Data Frame

This datatype is used to store tabular data.
They are represented by special type of list where every element of the list has to have the same length known as columns.
Unlike matrices Data frame can store different classes of objects like numeric, factor or character.
Data frame are created by calling read.table() or read.csv().
Can be converted to a matrix by calling data.matrix().
Columns have the same length.
Row names should be unique.

Data Frame

Data Frame is created implicitly by the function read.table;
It is also possible to create a data frame with the function data.frame.

> data.frame(x,n)

X <- 1:4, n <- 10 ; M <- c(10,35) ; y <-2:4

> data.frame(x,n)

> data.frame(x,M)

> data.frame(x,y)

Which one will give error.

Data Frame

```
> x <-data.frame(val = 1:4, bool=c(T,T,T,F))
>x
   foo  bar
1  1 TRUE
2  2  TRUE
3  3  TRUE
4  4  FALSE
> nrow(x)
[1] 4
> ncol(x)
[1] 2
```

Data Frame

Whenever we upload data from a file it will create a data frame.

```
airquality<-read.table("hw1_data.csv",
header=TRUE,    sep=",",)
```

Cleaning Data

```
>airquality[1:6, ]
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
> good <- complete.cases(airquality) ##  It will select only those rows which dont have NA
>airquality_clean<-airquality[good, ]
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
```

Questions

Extract first two rows of data frame

airquality[1:2,1:6]

How may observations are in this data frame

dim(airquality)

What is the value of Ozone in 47th row?

airquality$Ozone[[47]]

Extract the rows where Ozone value is above 31 and temp value are above 90.

airquality[airquality$Temp>90&&airquality$Ozone>31,1:6]

Take the mean of Solar.R, use function mean.