Preparing data in R

ABHISEK MAITI

August 2022

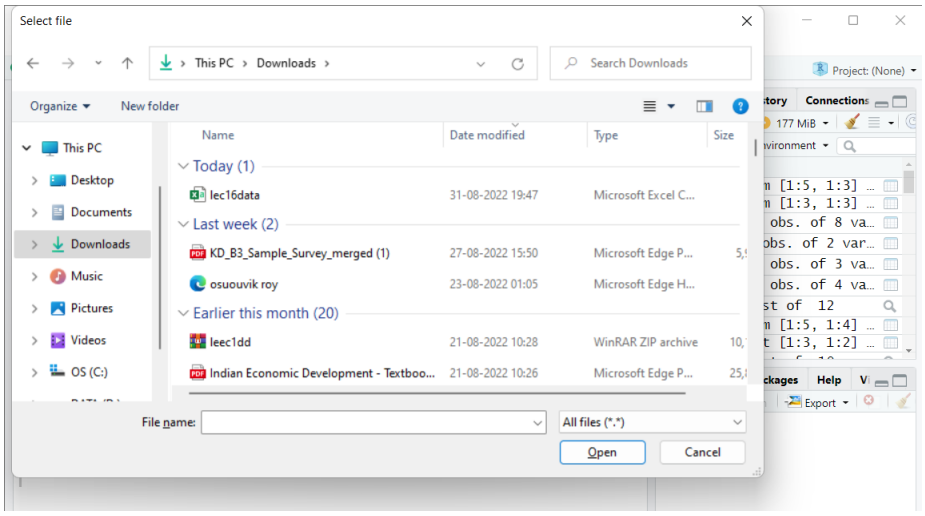
1 Introduction

Previously we have learn the theory of two factor model we have got a brief knowledge about interaction chart,additive model,non parallel interaction chart etc.Let's do some work in R.But before that we have prepare the data in R for ease of analysis.

2 Data read:

Our data should be store in computer in csv(Comma-Seperated Values).Then we can type command abc=read.csv("...") where our file name will be in place of ... another thing we can write our file location in the quotation but in that case precaution is that while in windows(commonly) coping path there will be back slash (\) but in command we have to write forward slash(/) another command is abc=read.csv(file.choose()) then a another window will open and file should be choosen by cursor double click.By default it will take header as true in other words coloums name of file will be taken in R otherwise not(by false command).After this command the file will be load as name abc.

```
> data=read.csv(file.choose())
```



a window like this will open and from this we can take our respected file. here we take lec16data.

3 Exploring Data:

we can print the data in R console by entering the name of the data as read.

dim("abc") by this cmd. we can know the no. of rows and coloumns in data.

```
> dim(data)
[1] 30  3
```

in our data there are 30 (data) row and 3 coloumn.

names("abc") by this we can know the coloumn names of the data.

```
> names(data)
[1] "variety" "tilt"    "yield"
```

this means our data set contains 3 coloumn with these names.

class("defg") by this we can know the nature of defg.

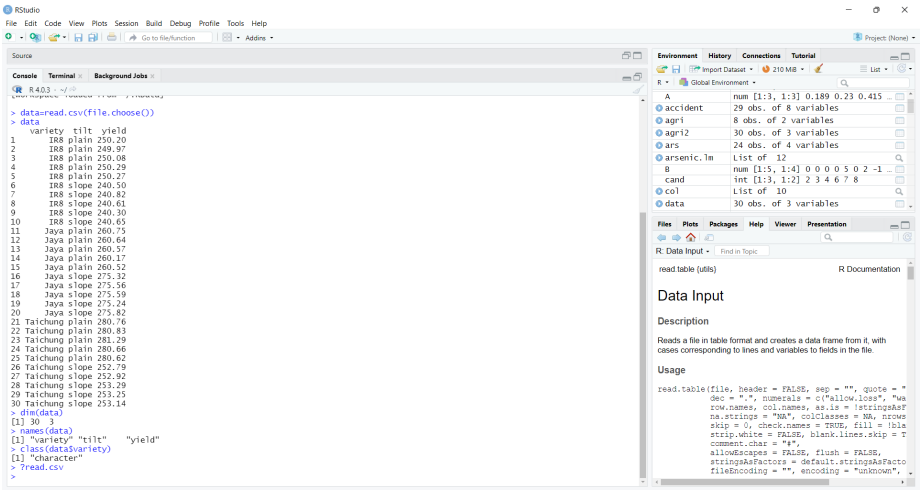
```
> class(data$variety)
[1] "character"
```

in our data this coloumn is taken as character. data$variety means variety coloumn of data variable.

as.is by this command we will change the nature of strings as factors if it taken false otherwise in original data form.

?cmd by this we can know the cmd and how to use this cmd etc.

```
> ?read.csv
```



in the right corner we can see about read.csv command.

4 as.is :

by default in read.csv command it is taken at false in r in r studio generally it taken as true so it is best to check it by class cmd. or by default.stringsAsFactors() .

By default.stringsAsFactors() this cmd. we can know whether in original data the coloumn was string and we taken any change or not .If change is taken then it is true or it is false.

```
> default.stringsAsFactors()
[1] FALSE
```

this means our strings are not taken as factors so for analysis we have to changed.

for change we have to use factor command.

```
> data$variety=factor(data$variety)
> class(data$variety)
[1] "factor"
```

so by this we can change form of string. But by this we have to individually change all the string so by as.is command we can take all strings as factor at one time.

```
> data=read.csv(file.choose(),as.is=FALSE)
> class(data$variety)
[1] "factor"
> class(data$tilt)
[1] "factor"
```

so we can see by this command we can take 2 strings as factors at a time.