Foundations of Data Analytics
Data Frames in R

Dr.C. Sweetlin Hemalatha

Vellore Institute of Technology, Chennai

July 22,2020

- What is data frame?
- Data frame creation
- Adding column
- Removing column
- Subsetting in data frame

What is Data frame?

- Data frame is a table of information like an Excel sheet
- Contains rows and columns containing different pieces of information
- Each column may be in different data types

Id	Name	Marks
1	Ramu	50
2	Raju	40
3	Ravi	25

Table 1: An example data frame

Data frame creation

Data frame is created using data.frame() function

Example:
my_df <- data.frame(Id = c(1, 2, 3),
                    Name = c("Ramu","Raju","Ravi"),
                    Marks = c(50, 40, 25))

Adding column

Columns can be added in different ways as follows:

my_df$perf <- 0 initialize to 0

my_df$perf <- c("Very good", "good", "needs improvement")

my_df[["perf"]] <- c("Very good", "good", "needs improvement")

my_df["perf"] <- c("Very good", "good", "needs improvement")

my_df[,"perf"] <- c("Very good", "good", "needs improvement")

Removing column

Columns can be removed in different ways as follows:

my_df$perf <- NULL

my_df[["perf"]] <- NULL

mydf["perf"] <- NULL

my_df[,"perf"] <- NULL

my_df[3] <- NULL

my_df[,3] <- NULL

my_df <- subset(my_df, select=-perf)

Subsetting in data frame

Let us consider the data set `mtcars` available in R

```
#loading data set
data(mtcars)

#Subsetting Using Indexing
      #Retrieving subset of rows
      mtcars[4:10, ]

       #Retrieving subset of columns
      mtcars[ ,c(1,3 ]

#Subsetting using column names
mtcars[ ,c("mpg","cyl" ]
mtcars[c("mpg","cyl" ]
```

Subsetting in data frame (contd.)

```
#Using subset() function to select desired columns
subset1 <- subset(mtcars, select= c(mpg, cyl))

#subsetting rows that satisfies a condition on a column
subset2 <- subset(mtcars, mpg>18)

#subsetting rows that satisfies more than one condition on columns
subset3 <- subset(mtcars, cyl>5 & mpg>16)

#To exclude the columns mpg and cyl prefix it with - sign
subset4 <- subset(mtcars, cyl>5 & mpg>16, select=c(-mpg,-cyl))
```