Introduction to Data Visualization — ggplot2

Christopher Maronga

March 26, 2019

Introduction to ggplot2

- ggplot2 is a powerful and awesome package implemented by **Hadley Wickham** for creating great and elegant graphics in R. The **gg** in ggplot2 stands for **Grammar of Graphics** an idea which describes a plot as being composed of *parts*.

- According to ggplot2 concept, every plot is divided into 3 parts i.e. **plot = data + Aesthetics + Geometry**

- **data** is the dataframe/dataset

- **Aesthetics** is the mapping driver for **x** and **y** variables. It is also used to control other plot features such as **color**, **shape**, **size** e.tc.

- **Geometry** this is the type of graphic output for the desired plot (barplot, histogram, boxplot e.t.c)

We are going to put focus on **ggplot2()** for building plots piece by piece in the next series of lectures.

Data Visualization types

The type of graphics to be created entirely depends on the format of the data as well the data types to be plotted. There are hundreds of visualization types that can be created using `ggplot2` package, some are beyond the scope of this lecture. We are going to focus on the major building blocks of all these visualizations.

Types of visualization:-

- One variable: Continuous
- One variable: Discrete
- Two variables: Continuous X, Continuous Y
- Two variables: Discrete X, Continuous Y
- Two variables: Discrete X, Discrete Y
- Graphical parameters

NOTE: Once you know how to plot all the above types of visualizations, you should feel confortable ploting other complex plots using similar concepts.