Experiment- 6

Working with larger data sets and introduction to *ggplot2* graphics

- Usually larger data sets are in the form of .csv or any other standard spreadsheet format. So in this experiment, we are going to deal with these files.
- First create a small .csv file having headers like **Sr., Name, Age, Gender, and Marks** of only 4 students and save in in your working directory (say **data.csv**)
- Now you can read and print the contents of **data.csv** using R program (say **expt6.R**)-

```
setwd("e:/temp")
fr = read.csv(file="data.csv")
print(fr)
```

Here **setwd(..)** is necessary to change the working directory where you have saved **data.csv .**

- Here check the **mode** and **class** of **fr** . Also try to display columns independently using command/statement **fr$Age**, **fr$Gender** for instance. Also check **mode** and **class** of these independent columns.
- Also observe and try to understand the output using one more argument in **read.csv(…, header=FALSE)**
- Now let's work with a larger data- here the file **lendingdata.csv** is already stored at http://famt.ac.in/eResource/it/lendingdata.csv . You can either use this URL as a file name in **read.csv(…)** function or you can download it in your working folder, and use only filename.
- Now view the data in frame **fr** inside R IDE itself (double click on environment variable **fr**).
- Know no. of records and columns in the frame **fr** using **ncol(..)** and **nrow(..)** function. Also explore few other methods used to work with data frame.
- Display the values of its columns independently; also check its **mode** and **class**.

- **ggplot2** is a data visualization package for the statistical programming. It uses semantic and grammar for visualization of graphics. It allows the user to add, remove or alter components in a plot using a language grammar.
- To use ggplot2 graphics, you require ggplot2 package to be installed using the command **install.packages("ggplot2")** at R IDE's prompt.
- After the package gets installed, you must import its library in the session using the function **library(ggplot2)** before using features of **ggplot2** for displaying the graphics.
- Here is a simple example to begin with **ggplot2.**

```
library(ggplot2)
ggplot(fr,aes(x=lender_count,y=loan_amount))+geom_point()+geom_smooth()
```

- Here **fr** is a dataframe obtained by reading .csv file, **lender_count** and **loan_amount** are variables(columns) of **fr**. **geom_point()** and **geom_smooth()** are used to set features of the graphics.
- Check additional arguments of **aes** function used to set some more features of the graph.
- For more explanation on **ggplot2**, visit- http://r-statistics.co/ggplot2-Tutorial-With-R.html

---