Sane's STATS Academy of Statistics

ggplot2

Enhanced Graphics Package

Sane's STATS Academy of Statistics

About ggplot2

- Package ggplot2 provides a method of creating innovative graphs based on graphical grammar
- There are four graphic systems in R currently.

Sane's STATS Academy of Statistics

Four Graphics Systems in R

1. The `base` Graphics System written by the Ross Ihaka included by default in every R installation
2. The `grid` graphics system written by Paul Murrell (2011)
3. The `lattice` graphics system written by Deepayan Sarkar (2008)
4. The `ggplot2` graphics system written by Hadley Wickham (2009)

Sane's STATS Academy of Statistics

Base Graphics

- We have already covered it in the previous sessions
- Composed of functions like plot(), boxplot(), barplot() etc.

Grid Graphics

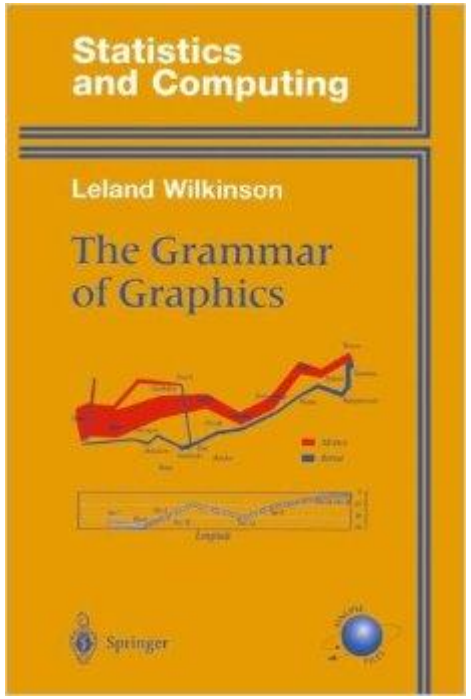
- Implemented by package `grid`
- Offers a low level alternative to the standard graphics system
- But the `grid` package doesn't provide functions for producing statistical graphics or complete plots

Lattice Graphics

- Implements *trellis* graphs with package `lattice`
- Provides a comprehensive system for creating statistical graphics
- Built using package `grid`

Ggplot2 Graphics

- The package *ggplot2* has been written by Hadley Wickham
- Provides a system for creating graphs based on the grammar of graphics described by Wilkinson and expanded by Wickham



Function qplot

- Before we look into the ggplot function, let us first have a look at the function qplot() (Quick Plot), which is a basic plotting function in package ggplot2.
- The function qplot() hides what goes on underneath (inside)

Syntax : qplot(x , y , data , …)

Where

x : variable to be considered on X-axis

y : variable to be considered on Y-axis (If not specified scatter plot won't be drawn)

data : data frame object

Function *ggplot*

- With ggplot function, the plots are created by putting together functions in a chain-like manner using plus (+) sign

Syntax :

ggplot(data,aes(x=,y=,…)) + geom function(s)

Where

data : data frame object

aes() : a function for specifying the role of variables

Sane's STATS Academy of Statistics

Basic Components of ggplot()

- **Data Frame**
- **Aesthetic mappings**: data mapping by color, shape etc.
- **Geoms**: Geometric object like points, shapes etc.
- **Facets** : Trellis plotting
- **Stats**: Statistical Transforms
- **Scales** : scale used by aesthetic map
- **Coordinate System**

Sane's STATS Academy of Statistics

Building Plots in ggplot()

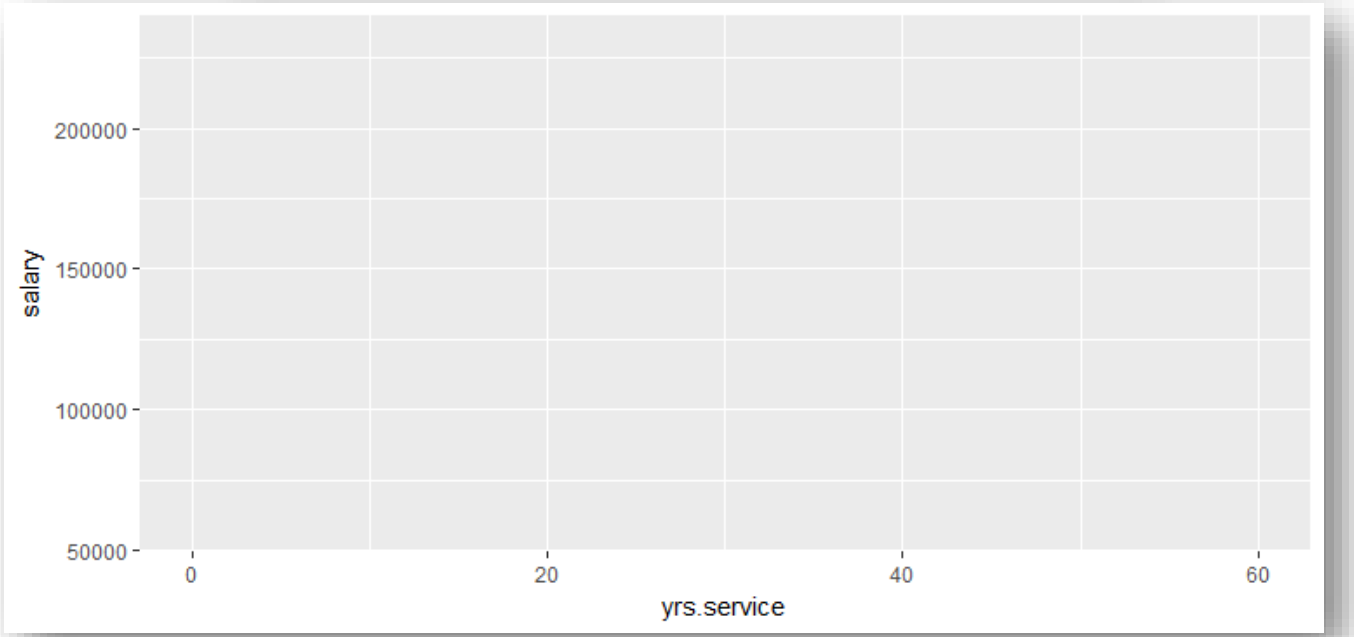
- Plots are built up in layers
  - Plotting the data
  - Overlaying the summary
  - Annotating the graph
- Let us have a simple example of displaying the scatter plot with *yrs.service* as X-axis and *salary* as Y-axis

Sane's STATS Academy of Statistics

Plotting by ggplot()

- We will find here that just specifying the dataset and the axes is not sufficient for generating graph.

```
p <- ggplot(Salaries,aes(yrs.service,salary))
print(p)
```



Sane's STATS Academy of Statistics

geom_*() functions

Geom	Graph Type
geom_point()	Scatter Plot
geom_line()	Line Graph
geom_histogram()	Histogram
geom_density()	Density Plot
geom_smooth()	Regression Line
geom_boxplot()	Boxplot
geom_bar()	Bar PLot