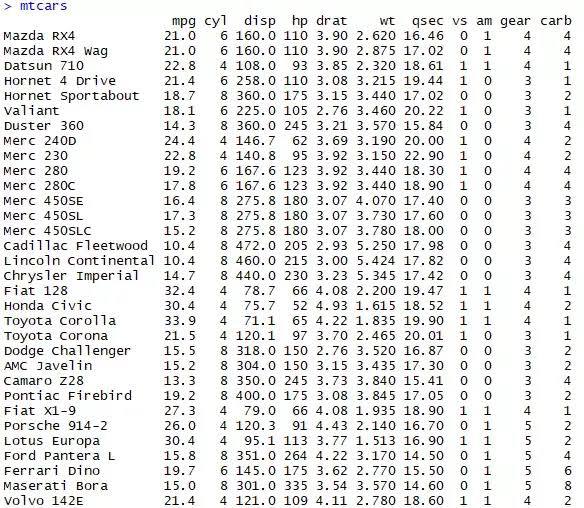
Task 2

Data manipulation in data.table for given data sets

By mauzum shamil a m

20bsc132

1} mtcars



## Last row content

*mtcars <-* ***as.data.table****(mtcars)*

*mtcars[.N]*

*Output;*

*## mpg cyl disp hp drat wt qsec vs am gear carb*

*## 1: 21.4 4 121 109 4.11 2.78 18.6 1 1 4 2*

# Subset rows where mpg is greater than 15

*mtcars[mpg* ***>*** *15* ***&*** *vs* ***!=*** *0]*

*Output;*

*## mpg cyl disp hp drat wt qsec vs am gear carb*

*## 1: 22.8 4 108.0 93 3.85 2.320 18.61 1 1 4 1*

*## 2: 21.4 6 258.0 110 3.08 3.215 19.44 1 0 3 1*

*## 3: 18.1 6 225.0 105 2.76 3.460 20.22 1 0 3 1*

*## 4: 24.4 4 146.7 62 3.69 3.190 20.00 1 0 4 2*

*## 5: 22.8 4 140.8 95 3.92 3.150 22.90 1 0 4 2*

*## 6: 19.2 6 167.6 123 3.92 3.440 18.30 1 0 4 4*

*## 7: 17.8 6 167.6 123 3.92 3.440 18.90 1 0 4 4*

*## 8: 32.4 4 78.7 66 4.08 2.200 19.47 1 1 4 1*

*## 9: 30.4 4 75.7 52 4.93 1.615 18.52 1 1 4 2*

*## 10: 33.9 4 71.1 65 4.22 1.835 19.90 1 1 4 1*

*## 11: 21.5 4 120.1 97 3.70 2.465 20.01 1 0 3 1*

*## 12: 27.3 4 79.0 66 4.08 1.935 18.90 1 1 4 1*

*## 13: 30.4 4 95.1 113 3.77 1.513 16.90 1 1 5 2*

*## 14: 21.4 4 121.0 109 4.11 2.780 18.60 1 1 4 2*

# If mtcars was only a data frame

>mtcars[mtcars$mpg > 15 & mtcars$vs != 0, ]

Output;

## mpg cyl disp hp drat wt qsec vs am gear carb

## 1: 22.8 4 108.0 93 3.85 2.320 18.61 1 1 4 1

## 2: 21.4 6 258.0 110 3.08 3.215 19.44 1 0 3 1

## 3: 18.1 6 225.0 105 2.76 3.460 20.22 1 0 3 1

## 4: 24.4 4 146.7 62 3.69 3.190 20.00 1 0 4 2

## 5: 22.8 4 140.8 95 3.92 3.150 22.90 1 0 4 2

## 6: 19.2 6 167.6 123 3.92 3.440 18.30 1 0 4 4

## 7: 17.8 6 167.6 123 3.92 3.440 18.90 1 0 4 4

## 8: 32.4 4 78.7 66 4.08 2.200 19.47 1 1 4 1

## 9: 30.4 4 75.7 52 4.93 1.615 18.52 1 1 4 2

## 10: 33.9 4 71.1 65 4.22 1.835 19.90 1 1 4 1

## 11: 21.5 4 120.1 97 3.70 2.465 20.01 1 0 3 1

## 12: 27.3 4 79.0 66 4.08 1.935 18.90 1 1 4 1

## 13: 30.4 4 95.1 113 3.77 1.513 16.90 1 1 5 2

## 14: 21.4 4 121.0 109 4.11 2.780 18.60 1 1 4 2

## selecting all but first two rows (! and - are equivalent to remove rows)

>mtcars[!1:2]

*Output*

*## mpg cyl disp hp drat wt qsec vs am gear carb*

*## 1: 22.8 4 108.0 93 3.85 2.320 18.61 1 1 4 1*

*## 2: 21.4 6 258.0 110 3.08 3.215 19.44 1 0 3 1*

*## 3: 18.7 8 360.0 175 3.15 3.440 17.02 0 0 3 2*

*## 4: 18.1 6 225.0 105 2.76 3.460 20.22 1 0 3 1*

*## 5: 14.3 8 360.0 245 3.21 3.570 15.84 0 0 3 4*

*## 6: 24.4 4 146.7 62 3.69 3.190 20.00 1 0 4 2*

*## 7: 22.8 4 140.8 95 3.92 3.150 22.90 1 0 4 2*

*## 8: 19.2 6 167.6 123 3.92 3.440 18.30 1 0 4 4*

*## 9: 17.8 6 167.6 123 3.92 3.440 18.90 1 0 4 4*

*## 10: 16.4 8 275.8 180 3.07 4.070 17.40 0 0 3 3*

*## 11: 17.3 8 275.8 180 3.07 3.730 17.60 0 0 3 3*

*## 12: 15.2 8 275.8 180 3.07 3.780 18.00 0 0 3 3*

*## 13: 10.4 8 472.0 205 2.93 5.250 17.98 0 0 3 4*

*## 14: 10.4 8 460.0 215 3.00 5.424 17.82 0 0 3 4*

*## 15: 14.7 8 440.0 230 3.23 5.345 17.42 0 0 3 4*

*## 16: 32.4 4 78.7 66 4.08 2.200 19.47 1 1 4 1*

*## 17: 30.4 4 75.7 52 4.93 1.615 18.52 1 1 4 2*

*## 18: 33.9 4 71.1 65 4.22 1.835 19.90 1 1 4 1*

*## 19: 21.5 4 120.1 97 3.70 2.465 20.01 1 0 3 1*

*## 20: 15.5 8 318.0 150 2.76 3.520 16.87 0 0 3 2*

*## 21: 15.2 8 304.0 150 3.15 3.435 17.30 0 0 3 2*

*## 22: 13.3 8 350.0 245 3.73 3.840 15.41 0 0 3 4*

*## 23: 19.2 8 400.0 175 3.08 3.845 17.05 0 0 3 2*

*## 24: 27.3 4 79.0 66 4.08 1.935 18.90 1 1 4 1*

*## 25: 26.0 4 120.3 91 4.43 2.140 16.70 0 1 5 2*

*## 26: 30.4 4 95.1 113 3.77 1.513 16.90 1 1 5 2*

*## 27: 15.8 8 351.0 264 4.22 3.170 14.50 0 1 5 4*

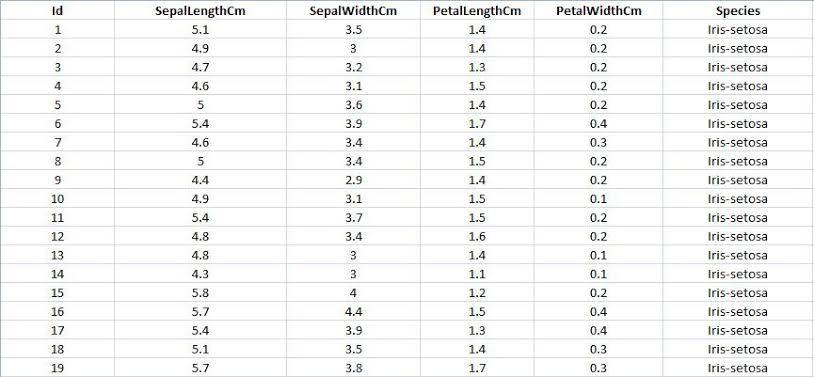
*## 28: 19.7 6 145.0 175 3.62 2.770 15.50 0 1 5 6*

*## 29: 15.0 8 301.0 335 3.54 3.570 14.60 0 1 5 8*

*## 30: 21.4 4 121.0 109 4.11 2.780 18.60 1 1 4 2*

*## mpg cyl disp hp drat wt qsec vs am gear carb*

*2} iris*

**

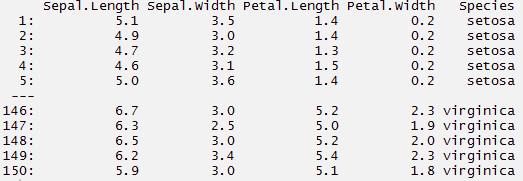
## to load

*> data(iris)*

*> myiris <- iris*

*> myiris <- data.table(myiris)*

*> myiris*

**

*#select columns with particular values*

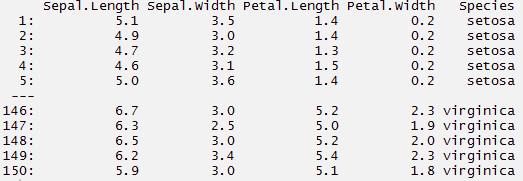
*> myiris[Species == 'setosa']*

**

*#select columns with multiple values. This will give you columns with Setosa*

*#and virginica species*

*> myiris[Species %in% c('setosa', 'virginica')]*

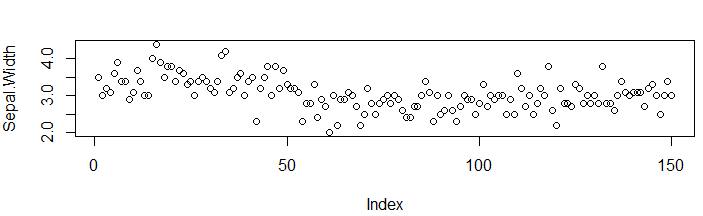
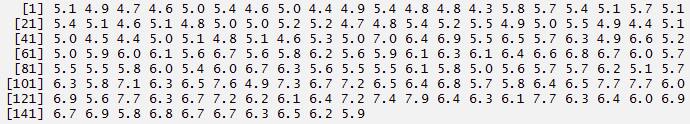
**

*#print and plot*

*> myiris[,{print(Sepal.Length)*

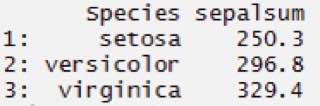
*> plot(Sepal.Width)*

*NULL}]*

**

*##grouping by a variable*

*> myiris[,.(sepalsum = sum(Sepal.Length)), by=Species]*

**

*#select a column for computation, hence need to set the key on column*

*> setkey(myiris, Species)*

*#selects all the rows associated with this data point*

*> myiris['setosa']*

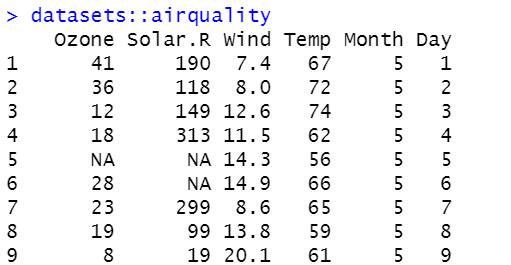
*> myiris[c('setosa', 'virginica')]*

*3} airquality*

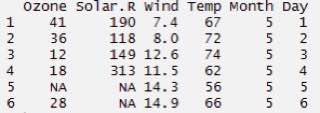
*#load data*

*> data("airquality")*

*> mydata <- airquality*

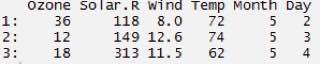
**

*> head(airquality,6)*

**

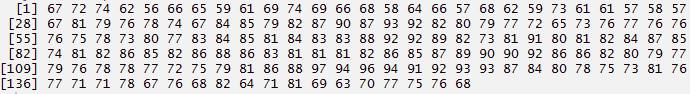
*#subset rows - select 2nd to 4th row*

*> mydata[2:4,]*

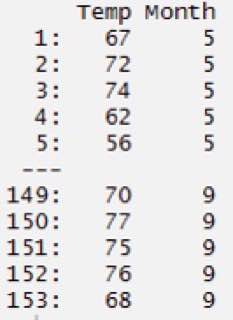
**

*#select columns. Returns a vector*

*> mydata[,Temp]*

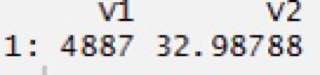
**

*> mydata[,.(Temp,Month)]*

**

*#returns sum and standard deviation*

*> mydata[,.(sum(Ozone, na.rm = TRUE), sd(Ozone, na.rm = TRUE))]*

**