

# Tutorial and Laboratories

## Week 10

The purpose of this week's tutorial and lab exercises is to get your hands on using ggplot2 to build different type of graphs:

The *objectives* of this tutorial and lab include:

- 1- Understanding the grammar of the visualisations in R.
- 2- Getting familiar with using the different geometrical functions in the ggplot2 library.

### The grammar of the data visualisation in R

To build a graph in R, you need to specify three components:

1. **Data**: the records and variables, which we need to represent with the graph to understand the relationship between them.
2. **Geometry**: the plot type, usually it is a function such as (scatterplot, boxplot, barplot, histogram, smooth density, etc.)
3. **Aesthetic mapping**: represents mapping the variables onto visual cues, such as the x-axis, y-axis, size, scale and color of the graph components.

The basic graph template with ggplot2 is as following:

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping=aes(<MAPPINGS>))
```

Also the following components can be added when needed:

- **Position adjustment** for locating each geometric object on the plot
- **Scale** (e.g., range of values) for each aesthetic mapping used
- **Coordinate system** used to organize the geometric objects
- **Facets** or groups of data shown in different plots

We will practice build different graphs using the following exercises:

### Prerequisites:

In the following exercises, we will use the “Fuel economy data from 1999 and 2008 for 38 popular models of car”, (mpg). It is a built in dataset with the basic packages of R. Get yourself familiar with this dataset by typing (?mpg) in the terminal window of R-Studio.

**Exercise 1**

1. Run `ggplot(data = mpg)`. What do you see?
2. Make a scatterplot of `hwy` vs `cyl`.
3. Change the colour of the points to discriminate each class of the car with different colour.
4. What happens if you make a scatterplot of `class` vs `drv`? Why is the plot not useful?

**Exercise 2**

- 1- Create a scatter plot between the engine displacement (i.e. `displ`) and highway miles per gallon (i.e. `hwy`).
- 2- Add a smooth line to show the correlation between the variables in the last graph
- 3- Create a bar chart to show the frequencies of the different classes of the cars.
- 4- Build a histogram for the distribution of the highway miles per gallon variable.
- 5- Change the ticks of the x-axis
- 6- Add a vertical line to represent the median value of the variable on the distribution

**Exercise 3**

- 1- Create a bar chart for the distribution of car classes
- 2- Change the fill of the bars to be aligned with the car type of drive
- 3- How to change the counts on the y-axis to percentages?

**Exercise 4 (unsupervised activity)**

- 1- Create a “row” of subplots, one for each car class, where each subplot shows a scatter plot for the displacement and highway miles per gallon variables.
- 2- Add two variables (year of manufacturing and number of cylinders) to the last graph, where each subplot still shows the relationship between the displacement and the gallons consumed per mile on the highways.