

Data Manipulations

sumendar

August 25, 2016

Using dplyr Package

1. filter - It filters the data based on a condition
2. select - It is used to select columns of interest from a data set
3. arrange - It is used to arrange data set values on ascending or descending order
4. mutate - It is used to create new variables from existing variables
5. summarise (with group_by) - It is used to perform analysis by commonly used operations such as min, max, mean count etc

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
data("mtcars")
data('iris')
# take the tem object mydata and store mtcars data
mydata <- mtcars
dim(mydata)
```

```
## [1] 32 11
```

```
head(mydata,n=8)
```

```
##           mpg  cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4      21.0   6 160.0 110 3.90 2.620 16.46 0  1    4    4
## Mazda RX4 Wag  21.0   6 160.0 110 3.90 2.875 17.02 0  1    4    4
## Datsun 710      22.8   4 108.0  93 3.85 2.320 18.61 1  1    4    1
## Hornet 4 Drive  21.4   6 258.0 110 3.08 3.215 19.44 1  0    3    1
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02 0  0    3    2
## Valiant         18.1   6 225.0 105 2.76 3.460 20.22 1  0    3    1
## Duster 360      14.3   8 360.0 245 3.21 3.570 15.84 0  0    3    4
## Merc 240D       24.4   4 146.7  62 3.69 3.190 20.00 1  0    4    2
```

```
mydata
```

```
##           mpg  cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4      21.0   6 160.0 110 3.90 2.620 16.46 0  1    4    4
## Mazda RX4 Wag  21.0   6 160.0 110 3.90 2.875 17.02 0  1    4    4
## Datsun 710      22.8   4 108.0  93 3.85 2.320 18.61 1  1    4    1
## Hornet 4 Drive  21.4   6 258.0 110 3.08 3.215 19.44 1  0    3    1
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02 0  0    3    2
## Valiant         18.1   6 225.0 105 2.76 3.460 20.22 1  0    3    1
```

```
## Duster 360      14.3   8 360.0 245 3.21 3.570 15.84 0 0   3   4
## Merc 240D      24.4   4 146.7  62 3.69 3.190 20.00 1 0   4   2
## Merc 230       22.8   4 140.8  95 3.92 3.150 22.90 1 0   4   2
## Merc 280       19.2   6 167.6 123 3.92 3.440 18.30 1 0   4   4
## Merc 280C      17.8   6 167.6 123 3.92 3.440 18.90 1 0   4   4
## Merc 450SE     16.4   8 275.8 180 3.07 4.070 17.40 0 0   3   3
## Merc 450SL     17.3   8 275.8 180 3.07 3.730 17.60 0 0   3   3
## Merc 450SLC    15.2   8 275.8 180 3.07 3.780 18.00 0 0   3   3
## Cadillac Fleetwood 10.4  8 472.0 205 2.93 5.250 17.98 0 0   3   4
## Lincoln Continental 10.4  8 460.0 215 3.00 5.424 17.82 0 0   3   4
## Chrysler Imperial 14.7  8 440.0 230 3.23 5.345 17.42 0 0   3   4
## Fiat 128       32.4   4  78.7  66 4.08 2.200 19.47 1 1   4   1
## Honda Civic    30.4   4  75.7  52 4.93 1.615 18.52 1 1   4   2
## Toyota Corolla 33.9   4  71.1  65 4.22 1.835 19.90 1 1   4   1
## Toyota Corona  21.5   4 120.1  97 3.70 2.465 20.01 1 0   3   1
## Dodge Challenger 15.5  8 318.0 150 2.76 3.520 16.87 0 0   3   2
## AMC Javelin    15.2   8 304.0 150 3.15 3.435 17.30 0 0   3   2
## Camaro Z28     13.3   8 350.0 245 3.73 3.840 15.41 0 0   3   4
## Pontiac Firebird 19.2  8 400.0 175 3.08 3.845 17.05 0 0   3   2
## Fiat X1-9      27.3   4  79.0  66 4.08 1.935 18.90 1 1   4   1
## Porsche 914-2  26.0   4 120.3  91 4.43 2.140 16.70 0 1   5   2
## Lotus Europa   30.4   4  95.1 113 3.77 1.513 16.90 1 1   5   2
## Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.50 0 1   5   4
## Ferrari Dino   19.7   6 145.0 175 3.62 2.770 15.50 0 1   5   6
## Maserati Bora   15.0   8 301.0 335 3.54 3.570 14.60 0 1   5   8
## Volvo 142E     21.4   4 121.0 109 4.11 2.780 18.60 1 1   4   2
```

```
#creating a local dataframe. Local data frame are easier to read
mynewdata <- tbl_df(mydata)
#now data will be in tabular structure
mynewdata
```

```
## Source: local data frame [32 x 11]
##
##      mpg    cyl  disp    hp  drat    wt  qsec    vs    am  gear  carb
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1   21.0     6 160.0   110  3.90  2.620 16.46     0     1     4     4
## 2   21.0     6 160.0   110  3.90  2.875 17.02     0     1     4     4
## 3   22.8     4 108.0    93  3.85  2.320 18.61     1     1     4     1
## 4   21.4     6 258.0   110  3.08  3.215 19.44     1     0     3     1
## 5   18.7     8 360.0   175  3.15  3.440 17.02     0     0     3     2
## 6   18.1     6 225.0   105  2.76  3.460 20.22     1     0     3     1
## 7   14.3     8 360.0   245  3.21  3.570 15.84     0     0     3     4
## 8   24.4     4 146.7    62  3.69  3.190 20.00     1     0     4     2
## 9   22.8     4 140.8    95  3.92  3.150 22.90     1     0     4     2
## 10  19.2     6 167.6   123  3.92  3.440 18.30     1     0     4     4
## ..     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...
```

```
myirisdata <- tbl_df(iris)
myirisdata
```

```
## Source: local data frame [150 x 5]
##
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
##            <dbl>         <dbl>         <dbl>         <dbl>   <fctr>
```

```
## 1      5.1      3.5      1.4      0.2 setosa
## 2      4.9      3.0      1.4      0.2 setosa
## 3      4.7      3.2      1.3      0.2 setosa
## 4      4.6      3.1      1.5      0.2 setosa
## 5      5.0      3.6      1.4      0.2 setosa
## 6      5.4      3.9      1.7      0.4 setosa
## 7      4.6      3.4      1.4      0.3 setosa
## 8      5.0      3.4      1.5      0.2 setosa
## 9      4.4      2.9      1.4      0.2 setosa
## 10     4.9      3.1      1.5      0.1 setosa
## ..      ...      ...      ...      ...      ...
```

```
#use filter to filter data with required condition
filter(mynewdata, cyl > 4 & gear > 4)
```

```
## Source: local data frame [3 x 11]
##
```

```
##      mpg   cyl  disp    hp  drat    wt  qsec    vs  am  gear  carb
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  15.8   8    351    264  4.22  3.17  14.5     0   1     5     4
## 2  19.7   6    145    175  3.62  2.77  15.5     0   1     5     6
## 3  15.0   8    301    335  3.54  3.57  14.6     0   1     5     8
```

```
filter(mydata, cyl > 4 & gear > 4)
```

```
##      mpg cyl disp  hp drat   wt  qsec vs am gear carb
## 1 15.8   8  351  264 4.22 3.17 14.5  0  1   5   4
## 2 19.7   6  145  175 3.62 2.77 15.5  0  1   5   6
## 3 15.0   8  301  335 3.54 3.57 14.6  0  1   5   8
```

```
filter(mynewdata, cyl > 4)
```

```
## Source: local data frame [21 x 11]
##
```

```
##      mpg   cyl  disp    hp  drat    wt  qsec    vs  am  gear  carb
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  21.0   6  160.0   110  3.90  2.620 16.46     0   1     4     4
## 2  21.0   6  160.0   110  3.90  2.875 17.02     0   1     4     4
## 3  21.4   6  258.0   110  3.08  3.215 19.44     1   0     3     1
## 4  18.7   8  360.0   175  3.15  3.440 17.02     0   0     3     2
## 5  18.1   6  225.0   105  2.76  3.460 20.22     1   0     3     1
## 6  14.3   8  360.0   245  3.21  3.570 15.84     0   0     3     4
## 7  19.2   6  167.6   123  3.92  3.440 18.30     1   0     4     4
## 8  17.8   6  167.6   123  3.92  3.440 18.90     1   0     4     4
## 9  16.4   8  275.8   180  3.07  4.070 17.40     0   0     3     3
## 10 17.3   8  275.8   180  3.07  3.730 17.60     0   0     3     3
## ..   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

```
filter(myirisdata, Species %in% c('setosa', 'virginica'))
```

```
## Source: local data frame [100 x 5]
##
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
##           <dbl>         <dbl>         <dbl>         <dbl>   <fctr>
## 1           5.1           3.5           1.4           0.2   setosa
## 2           4.9           3.0           1.4           0.2   setosa
## 3           4.7           3.2           1.3           0.2   setosa
```

```
## 4          4.6          3.1          1.5          0.2 setosa
## 5          5.0          3.6          1.4          0.2 setosa
## 6          5.4          3.9          1.7          0.4 setosa
## 7          4.6          3.4          1.4          0.3 setosa
## 8          5.0          3.4          1.5          0.2 setosa
## 9          4.4          2.9          1.4          0.2 setosa
## 10         4.9          3.1          1.5          0.1 setosa
## ..          ...          ...          ...          ...      ...
```

```
library(dplyr)
#use select to pick columns by name
select(mynewdata,cyl,mpg,hp)
```

```
## Source: local data frame [32 x 3]
##
##      cyl  mpg   hp
##    <dbl> <dbl> <dbl>
## 1      6 21.0  110
## 2      6 21.0  110
## 3      4 22.8   93
## 4      6 21.4  110
## 5      8 18.7  175
## 6      6 18.1  105
## 7      8 14.3  245
## 8      4 24.4   62
## 9      4 22.8   95
## 10     6 19.2  123
## ..    ...    ...    ...
```

```
# using base function selecting the columns
subset(x = mydata,mydata$cyl>4)
```

```
##      mpg cyl  disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4      21.0   6 160.0 110 3.90 2.620 16.46 0 1   4    4
## Mazda RX4 Wag  21.0   6 160.0 110 3.90 2.875 17.02 0 1   4    4
## Hornet 4 Drive  21.4   6 258.0 110 3.08 3.215 19.44 1 0   3    1
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02 0 0   3    2
## Valiant        18.1   6 225.0 105 2.76 3.460 20.22 1 0   3    1
## Duster 360     14.3   8 360.0 245 3.21 3.570 15.84 0 0   3    4
## Merc 280       19.2   6 167.6 123 3.92 3.440 18.30 1 0   4    4
## Merc 280C      17.8   6 167.6 123 3.92 3.440 18.90 1 0   4    4
## Merc 450SE     16.4   8 275.8 180 3.07 4.070 17.40 0 0   3    3
## Merc 450SL     17.3   8 275.8 180 3.07 3.730 17.60 0 0   3    3
## Merc 450SLC    15.2   8 275.8 180 3.07 3.780 18.00 0 0   3    3
## Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98 0 0   3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82 0 0   3    4
## Chrysler Imperial 14.7   8 440.0 230 3.23 5.345 17.42 0 0   3    4
## Dodge Challenger 15.5   8 318.0 150 2.76 3.520 16.87 0 0   3    2
## AMC Javelin     15.2   8 304.0 150 3.15 3.435 17.30 0 0   3    2
## Camaro Z28      13.3   8 350.0 245 3.73 3.840 15.41 0 0   3    4
## Pontiac Firebird 19.2   8 400.0 175 3.08 3.845 17.05 0 0   3    2
## Ford Pantera L  15.8   8 351.0 264 4.22 3.170 14.50 0 1   5    4
## Ferrari Dino    19.7   6 145.0 175 3.62 2.770 15.50 0 1   5    6
## Maserati Bora   15.0   8 301.0 335 3.54 3.570 14.60 0 1   5    8
```

```
mynewdata[,c(2,1,4)]
```

```
## Source: local data frame [32 x 3]
```

```
##  
##      cyl    mpg    hp  
##    <dbl> <dbl> <dbl>  
## 1      6  21.0  110  
## 2      6  21.0  110  
## 3      4  22.8   93  
## 4      6  21.4  110  
## 5      8  18.7  175  
## 6      6  18.1  105  
## 7      8  14.3  245  
## 8      4  24.4   62  
## 9      4  22.8   95  
## 10     6  19.2  123  
## ..    ...    ...    ...
```

```
#here you can use (-) to hide columns
```

```
select(mynewdata, -cyl, -mpg )
```

```
## Source: local data frame [32 x 9]
```

```
##  
##      disp    hp  drat    wt  qsec    vs    am  gear  carb  
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1  160.0   110  3.90 2.620 16.46     0     1     4     4  
## 2  160.0   110  3.90 2.875 17.02     0     1     4     4  
## 3  108.0    93  3.85 2.320 18.61     1     1     4     1  
## 4  258.0   110  3.08 3.215 19.44     1     0     3     1  
## 5  360.0   175  3.15 3.440 17.02     0     0     3     2  
## 6  225.0   105  2.76 3.460 20.22     1     0     3     1  
## 7  360.0   245  3.21 3.570 15.84     0     0     3     4  
## 8  146.7    62  3.69 3.190 20.00     1     0     4     2  
## 9  140.8    95  3.92 3.150 22.90     1     0     4     2  
## 10 167.6   123  3.92 3.440 18.30     1     0     4     4  
## ..    ...    ...    ...    ...    ...    ...    ...    ...
```

```
#hide a range of columns
```

```
select(mynewdata, -c(cyl,mpg,vs,gear))
```

```
## Source: local data frame [32 x 7]
```

```
##  
##      disp    hp  drat    wt  qsec    am  carb  
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1  160.0   110  3.90 2.620 16.46     1     4  
## 2  160.0   110  3.90 2.875 17.02     1     4  
## 3  108.0    93  3.85 2.320 18.61     1     1  
## 4  258.0   110  3.08 3.215 19.44     0     1  
## 5  360.0   175  3.15 3.440 17.02     0     2  
## 6  225.0   105  2.76 3.460 20.22     0     1  
## 7  360.0   245  3.21 3.570 15.84     0     4  
## 8  146.7    62  3.69 3.190 20.00     0     2  
## 9  140.8    95  3.92 3.150 22.90     0     2  
## 10 167.6   123  3.92 3.440 18.30     0     4  
## ..    ...    ...    ...    ...    ...    ...
```

```
#select series of columns
select(mynewdata, cyl:gear) #mpg and last variable is hided
```

```
## Source: local data frame [32 x 9]
##
##      cyl  disp    hp  drat    wt   qsec    vs    am  gear
##      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1      6 160.0   110  3.90  2.620 16.46    0     1     4
## 2      6 160.0   110  3.90  2.875 17.02    0     1     4
## 3      4 108.0    93  3.85  2.320 18.61    1     1     4
## 4      6 258.0   110  3.08  3.215 19.44    1     0     3
## 5      8 360.0   175  3.15  3.440 17.02    0     0     3
## 6      6 225.0   105  2.76  3.460 20.22    1     0     3
## 7      8 360.0   245  3.21  3.570 15.84    0     0     3
## 8      4 146.7    62  3.69  3.190 20.00    1     0     4
## 9      4 140.8    95  3.92  3.150 22.90    1     0     4
## 10     6 167.6   123  3.92  3.440 18.30    1     0     4
## ..    ...    ...    ...    ...    ...    ...    ...    ...
```

```
#chaining or pipelining - a way to perform multiple operations in one line
```

```
mynewdata %>%select(cyl, wt, gear)%>% filter(wt > 2)
```

```
## Source: local data frame [28 x 3]
##
##      cyl    wt  gear
##      <dbl> <dbl> <dbl>
## 1      6  2.620    4
## 2      6  2.875    4
## 3      4  2.320    4
## 4      6  3.215    3
## 5      8  3.440    3
## 6      6  3.460    3
## 7      8  3.570    3
## 8      4  3.190    4
## 9      4  3.150    4
## 10     6  3.440    4
## ..    ...    ...    ...
```

```
#arrange can be used to reorder rows
mynewdata%>%select(cyl, wt, gear)%>%arrange(wt)
```

```
## Source: local data frame [32 x 3]
##
##      cyl    wt  gear
##      <dbl> <dbl> <dbl>
## 1      4  1.513    5
## 2      4  1.615    4
## 3      4  1.835    4
## 4      4  1.935    4
## 5      4  2.140    5
## 6      4  2.200    4
## 7      4  2.320    4
## 8      4  2.465    3
```

```
## 9      6 2.620      4
## 10     6 2.770      5
## ..    ...    ...    ...
```

```
min(mynewdata$wt)
```

```
## [1] 1.513
```

```
#descending order
```

```
mynewdata%>%select(cyl, wt, gear)%>%arrange(desc(wt))
```

```
## Source: local data frame [32 x 3]
```

```
##
##      cyl      wt  gear
##    <dbl> <dbl> <dbl>
## 1      8 5.424     3
## 2      8 5.345     3
## 3      8 5.250     3
## 4      8 4.070     3
## 5      8 3.845     3
## 6      8 3.840     3
## 7      8 3.780     3
## 8      8 3.730     3
## 9      8 3.570     3
## 10     8 3.570     5
## ..    ...    ...    ...
```

```
#mutate - create new variables
```

```
(mynewdata %>%select(mpg, cyl)%>%mutate(newvariable = mpg*cyl))
```

```
## Source: local data frame [32 x 3]
```

```
##
##      mpg      cyl newvariable
##    <dbl> <dbl>      <dbl>
## 1    21.0      6      126.0
## 2    21.0      6      126.0
## 3    22.8      4       91.2
## 4    21.4      6      128.4
## 5    18.7      8      149.6
## 6    18.1      6      108.6
## 7    14.3      8      114.4
## 8    24.4      4       97.6
## 9    22.8      4       91.2
## 10   19.2      6      115.2
## ..    ...    ...    ...
```

```
mynewdata
```

```
## Source: local data frame [32 x 11]
```

```
##
##      mpg      cyl  disp    hp  drat    wt  qsec    vs  am  gear  carb
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1    21.0      6  160.0   110  3.90  2.620  16.46    0    1    4     4
## 2    21.0      6  160.0   110  3.90  2.875  17.02    0    1    4     4
## 3    22.8      4  108.0    93  3.85  2.320  18.61    1    1    4     1
## 4    21.4      6  258.0   110  3.08  3.215  19.44    1    0    3     1
## 5    18.7      8  360.0   175  3.15  3.440  17.02    0    0    3     2
```

```
## 6  18.1    6 225.0   105  2.76 3.460 20.22    1    0    3    1
## 7  14.3    8 360.0   245  3.21 3.570 15.84    0    0    3    4
## 8  24.4    4 146.7    62  3.69 3.190 20.00    1    0    4    2
## 9  22.8    4 140.8    95  3.92 3.150 22.90    1    0    4    2
## 10 19.2    6 167.6   123  3.92 3.440 18.30    1    0    4    4
## ..    ...    ...    ...    ...    ...    ...    ...    ...    ...
```

```
newvariable <- mynewdata %>%mutate(newvariable = mpg*cyl)
```

#summarise - this is used to find insights from data

```
myirisdata%>%group_by(Species)%>%summarise(Average = mean(Sepal.Length, na.rm = TRUE))
```

```
## Source: local data frame [3 x 2]
```

```
##
```

```
##      Species Average
```

```
##      <fctr>   <dbl>
```

```
## 1    setosa   5.006
```

```
## 2 versicolor 5.936
```

```
## 3 virginica  6.588
```

```
myirisdata1<-myirisdata%>%group_by(Species)%>%summarise(Average = mean(Sepal.Length, na.rm = TRUE))
```

#mean with NA

```
vec <- c(2,4,5,7,NA)
```

```
vec
```

```
## [1]  2  4  5  7 NA
```

```
mean(vec,na.rm = TRUE)
```

```
## [1] 4.5
```

```
####
```

```
View(myirisdata1)
```

#or use summarise each

```
myirisdata%>%
```

```
  group_by(Species)%>%
```

```
  summarise_each(funs(mean, n()), Sepal.Length, Sepal.Width)
```

```
## Source: local data frame [3 x 5]
```

```
##
```

```
##      Species Sepal.Length_mean Sepal.Width_mean Sepal.Length_n
```

```
##      <fctr>           <dbl>           <dbl>           <int>
```

```
## 1    setosa           5.006           3.428           50
```

```
## 2 versicolor         5.936           2.770           50
```

```
## 3 virginica          6.588           2.974           50
```

```
## Variables not shown: Sepal.Width_n <int>.
```

#you can rename the variables using rename command

```
mynewdata %>% rename(miles = mpg)
```

```
## Source: local data frame [32 x 11]
```

```
##
```

```
##      miles   cyl disp   hp drat   wt  qsec    vs    am  gear  carb
```

```
##      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
```

```
## 1    21.0    6 160.0   110  3.90 2.620 16.46    0    1    4    4
```

```
## 2    21.0    6 160.0   110  3.90 2.875 17.02    0    1    4    4
```

```
## 3    22.8    4 108.0    93  3.85 2.320 18.61    1    1    4    1
```



```
## 4  21.4    6 258.0   110 3.08 3.215 19.44    1    0    3    1
## 5  18.7    8 360.0   175 3.15 3.440 17.02    0    0    3    2
## 6  18.1    6 225.0   105 2.76 3.460 20.22    1    0    3    1
## 7  14.3    8 360.0   245 3.21 3.570 15.84    0    0    3    4
## 8  24.4    4 146.7    62 3.69 3.190 20.00    1    0    4    2
## 9  22.8    4 140.8    95 3.92 3.150 22.90    1    0    4    2
## 10 19.2    6 167.6   123 3.92 3.440 18.30    1    0    4    4
## ..    ...    ...    ...    ...    ...    ...    ...    ...    ...
```

rename the column without dplyr

```
#method 1
names(mynewdata)[1] <- "miles"
#method 2
names(mynewdata) <- sub("^miles$", "mpg", names(mynewdata))
#method 3
names(mynewdata)[names(mynewdata) == 'miles'] <- 'mpg'
mynewdata
```

```
## Source: local data frame [32 x 11]
##
##      mpg    cyl  disp    hp  drat    wt   qsec    vs    am  gear  carb
##    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  21.0     6 160.0   110  3.90  2.620 16.46     0     1     4     4
## 2  21.0     6 160.0   110  3.90  2.875 17.02     0     1     4     4
## 3  22.8     4 108.0    93  3.85  2.320 18.61     1     1     4     1
## 4  21.4     6 258.0   110  3.08  3.215 19.44     1     0     3     1
## 5  18.7     8 360.0   175  3.15  3.440 17.02     0     0     3     2
## 6  18.1     6 225.0   105  2.76  3.460 20.22     1     0     3     1
## 7  14.3     8 360.0   245  3.21  3.570 15.84     0     0     3     4
## 8  24.4     4 146.7    62  3.69  3.190 20.00     1     0     4     2
## 9  22.8     4 140.8    95  3.92  3.150 22.90     1     0     4     2
## 10 19.2     6 167.6   123  3.92  3.440 18.30     1     0     4     4
## ..    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
```

data.table package

```
library(data.table)
#load data sets
data("airquality")
mydata <- airquality
head(mydata)
```

```
##      Ozone Solar.R Wind Temp Month Day
## 1     41     190   7.4   67     5    1
## 2     36     118   8.0   72     5    2
## 3     12     149  12.6   74     5    3
## 4     18     313  11.5   62     5    4
## 5     NA      NA  14.3   56     5    5
## 6     28      NA  14.9   66     5    6
```

```
data(iris)
myiris <- iris
class(mydata)
```

```
## [1] "data.frame"
```

```
mydata <- data.table(mydata)
class(mydata)
```

```
## [1] "data.table" "data.frame"
```

```
myiris <- data.table(myiris)
myiris
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width  Species
##  1:           5.1           3.5           1.4           0.2    setosa
##  2:           4.9           3.0           1.4           0.2    setosa
##  3:           4.7           3.2           1.3           0.2    setosa
##  4:           4.6           3.1           1.5           0.2    setosa
##  5:           5.0           3.6           1.4           0.2    setosa
##  ---
## 146:           6.7           3.0           5.2           2.3 virginica
## 147:           6.3           2.5           5.0           1.9 virginica
## 148:           6.5           3.0           5.2           2.0 virginica
## 149:           6.2           3.4           5.4           2.3 virginica
## 150:           5.9           3.0           5.1           1.8 virginica
```

```
#subset rows - select 2nd to 4th row
mydata[2:4,]
```

```
##      Ozone Solar.R Wind Temp Month Day
## 1:    36    118  8.0   72     5    2
## 2:    12    149 12.6   74     5    3
## 3:    18    313 11.5   62     5    4
```

```
#select columns with particular values
myiris[Species == 'setosa']
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
##  1:           5.1           3.5           1.4           0.2    setosa
##  2:           4.9           3.0           1.4           0.2    setosa
##  3:           4.7           3.2           1.3           0.2    setosa
##  4:           4.6           3.1           1.5           0.2    setosa
##  5:           5.0           3.6           1.4           0.2    setosa
##  6:           5.4           3.9           1.7           0.4    setosa
##  7:           4.6           3.4           1.4           0.3    setosa
##  8:           5.0           3.4           1.5           0.2    setosa
##  9:           4.4           2.9           1.4           0.2    setosa
## 10:           4.9           3.1           1.5           0.1    setosa
## 11:           5.4           3.7           1.5           0.2    setosa
## 12:           4.8           3.4           1.6           0.2    setosa
## 13:           4.8           3.0           1.4           0.1    setosa
## 14:           4.3           3.0           1.1           0.1    setosa
## 15:           5.8           4.0           1.2           0.2    setosa
## 16:           5.7           4.4           1.5           0.4    setosa
## 17:           5.4           3.9           1.3           0.4    setosa
## 18:           5.1           3.5           1.4           0.3    setosa
## 19:           5.7           3.8           1.7           0.3    setosa
## 20:           5.1           3.8           1.5           0.3    setosa
```

```
## 21:      5.4      3.4      1.7      0.2 setosa
## 22:      5.1      3.7      1.5      0.4 setosa
## 23:      4.6      3.6      1.0      0.2 setosa
## 24:      5.1      3.3      1.7      0.5 setosa
## 25:      4.8      3.4      1.9      0.2 setosa
## 26:      5.0      3.0      1.6      0.2 setosa
## 27:      5.0      3.4      1.6      0.4 setosa
## 28:      5.2      3.5      1.5      0.2 setosa
## 29:      5.2      3.4      1.4      0.2 setosa
## 30:      4.7      3.2      1.6      0.2 setosa
## 31:      4.8      3.1      1.6      0.2 setosa
## 32:      5.4      3.4      1.5      0.4 setosa
## 33:      5.2      4.1      1.5      0.1 setosa
## 34:      5.5      4.2      1.4      0.2 setosa
## 35:      4.9      3.1      1.5      0.2 setosa
## 36:      5.0      3.2      1.2      0.2 setosa
## 37:      5.5      3.5      1.3      0.2 setosa
## 38:      4.9      3.6      1.4      0.1 setosa
## 39:      4.4      3.0      1.3      0.2 setosa
## 40:      5.1      3.4      1.5      0.2 setosa
## 41:      5.0      3.5      1.3      0.3 setosa
## 42:      4.5      2.3      1.3      0.3 setosa
## 43:      4.4      3.2      1.3      0.2 setosa
## 44:      5.0      3.5      1.6      0.6 setosa
## 45:      5.1      3.8      1.9      0.4 setosa
## 46:      4.8      3.0      1.4      0.3 setosa
## 47:      5.1      3.8      1.6      0.2 setosa
## 48:      4.6      3.2      1.4      0.2 setosa
## 49:      5.3      3.7      1.5      0.2 setosa
## 50:      5.0      3.3      1.4      0.2 setosa
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
```

```
#Using normal data frame we can also get setosa species;
class(iris)
```

```
## [1] "data.frame"
```

```
subset(iris,iris$Species == 'setosa')
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1      5.1      3.5      1.4      0.2 setosa
## 2      4.9      3.0      1.4      0.2 setosa
## 3      4.7      3.2      1.3      0.2 setosa
## 4      4.6      3.1      1.5      0.2 setosa
## 5      5.0      3.6      1.4      0.2 setosa
## 6      5.4      3.9      1.7      0.4 setosa
## 7      4.6      3.4      1.4      0.3 setosa
## 8      5.0      3.4      1.5      0.2 setosa
## 9      4.4      2.9      1.4      0.2 setosa
## 10     4.9      3.1      1.5      0.1 setosa
## 11     5.4      3.7      1.5      0.2 setosa
## 12     4.8      3.4      1.6      0.2 setosa
## 13     4.8      3.0      1.4      0.1 setosa
## 14     4.3      3.0      1.1      0.1 setosa
## 15     5.8      4.0      1.2      0.2 setosa
## 16     5.7      4.4      1.5      0.4 setosa
```

```
## 17      5.4      3.9      1.3      0.4 setosa
## 18      5.1      3.5      1.4      0.3 setosa
## 19      5.7      3.8      1.7      0.3 setosa
## 20      5.1      3.8      1.5      0.3 setosa
## 21      5.4      3.4      1.7      0.2 setosa
## 22      5.1      3.7      1.5      0.4 setosa
## 23      4.6      3.6      1.0      0.2 setosa
## 24      5.1      3.3      1.7      0.5 setosa
## 25      4.8      3.4      1.9      0.2 setosa
## 26      5.0      3.0      1.6      0.2 setosa
## 27      5.0      3.4      1.6      0.4 setosa
## 28      5.2      3.5      1.5      0.2 setosa
## 29      5.2      3.4      1.4      0.2 setosa
## 30      4.7      3.2      1.6      0.2 setosa
## 31      4.8      3.1      1.6      0.2 setosa
## 32      5.4      3.4      1.5      0.4 setosa
## 33      5.2      4.1      1.5      0.1 setosa
## 34      5.5      4.2      1.4      0.2 setosa
## 35      4.9      3.1      1.5      0.2 setosa
## 36      5.0      3.2      1.2      0.2 setosa
## 37      5.5      3.5      1.3      0.2 setosa
## 38      4.9      3.6      1.4      0.1 setosa
## 39      4.4      3.0      1.3      0.2 setosa
## 40      5.1      3.4      1.5      0.2 setosa
## 41      5.0      3.5      1.3      0.3 setosa
## 42      4.5      2.3      1.3      0.3 setosa
## 43      4.4      3.2      1.3      0.2 setosa
## 44      5.0      3.5      1.6      0.6 setosa
## 45      5.1      3.8      1.9      0.4 setosa
## 46      4.8      3.0      1.4      0.3 setosa
## 47      5.1      3.8      1.6      0.2 setosa
## 48      4.6      3.2      1.4      0.2 setosa
## 49      5.3      3.7      1.5      0.2 setosa
## 50      5.0      3.3      1.4      0.2 setosa
```

#select columns with multiple values. This will give you columns with Setosa and virginica species

```
temp1<-myiris[Species %in% c('setosa', 'virginica')]
View(temp1)
```

#select columns. Returns a vector

#now take airquality data table

```
mydata[,Temp]
```

```
## [1] 67 72 74 62 56 66 65 59 61 69 74 69 66 68 58 64 66 57 68 62 59 73 61
## [24] 61 57 58 57 67 81 79 76 78 74 67 84 85 79 82 87 90 87 93 92 82 80 79
## [47] 77 72 65 73 76 77 76 76 76 75 78 73 80 77 83 84 85 81 84 83 83 88 92
## [70] 92 89 82 73 81 91 80 81 82 84 87 85 74 81 82 86 85 82 86 88 86 83 81
## [93] 81 81 82 86 85 87 89 90 90 92 86 86 82 80 79 77 79 76 78 78 77 72 75
## [116] 79 81 86 88 97 94 96 94 91 92 93 93 87 84 80 78 75 73 81 76 77 71 71
## [139] 78 67 76 68 82 64 71 81 69 63 70 77 75 76 68
```

#retrieving two columns using data table

```
mydata[,.(Temp,Month)]
```

```
##      Temp Month
## 1:    67     5
```

```
## 2: 72 5
## 3: 74 5
## 4: 62 5
## 5: 56 5
## ---
## 149: 70 9
## 150: 77 9
## 151: 75 9
## 152: 76 9
## 153: 68 9
```

```
#same above operation on normal data frame
airquality[,c("Temp","Month")]
```

```
##      Temp Month
## 1      67     5
## 2      72     5
## 3      74     5
## 4      62     5
## 5      56     5
## 6      66     5
## 7      65     5
## 8      59     5
## 9      61     5
## 10     69     5
## 11     74     5
## 12     69     5
## 13     66     5
## 14     68     5
## 15     58     5
## 16     64     5
## 17     66     5
## 18     57     5
## 19     68     5
## 20     62     5
## 21     59     5
## 22     73     5
## 23     61     5
## 24     61     5
## 25     57     5
## 26     58     5
## 27     57     5
## 28     67     5
## 29     81     5
## 30     79     5
## 31     76     5
## 32     78     6
## 33     74     6
## 34     67     6
## 35     84     6
## 36     85     6
## 37     79     6
## 38     82     6
## 39     87     6
## 40     90     6
```

## 41	87	6
## 42	93	6
## 43	92	6
## 44	82	6
## 45	80	6
## 46	79	6
## 47	77	6
## 48	72	6
## 49	65	6
## 50	73	6
## 51	76	6
## 52	77	6
## 53	76	6
## 54	76	6
## 55	76	6
## 56	75	6
## 57	78	6
## 58	73	6
## 59	80	6
## 60	77	6
## 61	83	6
## 62	84	7
## 63	85	7
## 64	81	7
## 65	84	7
## 66	83	7
## 67	83	7
## 68	88	7
## 69	92	7
## 70	92	7
## 71	89	7
## 72	82	7
## 73	73	7
## 74	81	7
## 75	91	7
## 76	80	7
## 77	81	7
## 78	82	7
## 79	84	7
## 80	87	7
## 81	85	7
## 82	74	7
## 83	81	7
## 84	82	7
## 85	86	7
## 86	85	7
## 87	82	7
## 88	86	7
## 89	88	7
## 90	86	7
## 91	83	7
## 92	81	7
## 93	81	8
## 94	81	8

## 95	82	8
## 96	86	8
## 97	85	8
## 98	87	8
## 99	89	8
## 100	90	8
## 101	90	8
## 102	92	8
## 103	86	8
## 104	86	8
## 105	82	8
## 106	80	8
## 107	79	8
## 108	77	8
## 109	79	8
## 110	76	8
## 111	78	8
## 112	78	8
## 113	77	8
## 114	72	8
## 115	75	8
## 116	79	8
## 117	81	8
## 118	86	8
## 119	88	8
## 120	97	8
## 121	94	8
## 122	96	8
## 123	94	8
## 124	91	9
## 125	92	9
## 126	93	9
## 127	93	9
## 128	87	9
## 129	84	9
## 130	80	9
## 131	78	9
## 132	75	9
## 133	73	9
## 134	81	9
## 135	76	9
## 136	77	9
## 137	71	9
## 138	71	9
## 139	78	9
## 140	67	9
## 141	76	9
## 142	68	9
## 143	82	9
## 144	64	9
## 145	71	9
## 146	81	9
## 147	69	9
## 148	63	9

```

## 149    70      9
## 150    77      9
## 151    75      9
## 152    76      9
## 153    68      9

#returns sum of selected column
mydata[,sum(Ozone, na.rm = TRUE)]

## [1] 4887

#returns sum and standard deviation
mydata[,.(sum(Ozone, na.rm = TRUE), sd(Ozone, na.rm = TRUE))]

##          V1          V2
## 1: 4887 32.98788

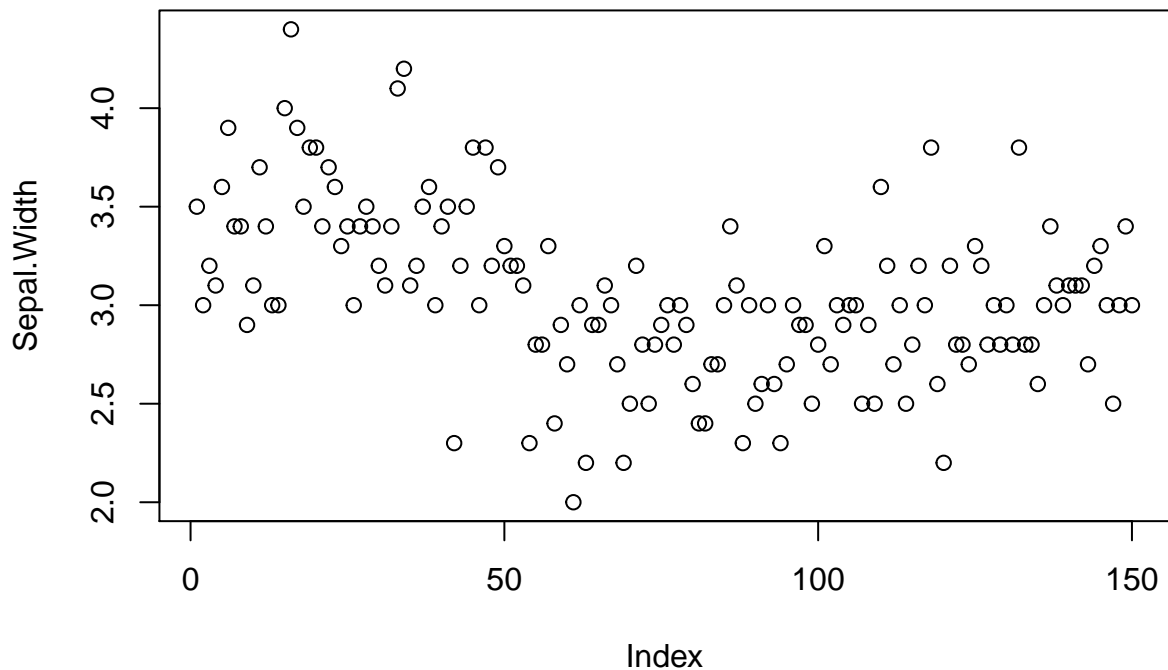
#print and plot
myiris[,{print(Sepal.Length)}]

## [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4 4.8 4.8 4.3 5.8 5.7 5.4
## [18] 5.1 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2 4.7 4.8 5.4 5.2 5.5
## [35] 4.9 5.0 5.5 4.9 4.4 5.1 5.0 4.5 4.4 5.0 5.1 4.8 5.1 4.6 5.3 5.0 7.0
## [52] 6.4 6.9 5.5 6.5 5.7 6.3 4.9 6.6 5.2 5.0 5.9 6.0 6.1 5.6 6.7 5.6 5.8
## [69] 6.2 5.6 5.9 6.1 6.3 6.1 6.4 6.6 6.8 6.7 6.0 5.7 5.5 5.5 5.8 6.0 5.4
## [86] 6.0 6.7 6.3 5.6 5.5 5.5 6.1 5.8 5.0 5.6 5.7 5.7 6.2 5.1 5.7 6.3 5.8
## [103] 7.1 6.3 6.5 7.6 4.9 7.3 6.7 7.2 6.5 6.4 6.8 5.7 5.8 6.4 6.5 7.7 7.7
## [120] 6.0 6.9 5.6 7.7 6.3 6.7 7.2 6.2 6.1 6.4 7.2 7.4 7.9 6.4 6.3 6.1 7.7
## [137] 6.3 6.4 6.0 6.9 6.7 6.9 5.8 6.8 6.7 6.7 6.3 6.5 6.2 5.9

## [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4 4.8 4.8 4.3 5.8 5.7 5.4
## [18] 5.1 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2 4.7 4.8 5.4 5.2 5.5
## [35] 4.9 5.0 5.5 4.9 4.4 5.1 5.0 4.5 4.4 5.0 5.1 4.8 5.1 4.6 5.3 5.0 7.0
## [52] 6.4 6.9 5.5 6.5 5.7 6.3 4.9 6.6 5.2 5.0 5.9 6.0 6.1 5.6 6.7 5.6 5.8
## [69] 6.2 5.6 5.9 6.1 6.3 6.1 6.4 6.6 6.8 6.7 6.0 5.7 5.5 5.5 5.8 6.0 5.4
## [86] 6.0 6.7 6.3 5.6 5.5 5.5 6.1 5.8 5.0 5.6 5.7 5.7 6.2 5.1 5.7 6.3 5.8
## [103] 7.1 6.3 6.5 7.6 4.9 7.3 6.7 7.2 6.5 6.4 6.8 5.7 5.8 6.4 6.5 7.7 7.7
## [120] 6.0 6.9 5.6 7.7 6.3 6.7 7.2 6.2 6.1 6.4 7.2 7.4 7.9 6.4 6.3 6.1 7.7
## [137] 6.3 6.4 6.0 6.9 6.7 6.9 5.8 6.8 6.7 6.7 6.3 6.5 6.2 5.9

myiris[,{plot(Sepal.Width)}]

```

```
## NULL
```

```
#grouping by a variable
```

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

```
##      Species sepalsum
```

```
## 1:   setosa    250.3
```

```
## 2: versicolor  296.8
```

```
## 3:  virginica  329.4
```

```
#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']
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1:         5.1         3.5         1.4         0.2   setosa
## 2:         4.9         3.0         1.4         0.2   setosa
## 3:         4.7         3.2         1.3         0.2   setosa
## 4:         4.6         3.1         1.5         0.2   setosa
## 5:         5.0         3.6         1.4         0.2   setosa
## 6:         5.4         3.9         1.7         0.4   setosa
## 7:         4.6         3.4         1.4         0.3   setosa
## 8:         5.0         3.4         1.5         0.2   setosa
## 9:         4.4         2.9         1.4         0.2   setosa
## 10:        4.9         3.1         1.5         0.1   setosa
## 11:        5.4         3.7         1.5         0.2   setosa
## 12:        4.8         3.4         1.6         0.2   setosa
```

```
## 13:      4.8      3.0      1.4      0.1 setosa
## 14:      4.3      3.0      1.1      0.1 setosa
## 15:      5.8      4.0      1.2      0.2 setosa
## 16:      5.7      4.4      1.5      0.4 setosa
## 17:      5.4      3.9      1.3      0.4 setosa
## 18:      5.1      3.5      1.4      0.3 setosa
## 19:      5.7      3.8      1.7      0.3 setosa
## 20:      5.1      3.8      1.5      0.3 setosa
## 21:      5.4      3.4      1.7      0.2 setosa
## 22:      5.1      3.7      1.5      0.4 setosa
## 23:      4.6      3.6      1.0      0.2 setosa
## 24:      5.1      3.3      1.7      0.5 setosa
## 25:      4.8      3.4      1.9      0.2 setosa
## 26:      5.0      3.0      1.6      0.2 setosa
## 27:      5.0      3.4      1.6      0.4 setosa
## 28:      5.2      3.5      1.5      0.2 setosa
## 29:      5.2      3.4      1.4      0.2 setosa
## 30:      4.7      3.2      1.6      0.2 setosa
## 31:      4.8      3.1      1.6      0.2 setosa
## 32:      5.4      3.4      1.5      0.4 setosa
## 33:      5.2      4.1      1.5      0.1 setosa
## 34:      5.5      4.2      1.4      0.2 setosa
## 35:      4.9      3.1      1.5      0.2 setosa
## 36:      5.0      3.2      1.2      0.2 setosa
## 37:      5.5      3.5      1.3      0.2 setosa
## 38:      4.9      3.6      1.4      0.1 setosa
## 39:      4.4      3.0      1.3      0.2 setosa
## 40:      5.1      3.4      1.5      0.2 setosa
## 41:      5.0      3.5      1.3      0.3 setosa
## 42:      4.5      2.3      1.3      0.3 setosa
## 43:      4.4      3.2      1.3      0.2 setosa
## 44:      5.0      3.5      1.6      0.6 setosa
## 45:      5.1      3.8      1.9      0.4 setosa
## 46:      4.8      3.0      1.4      0.3 setosa
## 47:      5.1      3.8      1.6      0.2 setosa
## 48:      4.6      3.2      1.4      0.2 setosa
## 49:      5.3      3.7      1.5      0.2 setosa
## 50:      5.0      3.3      1.4      0.2 setosa
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
```

```
myiris[c('setosa', 'virginica')]
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1:      5.1      3.5      1.4      0.2 setosa
## 2:      4.9      3.0      1.4      0.2 setosa
## 3:      4.7      3.2      1.3      0.2 setosa
## 4:      4.6      3.1      1.5      0.2 setosa
## 5:      5.0      3.6      1.4      0.2 setosa
## 6:      5.4      3.9      1.7      0.4 setosa
## 7:      4.6      3.4      1.4      0.3 setosa
## 8:      5.0      3.4      1.5      0.2 setosa
## 9:      4.4      2.9      1.4      0.2 setosa
## 10:      4.9      3.1      1.5      0.1 setosa
## 11:      5.4      3.7      1.5      0.2 setosa
## 12:      4.8      3.4      1.6      0.2 setosa
```

## 13:	4.8	3.0	1.4	0.1	setosa
## 14:	4.3	3.0	1.1	0.1	setosa
## 15:	5.8	4.0	1.2	0.2	setosa
## 16:	5.7	4.4	1.5	0.4	setosa
## 17:	5.4	3.9	1.3	0.4	setosa
## 18:	5.1	3.5	1.4	0.3	setosa
## 19:	5.7	3.8	1.7	0.3	setosa
## 20:	5.1	3.8	1.5	0.3	setosa
## 21:	5.4	3.4	1.7	0.2	setosa
## 22:	5.1	3.7	1.5	0.4	setosa
## 23:	4.6	3.6	1.0	0.2	setosa
## 24:	5.1	3.3	1.7	0.5	setosa
## 25:	4.8	3.4	1.9	0.2	setosa
## 26:	5.0	3.0	1.6	0.2	setosa
## 27:	5.0	3.4	1.6	0.4	setosa
## 28:	5.2	3.5	1.5	0.2	setosa
## 29:	5.2	3.4	1.4	0.2	setosa
## 30:	4.7	3.2	1.6	0.2	setosa
## 31:	4.8	3.1	1.6	0.2	setosa
## 32:	5.4	3.4	1.5	0.4	setosa
## 33:	5.2	4.1	1.5	0.1	setosa
## 34:	5.5	4.2	1.4	0.2	setosa
## 35:	4.9	3.1	1.5	0.2	setosa
## 36:	5.0	3.2	1.2	0.2	setosa
## 37:	5.5	3.5	1.3	0.2	setosa
## 38:	4.9	3.6	1.4	0.1	setosa
## 39:	4.4	3.0	1.3	0.2	setosa
## 40:	5.1	3.4	1.5	0.2	setosa
## 41:	5.0	3.5	1.3	0.3	setosa
## 42:	4.5	2.3	1.3	0.3	setosa
## 43:	4.4	3.2	1.3	0.2	setosa
## 44:	5.0	3.5	1.6	0.6	setosa
## 45:	5.1	3.8	1.9	0.4	setosa
## 46:	4.8	3.0	1.4	0.3	setosa
## 47:	5.1	3.8	1.6	0.2	setosa
## 48:	4.6	3.2	1.4	0.2	setosa
## 49:	5.3	3.7	1.5	0.2	setosa
## 50:	5.0	3.3	1.4	0.2	setosa
## 51:	6.3	3.3	6.0	2.5	virginica
## 52:	5.8	2.7	5.1	1.9	virginica
## 53:	7.1	3.0	5.9	2.1	virginica
## 54:	6.3	2.9	5.6	1.8	virginica
## 55:	6.5	3.0	5.8	2.2	virginica
## 56:	7.6	3.0	6.6	2.1	virginica
## 57:	4.9	2.5	4.5	1.7	virginica
## 58:	7.3	2.9	6.3	1.8	virginica
## 59:	6.7	2.5	5.8	1.8	virginica
## 60:	7.2	3.6	6.1	2.5	virginica
## 61:	6.5	3.2	5.1	2.0	virginica
## 62:	6.4	2.7	5.3	1.9	virginica
## 63:	6.8	3.0	5.5	2.1	virginica
## 64:	5.7	2.5	5.0	2.0	virginica
## 65:	5.8	2.8	5.1	2.4	virginica
## 66:	6.4	3.2	5.3	2.3	virginica

## 67:	6.5	3.0	5.5	1.8 virginica	
## 68:	7.7	3.8	6.7	2.2 virginica	
## 69:	7.7	2.6	6.9	2.3 virginica	
## 70:	6.0	2.2	5.0	1.5 virginica	
## 71:	6.9	3.2	5.7	2.3 virginica	
## 72:	5.6	2.8	4.9	2.0 virginica	
## 73:	7.7	2.8	6.7	2.0 virginica	
## 74:	6.3	2.7	4.9	1.8 virginica	
## 75:	6.7	3.3	5.7	2.1 virginica	
## 76:	7.2	3.2	6.0	1.8 virginica	
## 77:	6.2	2.8	4.8	1.8 virginica	
## 78:	6.1	3.0	4.9	1.8 virginica	
## 79:	6.4	2.8	5.6	2.1 virginica	
## 80:	7.2	3.0	5.8	1.6 virginica	
## 81:	7.4	2.8	6.1	1.9 virginica	
## 82:	7.9	3.8	6.4	2.0 virginica	
## 83:	6.4	2.8	5.6	2.2 virginica	
## 84:	6.3	2.8	5.1	1.5 virginica	
## 85:	6.1	2.6	5.6	1.4 virginica	
## 86:	7.7	3.0	6.1	2.3 virginica	
## 87:	6.3	3.4	5.6	2.4 virginica	
## 88:	6.4	3.1	5.5	1.8 virginica	
## 89:	6.0	3.0	4.8	1.8 virginica	
## 90:	6.9	3.1	5.4	2.1 virginica	
## 91:	6.7	3.1	5.6	2.4 virginica	
## 92:	6.9	3.1	5.1	2.3 virginica	
## 93:	5.8	2.7	5.1	1.9 virginica	
## 94:	6.8	3.2	5.9	2.3 virginica	
## 95:	6.7	3.3	5.7	2.5 virginica	
## 96:	6.7	3.0	5.2	2.3 virginica	
## 97:	6.3	2.5	5.0	1.9 virginica	
## 98:	6.5	3.0	5.2	2.0 virginica	
## 99:	6.2	3.4	5.4	2.3 virginica	
## 100:	5.9	3.0	5.1	1.8 virginica	
##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species