Datasets in R

R has a lot inbuilt Datasets that one can use. The command :

```
> data()
```

will list currently installed data sets.

Datasets in R

- R stores many datasets as data frame (often).
- A data frame is a rectangular collection of variables (in the columns) and observations (in the rows).

airquality in R

Let us learn about real data stored as data frame.

```
> ?airquality
```

airquality in R

Let us learn about `airquality` dataset a bit more.

- we could print the entire data set on the screen

  `>airquality`

  but this is too much information.

- Let us try the `head()` function

  `> head(airquality)`

  This provides the first six rows.

airquality in R

Let us learn about `airquality` dataset a bit more.

- Let us try the `tail()` function

  `> tail(airquality)`

  This provides the last six rows.

airquality in R

- Below provides the first ten rows.

```
> head(airquality, n = 10)
```

- Data can be called using row and column number

```
> airquality[148,4]

[1] 63
```

- We can use the variable name for the given column and call it by its position.

```
> airquality$Temp[148]

[1] 63
```

airquality in R

- Provides an entire row

  ```
  > airquality[148,]
  ```

- Provides `Ozone Temp` columns

  ```
  > airquality[,c(1,4)]
  ```

  using `c()` function we can form any vector and that will enable display of the respective columns. We did not specify the row, so all rows will be displayed.

Five Number Summary and Histograms

```
> summary(airquality$Temp)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  56.00   72.00   79.00   77.88   85.00   97.00
> hist(airquality$Temp)
```
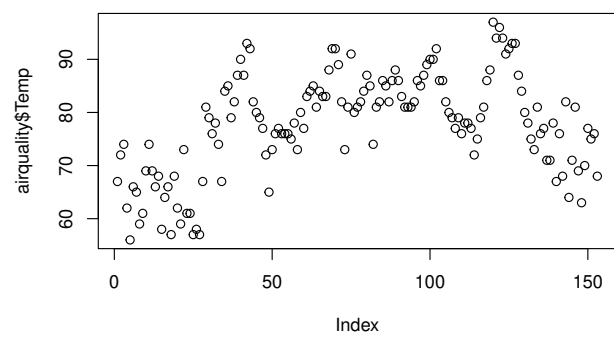
Histogram of airquality$Temp
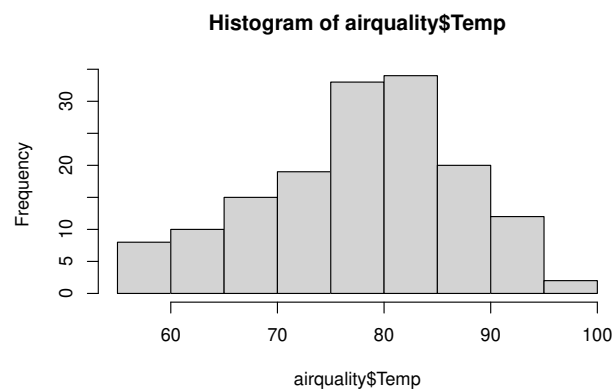
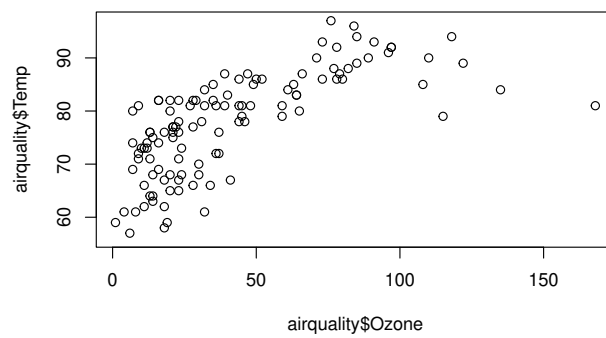# Plot

We can use the `plot` function to just plot.
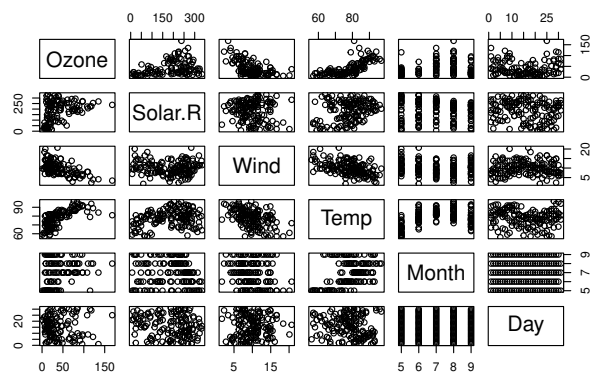
```
> plot(airquality$Temp)
```

Scatter Plot

We can use the `plot` function to get a Scatter plot.

```
> plot(airquality$Ozone, airquality$Temp)
```

Plot!

```
> plot(airquality)
```

External Packages in R

R has can be enhanced with a lot of external packages that are available. The package `UsingR` has many datasets loaded in it.

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

Once installed then to add to current workspace

```
> library("UsingR")
```

ggplot2- Data Visualisation

ggplot2 implements grammar of graphics

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

Once installed then to add to current workspace

```
> library("tidyverse")
```
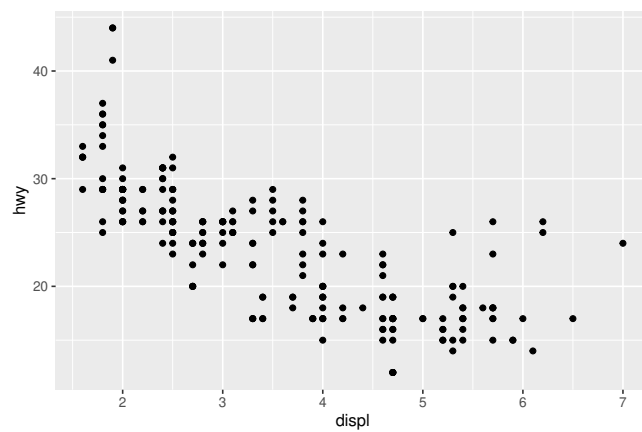
ggplot2- Data Visualisation

Dataset in `tidyverse`

`> mpg`

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

ggplot2- Data Visualisation

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


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

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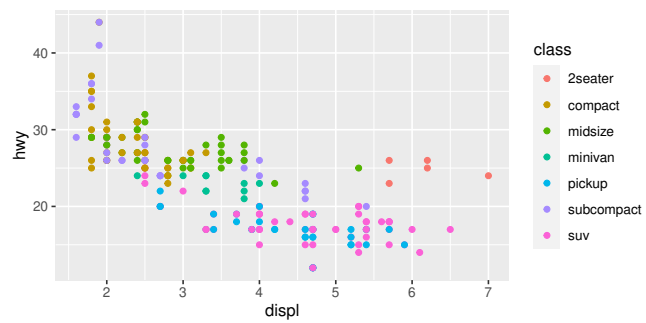

Added a third variable called `class` to a 2−D scatter plot by mapping it to an *a*esthetic.

ggplot2- Scaling

- Map an aesthetic to a variable
- Associate the `name` of the aesthetic to the name of the `variable.`
- Above example Name=colour and Variable=class.
- Scaling: `ggplot2()` will assign a unique level of the aesthetic `colour` to a unique level to the variable `class`.
- `ggplot2` will also add a legend explaining the levels
- Other aesthetics include : shape and size.

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/.
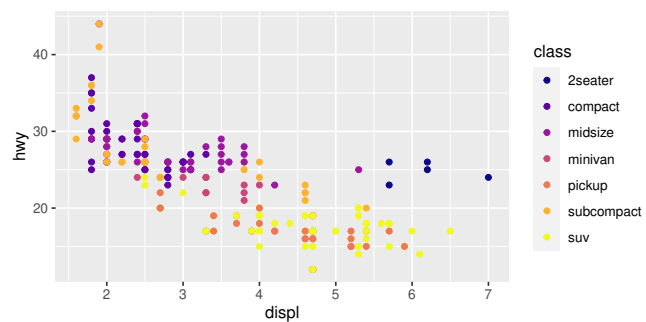
ggplot()-viridis options

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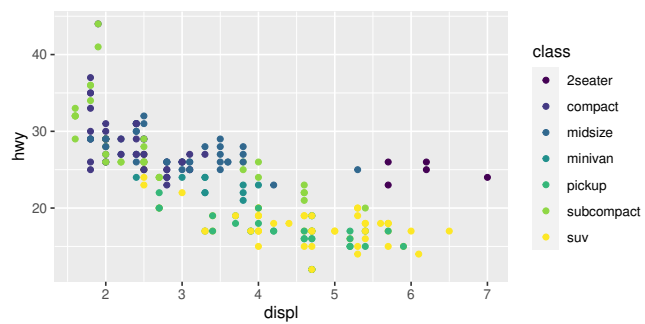

Using colour palette from viridis package (colour blind colours).

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).