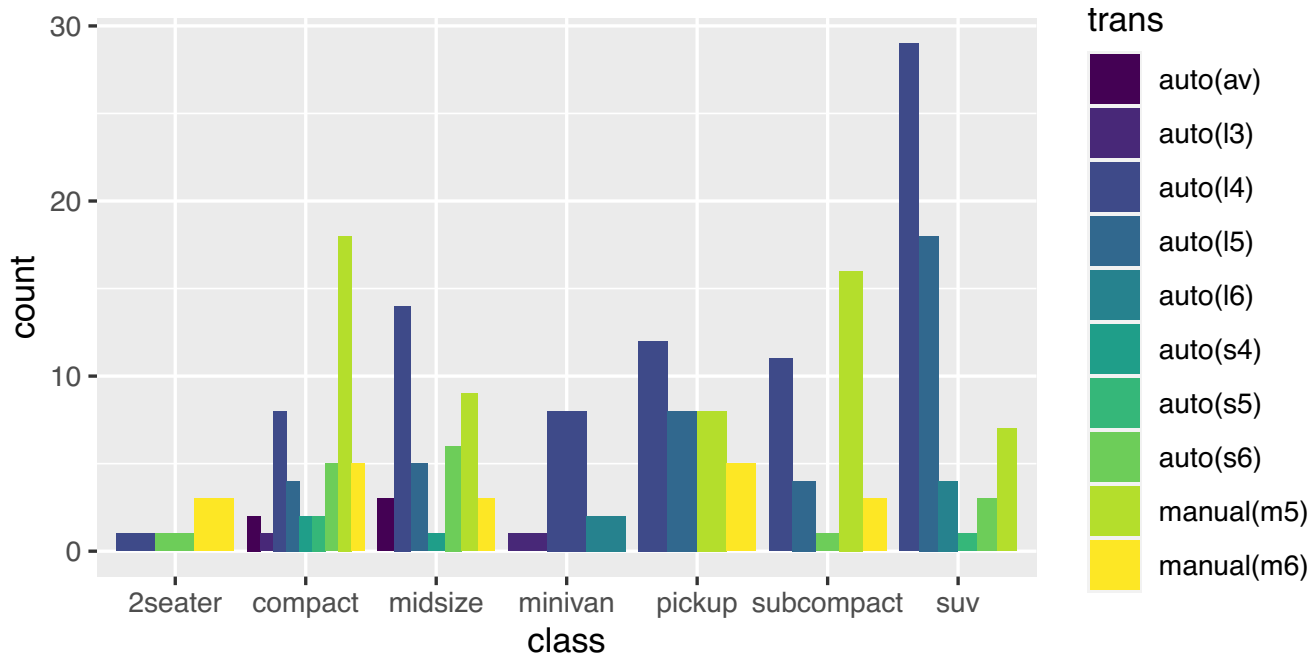
Stacked

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



Stacked

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

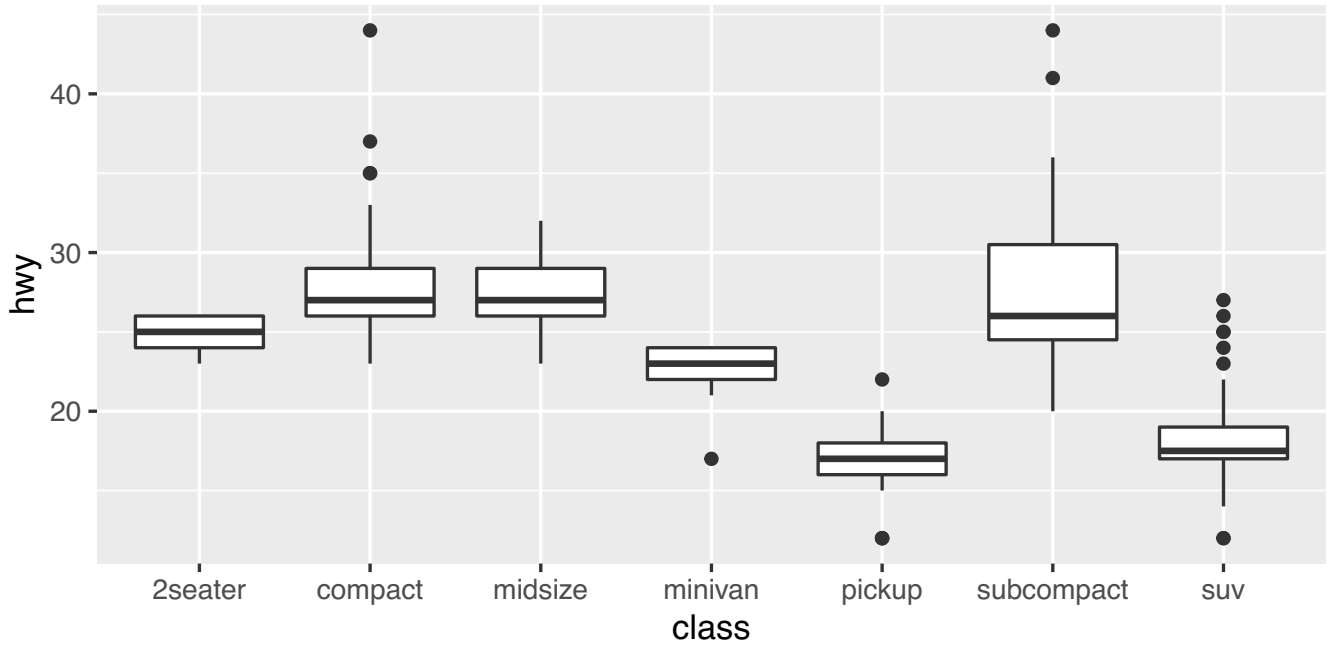


Box-Plot – common plot that is used

- The boxplot is used to summarize data using the five number summary.
- From the display one can check easily if the data is symmetric or has suspected outliers.
- Its simplicity is its feature.

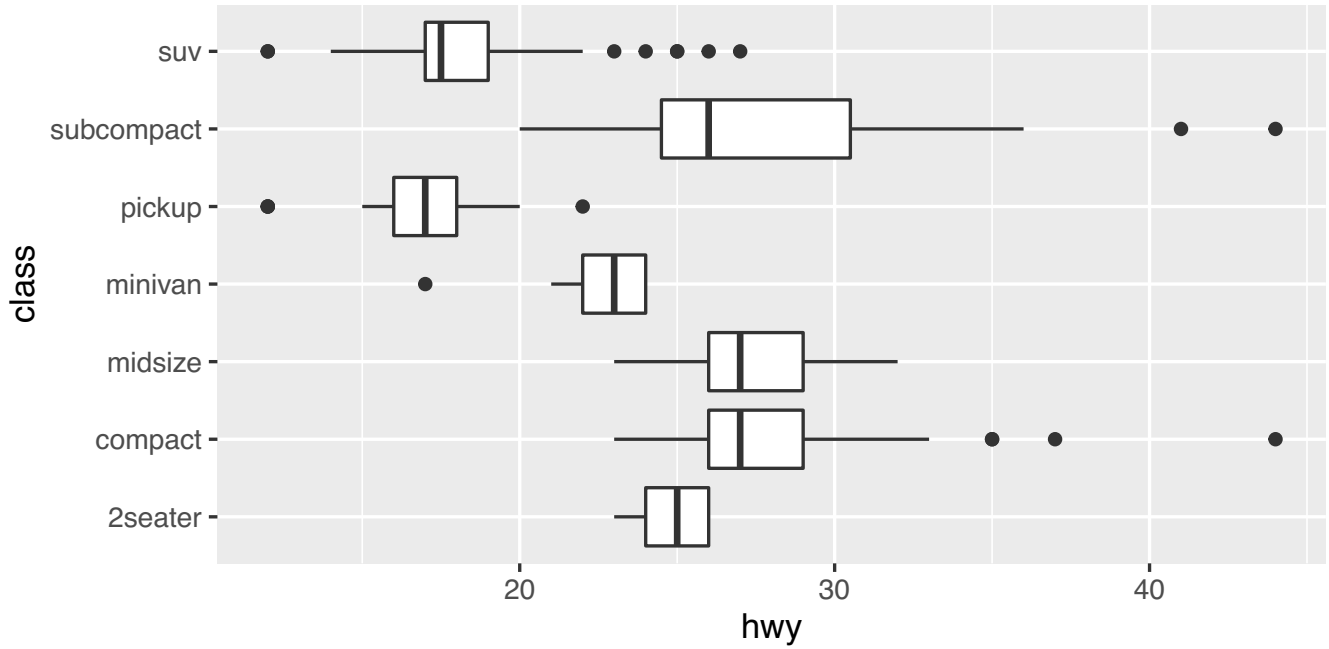
BoxPlot

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



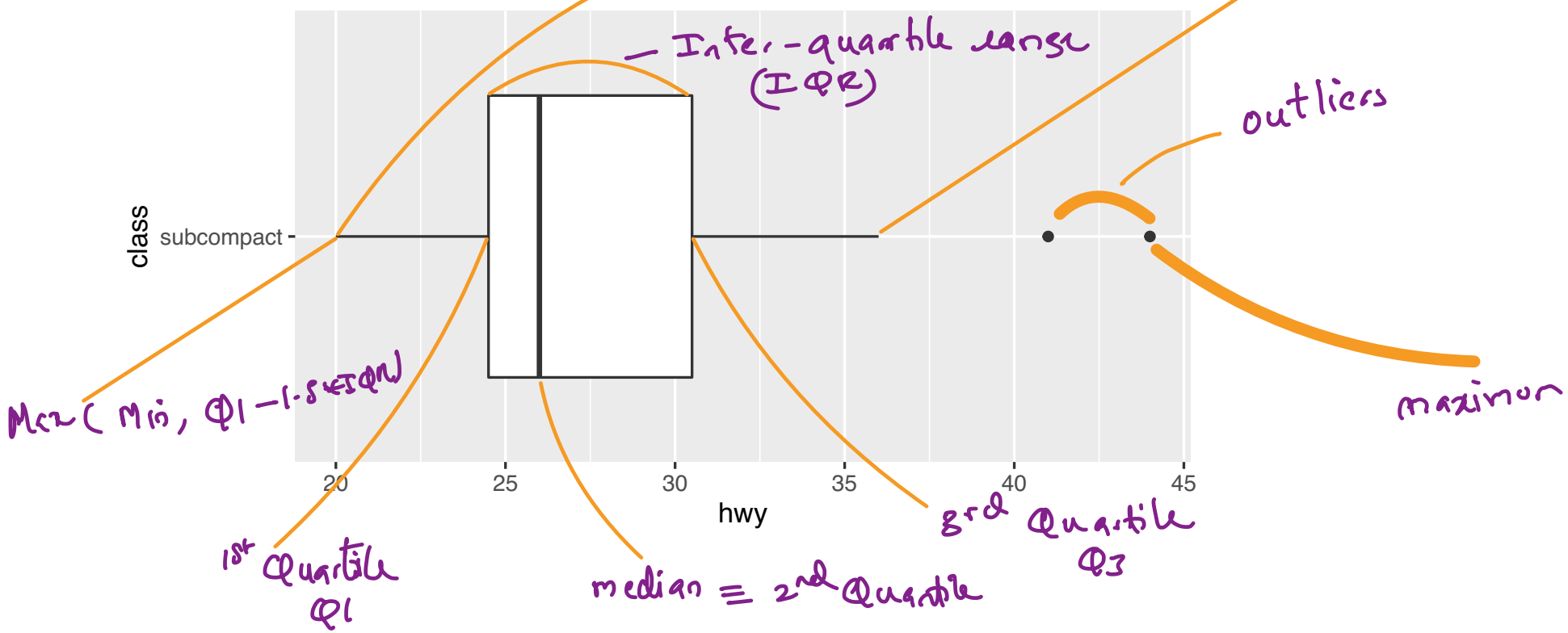
BoxPlot-Coordinates Flipped

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



BoxPlot- hwy for subcompact

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

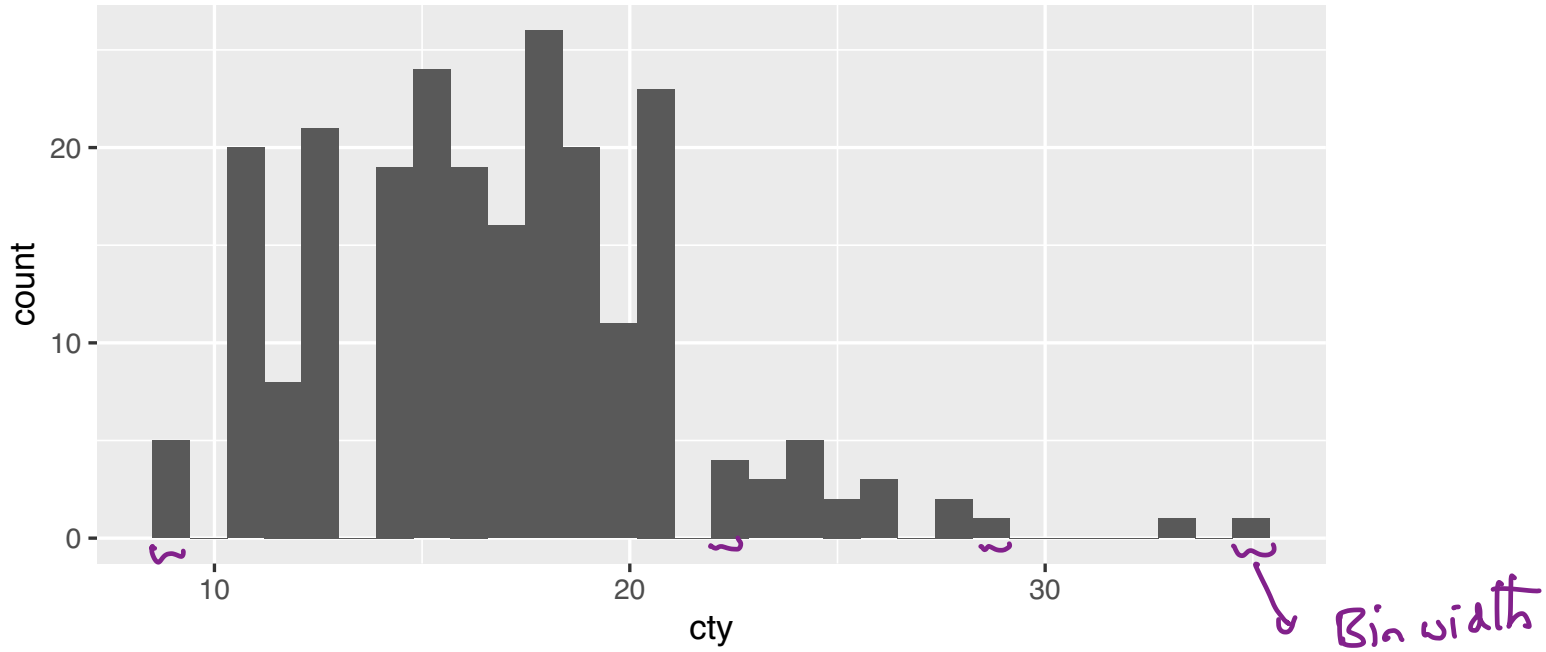


Histogram

- First specifies a sequence of points, called breaks.
- It counts the number of observation between the breaks, called bins.
- Places a bar in each bin with
  - base being the length of the bin and
  - height being either the frequency or proportion of observations in the bin.

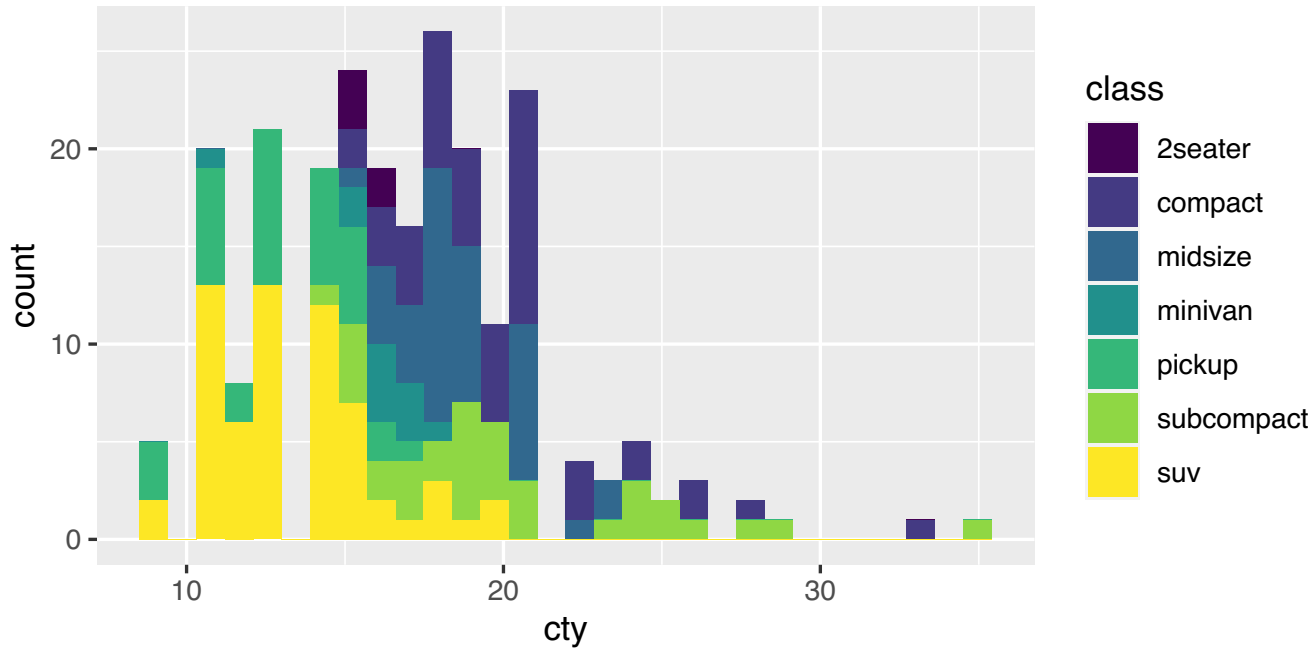
Histogram

```
> ggplot(data = mpg,  mapping = aes(x = cty)) +
+   geom_histogram()
```



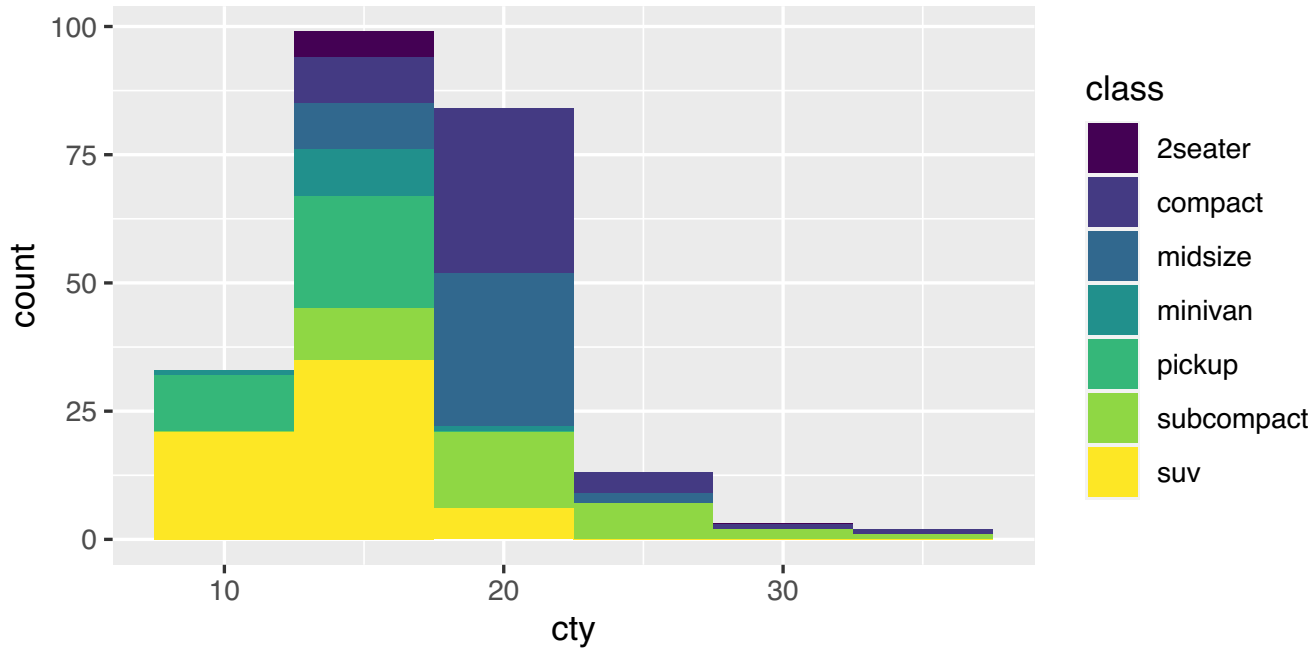
Histogram

```
> ggplot(data = mpg,  mapping = aes(x = cty,fill=class)) +
+   geom_histogram() +
+   scale_fill_viridis(discrete = TRUE)
```



Histogram

```
> ggplot(data = mpg,  mapping = aes(x = cty,fill=class)) +
+   geom_histogram(binwidth = 5) +
+   scale_fill_viridis(discrete = TRUE)
```



Histogram

```
> ggplot(data = mpg,  mapping = aes(x = cty,fill=class)) +
+   geom_histogram(binwidth = 5) +
+   scale_fill_viridis(discrete = TRUE)
```