Data Visualisation using R

Center for Data Research and Analytics

Session: ggplot2

What is ggplot2?

- A data visualization library in R; it is based on "**grammar of graphics**"
- This is an implementation of the book "Grammar of Graphics" by Leland Wilkinson, this library implemented by Hadley Wickham in 2005
- It defines structuring mathematic and aesthetic elements into a meaningful graph
- It uses **layered concept of graphing** to build component by component in each layer, rather than producing premade graphics
- Users can create their own visualization based on their concept and **it is flexible enough** to create any type of graphs from the data

What is ggplot2? (cont.)

- It breaks the data visualization into semantic components such as:
  - Data
  - Aesthetic mapping
  - Geometric object
  - Statistical transformations
  - Scales
  - Coordinate system
  - Position adjustments
  - Faceting

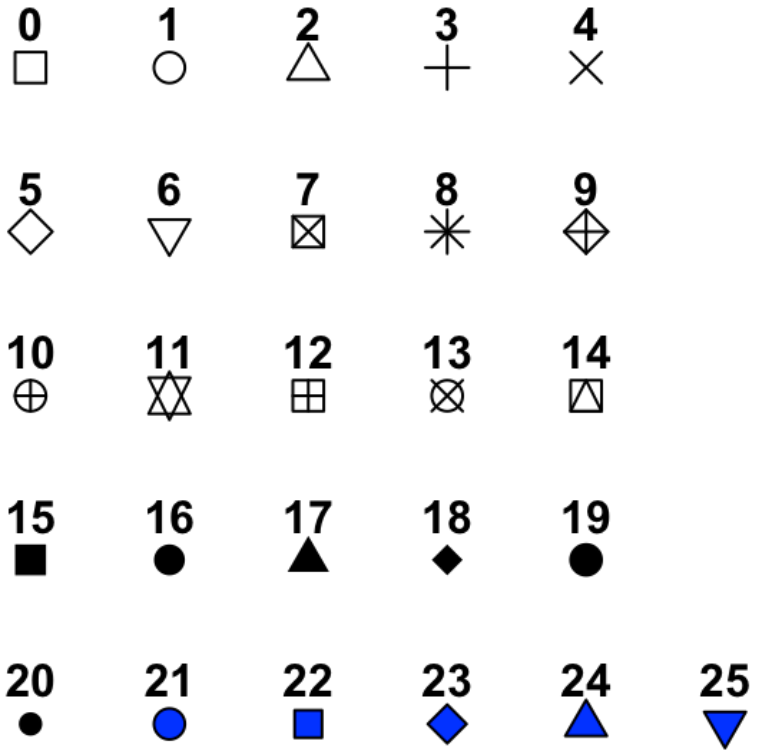
How ggplot2 works?

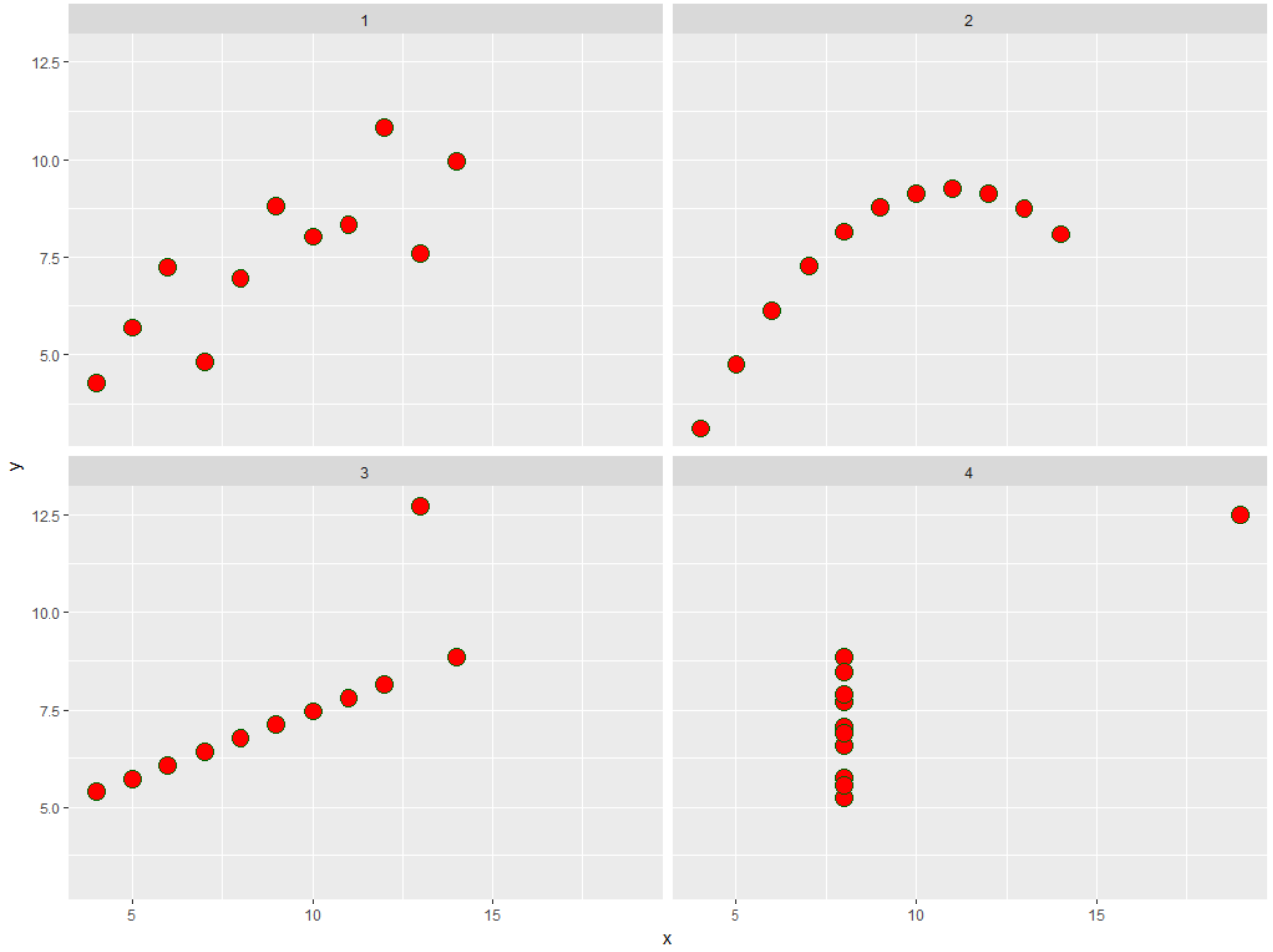
- It always take primary input as a **dataframe**, ggplot2 is not designed to take a vector input
- The necessary variables should be part of the dataframe
- User can add layer to enhance the visualization
- The first and mandatory function is ggplot(), without calling it you cannot create a plot in ggplot2 framework

```
ggplot(data = inputdataframe, aes(x = xaxis, y = yaxis))
```

Customize point properties

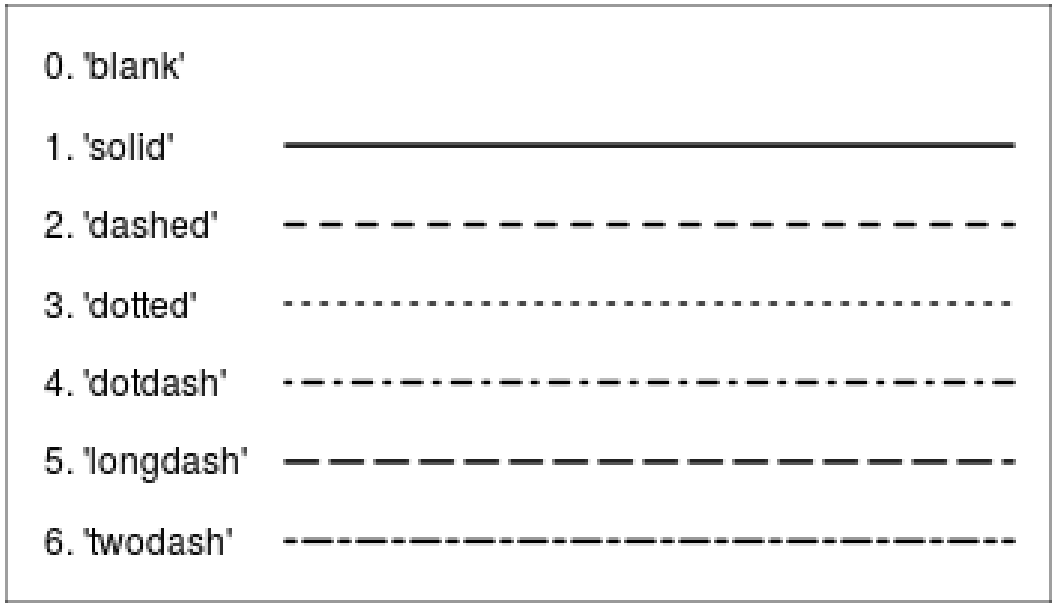
```
plot1 <- plot1 + geom_point(size = 5, fill = "red", color = "darkgreen", shape = 21)
print(plot1)
```

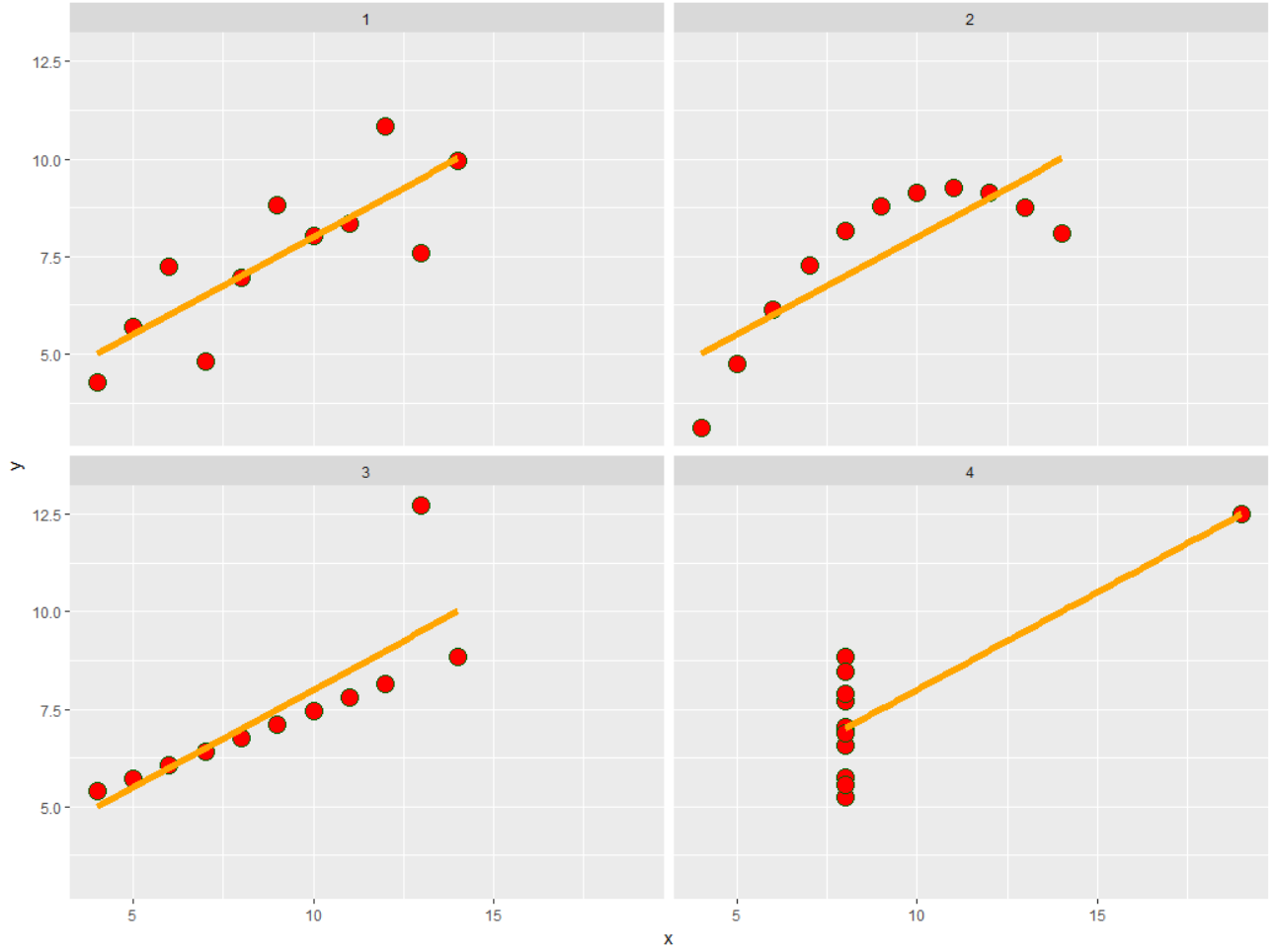




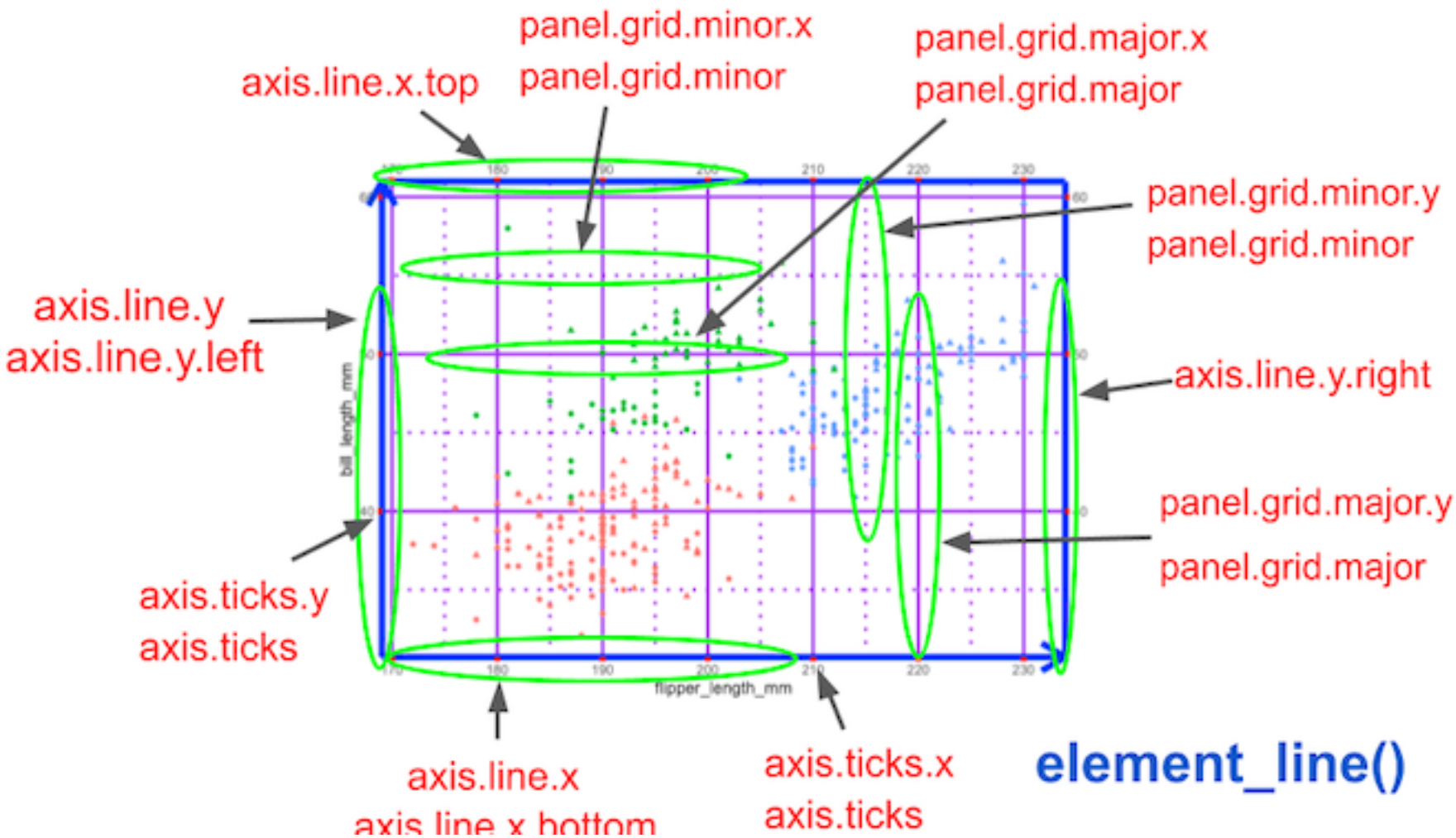
Customize line properties

```
plot1 <- plot1 + geom_smooth(method = "lm",
                             se = F, linetype = 1, size = 2, col = "orange")
print(plot1)
```





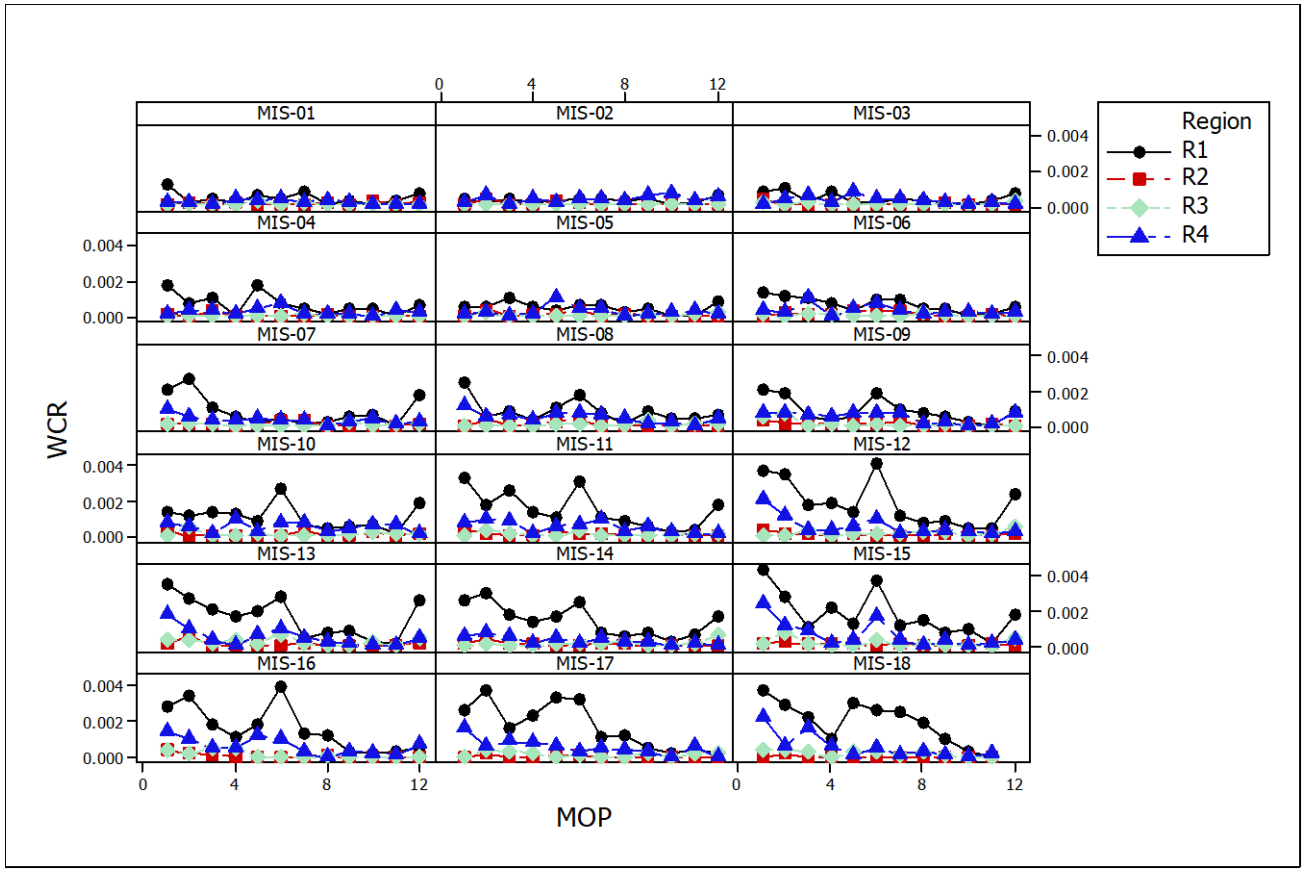
Anatomy of ggplot areas



Example Plot

- Aim is to create a graph like the following by R based on the four variables `MOP`, `WCR`, `MIS`, and `Region`

- The data are given in the csv file named `MOP_MIS_graph_data.csv`



Let’s follow the practical example