```
R version 3.3.3 (2017-03-06) -- "Another Canoe"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> getwd()
[1] "C:/Users/bhaskarnilh/Documents/Official Learning centre/Edx/MITx/The Analytics Edge/2nd week
-LR"
> data=read.csv("climate_change.csv")
> str(data)
'data.frame':   308 obs. of  11 variables:
 $ Year    : int  1983 1983 1983 1983 1983 1983 1983 1983 1984 1984 ...
 $ Month   : int  5 6 7 8 9 10 11 12 1 2 ...
 $ MEI     : num  2.556 2.167 1.741 1.13 0.428 ...
 $ CO2     : num  346 346 344 342 340 ...
 $ CH4     : num  1639 1634 1633 1631 1648 ...
 $ N2O     : num  304 304 304 304 304 ...
 $ CFC.11  : num  191 192 193 194 194 ...
 $ CFC.12  : num  350 352 354 356 357 ...
 $ TSI     : num  1366 1366 1366 1366 1366 ...
 $ Aerosols: num  0.0863 0.0794 0.0731 0.0673 0.0619 0.0569 0.0524 0.0486 0.0451 0.0416 ...
 $ Temp    : num  0.109 0.118 0.137 0.176 0.149 0.093 0.232 0.078 0.089 0.013 ...
>     dt = sort(sample(nrow(data), nrow(data)*.7))
>     train<-data[dt,]
>     test<-data[-dt,]
>
>     climate_train<-data[dt,]
> climate_test<-dat[-dt,]
Error: object 'dat' not found
> climate_test<-data[-dt,]
> str(climate_train)
'data.frame':   215 obs. of  11 variables:
 $ Year    : int  1983 1983 1983 1983 1983 1983 1984 1984 1984 1984 ...
 $ Month   : int  5 6 7 8 9 11 1 2 4 5 ...
 $ MEI     : num  2.556 2.167 1.741 1.13 0.428 ...
 $ CO2     : num  346 346 344 342 340 ...
 $ CH4     : num  1639 1634 1633 1631 1648 ...
 $ N2O     : num  304 304 304 304 304 ...
 $ CFC.11  : num  191 192 193 194 194 ...
 $ CFC.12  : num  350 352 354 356 357 ...
 $ TSI     : num  1366 1366 1366 1366 1366 ...
 $ Aerosols: num  0.0863 0.0794 0.0731 0.0673 0.0619 0.0524 0.0451 0.0416 0.0352 0.0324 ...
 $ Temp    : num  0.109 0.118 0.137 0.176 0.149 0.232 0.089 0.013 -0.019 0.065 ...
> model1=lm(Temp~MEI+CO2+CH4+N2O+CFC.11+CFC.12+TSI+Aerosols, data=climate_train)
> summary(model1)

Call:
lm(formula = Temp ~ MEI + CO2 + CH4 + N2O + CFC.11 + CFC.12 +
    TSI + Aerosols, data = climate_train)

Residuals:
      Min        1Q    Median        3Q       Max
-0.257292 -0.068086  0.001382  0.062702  0.316031
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.127e+02  2.394e+01  -4.709 4.56e-06 ***
MEI          6.688e-02  7.911e-03   8.454 5.13e-15 ***
CO2          4.745e-03  2.751e-03   1.725 0.086036 .
CH4         -2.518e-04  6.252e-04  -0.403 0.687612
N2O         -1.484e-02  9.524e-03  -1.558 0.120762
CFC.11      -7.039e-03  1.797e-03  -3.918 0.000122 ***
CFC.12       4.364e-03  1.084e-03   4.024 8.03e-05 ***
TSI          8.487e-02  1.739e-02   4.882 2.10e-06 ***
Aerosols    -1.645e+00  2.467e-01  -6.668 2.34e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09583 on 206 degrees of freedom
Multiple R-squared:  0.7317,    Adjusted R-squared:  0.7212
F-statistic: 70.21 on 8 and 206 DF,  p-value: < 2.2e-16

> dataTrain=subset(data,data$Year <=2006)
> dataTest=subset(data,data$Year>2006)
> summary(datTrain)
Error in summary(datTrain) : object 'datTrain' not found
> summary(dataTrain)
      Year          Month             MEI                CO2
 Min.   :1983   Min.   : 1.000   Min.   :-1.5860   Min.   :340.2
 1st Qu.:1989   1st Qu.: 4.000   1st Qu.:-0.3230   1st Qu.:352.3
 Median :1995   Median : 7.000   Median : 0.3085   Median :359.9
 Mean   :1995   Mean   : 6.556   Mean   : 0.3419   Mean   :361.4
 3rd Qu.:2001   3rd Qu.:10.000   3rd Qu.: 0.8980   3rd Qu.:370.6
 Max.   :2006   Max.   :12.000   Max.   : 3.0010   Max.   :385.0
      CH4            N2O            CFC.11          CFC.12           TSI
 Min.   :1630   Min.   :303.7   Min.   :191.3   Min.   :350.1   Min.   :1365
 1st Qu.:1716   1st Qu.:307.7   1st Qu.:249.6   1st Qu.:462.5   1st Qu.:1366
 Median :1759   Median :310.8   Median :260.4   Median :522.1   Median :1366
 Mean   :1746   Mean   :311.7   Mean   :252.5   Mean   :494.2   Mean   :1366
 3rd Qu.:1782   3rd Qu.:316.1   3rd Qu.:267.4   3rd Qu.:541.0   3rd Qu.:1366
 Max.   :1808   Max.   :320.5   Max.   :271.5   Max.   :543.8   Max.   :1367
    Aerosols            Temp
 Min.   :0.00160   Min.   :-0.2820
 1st Qu.:0.00270   1st Qu.: 0.1180
 Median :0.00620   Median : 0.2325
 Mean   :0.01772   Mean   : 0.2478
 3rd Qu.:0.01400   3rd Qu.: 0.4065
 Max.   :0.14940   Max.   : 0.7390
> model1=lm(Temp~MEI+CO2+CH4+N2O+CFC.11+CFC.12+TSI+Aerosols)
Error in eval(expr, envir, enclos) : object 'Temp' not found
> model1=lm(Temp~MEI+CO2+CH4+N2O+CFC.11+CFC.12+TSI+Aerosols, data=dataTrain)
> summary(dataTrain)
      Year          Month             MEI                CO2
 Min.   :1983   Min.   : 1.000   Min.   :-1.5860   Min.   :340.2
 1st Qu.:1989   1st Qu.: 4.000   1st Qu.:-0.3230   1st Qu.:352.3
 Median :1995   Median : 7.000   Median : 0.3085   Median :359.9
 Mean   :1995   Mean   : 6.556   Mean   : 0.3419   Mean   :361.4
 3rd Qu.:2001   3rd Qu.:10.000   3rd Qu.: 0.8980   3rd Qu.:370.6
 Max.   :2006   Max.   :12.000   Max.   : 3.0010   Max.   :385.0
      CH4            N2O            CFC.11          CFC.12           TSI
 Min.   :1630   Min.   :303.7   Min.   :191.3   Min.   :350.1   Min.   :1365
 1st Qu.:1716   1st Qu.:307.7   1st Qu.:249.6   1st Qu.:462.5   1st Qu.:1366
 Median :1759   Median :310.8   Median :260.4   Median :522.1   Median :1366
 Mean   :1746   Mean   :311.7   Mean   :252.5   Mean   :494.2   Mean   :1366
 3rd Qu.:1782   3rd Qu.:316.1   3rd Qu.:267.4   3rd Qu.:541.0   3rd Qu.:1366
 Max.   :1808   Max.   :320.5   Max.   :271.5   Max.   :543.8   Max.   :1367
    Aerosols            Temp
 Min.   :0.00160   Min.   :-0.2820
 1st Qu.:0.00270   1st Qu.: 0.1180
 Median :0.00620   Median : 0.2325
 Mean   :0.01772   Mean   : 0.2478
 3rd Qu.:0.01400   3rd Qu.: 0.4065
 Max.   :0.14940   Max.   : 0.7390
> summary(model1)
```

```
Call:
lm(formula = Temp ~ MEI + CO2 + CH4 + N2O + CFC.11 + CFC.12 +
    TSI + Aerosols, data = dataTrain)

Residuals:
     Min       1Q   Median       3Q      Max
-0.25888 -0.05913 -0.00082  0.05649  0.32433

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.246e+02  1.989e+01  -6.265 1.43e-09 ***
MEI          6.421e-02  6.470e-03   9.923  < 2e-16 ***
CO2          6.457e-03  2.285e-03   2.826  0.00505 **
CH4          1.240e-04  5.158e-04   0.240  0.81015
N2O         -1.653e-02  8.565e-03  -1.930  0.05467 .
CFC.11      -6.631e-03  1.626e-03  -4.078 5.96e-05 ***
CFC.12       3.808e-03  1.014e-03   3.757  0.00021 ***
TSI          9.314e-02  1.475e-02   6.313 1.10e-09 ***
Aerosols    -1.538e+00  2.133e-01  -7.210 5.41e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09171 on 275 degrees of freedom
Multiple R-squared:  0.7509,    Adjusted R-squared:  0.7436
F-statistic: 103.6 on 8 and 275 DF,  p-value: < 2.2e-16

> cor(model1)
Error in cor(model1) : supply both 'x' and 'y' or a matrix-like 'x'
> cor(dataTrain)
                Year          Month            MEI         CO2         CH4
Year       1.00000000 -0.0279419602 -0.0369876842  0.98274939  0.91565945
Month     -0.02794196  1.0000000000  0.0008846905 -0.10673246  0.01856866
MEI       -0.03698768  0.0008846905  1.0000000000 -0.04114717 -0.03341930
CO2        0.98274939 -0.1067324607 -0.0411471651  1.00000000  0.87727963
CH4        0.91565945  0.0185686624 -0.0334193014  0.87727963  1.00000000
N2O        0.99384523  0.0136315303 -0.0508197755  0.97671982  0.89983864
CFC.11     0.56910643 -0.0131112236  0.0690004387  0.51405975  0.77990402
CFC.12     0.89701166  0.0006751102  0.0082855443  0.85268963  0.96361625
TSI        0.17030201 -0.0346061935 -0.1544919227  0.17742893  0.24552844
Aerosols  -0.34524670  0.0148895406  0.3402377871 -0.35615480 -0.26780919
Temp       0.78679714 -0.0998567411  0.1724707512  0.78852921  0.70325502
                 N2O      CFC.11        CFC.12         TSI    Aerosols
Year       0.99384523  0.56910643  0.8970116635  0.17030201 -0.34524670
Month      0.01363153 -0.01311122  0.0006751102 -0.03460619  0.01488954
MEI       -0.05081978  0.06900044  0.0082855443 -0.15449192  0.34023779
CO2        0.97671982  0.51405975  0.8526896272  0.17742893 -0.35615480
CH4        0.89983864  0.77990402  0.9636162478  0.24552844 -0.26780919
N2O        1.00000000  0.52247732  0.8679307757  0.19975668 -0.33705457
CFC.11     0.52247732  1.00000000  0.8689851828  0.27204596 -0.04392120
CFC.12     0.86793078  0.86898518  1.0000000000  0.25530281 -0.22513124
TSI        0.19975668  0.27204596  0.2553028138  1.00000000  0.05211651
Aerosols  -0.33705457 -0.04392120 -0.2251312440  0.05211651  1.00000000
Temp       0.77863893  0.40771029  0.6875575483  0.24338269 -0.38491375
                Temp
Year       0.78679714
Month     -0.09985674
MEI        0.17247075
CO2        0.78852921
CH4        0.70325502
N2O        0.77863893
CFC.11     0.40771029
CFC.12     0.68755755
TSI        0.24338269
Aerosols  -0.38491375
Temp       1.00000000
> model2=lm(N2O~MEI+TSI+Aerosols, data=dataTrain)
> summary(model2)

Call:
```

```
lm(formula = N2O ~ MEI + TSI + Aerosols, data = dataTrain)

Residuals:
    Min      1Q  Median      3Q     Max
-6.8849 -4.5183  0.4462  2.7020  8.8085

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3549.5201   896.8079  -3.958 9.59e-05 ***
MEI             0.6073     0.3009   2.018   0.0445 *
TSI             2.8271     0.6565   4.306 2.30e-05 ***
Aerosols      -61.8079     9.2217  -6.702 1.13e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.351 on 280 degrees of freedom
Multiple R-squared:  0.173,     Adjusted R-squared:  0.1641
F-statistic: 19.52 on 3 and 280 DF,  p-value: 1.6e-11

> model2=lm(Temp~N2O+MEI+TSI+Aerosols, data=dataTrain)
> summary(model2)

Call:
lm(formula = Temp ~ N2O + MEI + TSI + Aerosols, data = dataTrain)

Residuals:
     Min       1Q   Median       3Q      Max
-0.27916 -0.05975 -0.00595  0.05672  0.34195

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.162e+02  2.022e+01  -5.747 2.37e-08 ***
N2O          2.532e-02  1.311e-03  19.307  < 2e-16 ***
MEI          6.419e-02  6.652e-03   9.649  < 2e-16 ***
TSI          7.949e-02  1.487e-02   5.344 1.89e-07 ***
Aerosols    -1.702e+00  2.180e-01  -7.806 1.19e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09547 on 279 degrees of freedom
Multiple R-squared:  0.7261,    Adjusted R-squared:  0.7222
F-statistic: 184.9 on 4 and 279 DF,  p-value: < 2.2e-16

> newStepModel=step(model1)
Start:  AIC=-1348.16
Temp ~ MEI + CO2 + CH4 + N2O + CFC.11 + CFC.12 + TSI + Aerosols

           Df Sum of Sq    RSS     AIC
- CH4       1   0.00049 2.3135 -1350.1
<none>                 2.3130 -1348.2
- N2O       1   0.03132 2.3443 -1346.3
- CO2       1   0.06719 2.3802 -1342.0
- CFC.12    1   0.11874 2.4318 -1335.9
- CFC.11    1   0.13986 2.4529 -1333.5
- TSI       1   0.33516 2.6482 -1311.7
- Aerosols  1   0.43727 2.7503 -1301.0
- MEI       1   0.82823 3.1412 -1263.2

Step:  AIC=-1350.1
Temp ~ MEI + CO2 + N2O + CFC.11 + CFC.12 + TSI + Aerosols

           Df Sum of Sq    RSS     AIC
<none>                 2.3135 -1350.1
- N2O       1   0.03133 2.3448 -1348.3
- CO2       1   0.06672 2.3802 -1344.0
- CFC.12    1   0.13023 2.4437 -1336.5
- CFC.11    1   0.13938 2.4529 -1335.5
- TSI       1   0.33500 2.6485 -1313.7
- Aerosols  1   0.43987 2.7534 -1302.7
- MEI       1   0.83118 3.1447 -1264.9
```

```
> summary(newStepModel)

Call:
lm(formula = Temp ~ MEI + CO2 + N2O + CFC.11 + CFC.12 + TSI +
    Aerosols, data = dataTrain)

Residuals:
     Min       1Q   Median       3Q      Max
-0.25770 -0.05994 -0.00104  0.05588  0.32203

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.245e+02  1.985e+01  -6.273 1.37e-09 ***
MEI          6.407e-02  6.434e-03   9.958  < 2e-16 ***
CO2          6.402e-03  2.269e-03   2.821 0.005129 **
N2O         -1.602e-02  8.287e-03  -1.933 0.054234 .
CFC.11      -6.609e-03  1.621e-03  -4.078 5.95e-05 ***
CFC.12       3.868e-03  9.812e-04   3.942 0.000103 ***
TSI          9.312e-02  1.473e-02   6.322 1.04e-09 ***
Aerosols    -1.540e+00  2.126e-01  -7.244 4.36e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09155 on 276 degrees of freedom
Multiple R-squared:  0.7508,    Adjusted R-squared:  0.7445
F-statistic: 118.8 on 7 and 276 DF,  p-value: < 2.2e-16

> help(step)
starting httpd help server ... done
> predict_Temp=predict(newStepModel1,newdat=dataTest)
Error in predict(newStepModel1, newdat = dataTest) :
  object 'newStepModel1' not found
> predict_Temp=predict(newStepModel1,newdata=dataTest)
Error in predict(newStepModel1, newdata = dataTest) :
  object 'newStepModel1' not found
> predict_Temp=predict(newStepModel,newdata=dataTest)
> summary(predict_temp)
Error in summary(predict_temp) : object 'predict_temp' not found
> summary(predict_Temp)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3142  0.3418  0.3771  0.3832  0.4245  0.4678
> predict_Temp
      285       286       287       288       289       290       291       292
0.4677808 0.4435404 0.4265541 0.4299162 0.4455113 0.4151422 0.4097367 0.3839390
      293       294       295       296       297       298       299       300
0.3255595 0.3274147 0.3231401 0.3316704 0.3522134 0.3313129 0.3142112 0.3703410
      301       302       303       304       305       306       307       308
0.4162213 0.4391458 0.4237965 0.3913679 0.3587615 0.3451991 0.3607087 0.3638076
> summary(data_Test)
Error in summary(data_Test) : object 'data_Test' not found
> summary(dataTest)
      Year          Month            MEI               CO2
 Min.   :2007   Min.   : 1.00   Min.   :-1.6350   Min.   :380.9
 1st Qu.:2007   1st Qu.: 3.75   1st Qu.:-1.0437   1st Qu.:383.1
 Median :2008   Median : 6.50   Median :-0.5305   Median :384.5
 Mean   :2008   Mean   : 6.50   Mean   :-0.5098   Mean   :384.7
 3rd Qu.:2008   3rd Qu.: 9.25   3rd Qu.:-0.0360   3rd Qu.:386.1
 Max.   :2008   Max.   :12.00   Max.   : 0.9740   Max.   :388.5
      CH4            N2O           CFC.11          CFC.12           TSI
 Min.   :1772   Min.   :320.3   Min.   :244.1   Min.   :534.9   Min.   :1366
 1st Qu.:1792   1st Qu.:320.6   1st Qu.:244.6   1st Qu.:535.1   1st Qu.:1366
 Median :1798   Median :321.3   Median :246.2   Median :537.0   Median :1366
 Mean   :1797   Mean   :321.1   Mean   :245.9   Mean   :536.7   Mean   :1366
 3rd Qu.:1804   3rd Qu.:321.4   3rd Qu.:246.6   3rd Qu.:537.4   3rd Qu.:1366
 Max.   :1814   Max.   :322.2   Max.   :248.4   Max.   :539.2   Max.   :1366
    Aerosols            Temp
 Min.   :0.003100   Min.   :0.074
 1st Qu.:0.003600   1st Qu.:0.307
 Median :0.004100   Median :0.380
 Mean   :0.004071   Mean   :0.363
```

```
 3rd Qu.:0.004500   3rd Qu.:0.414
 Max.   :0.005400   Max.   :0.601
> model5=lm(Temp~MEI+CO2+CH4+N2O+CFC.11+CFC.12+TSI+Aerosols, data=dataTest)
> summary(model5)

Call:
lm(formula = Temp ~ MEI + CO2 + CH4 + N2O + CFC.11 + CFC.12 +
    TSI + Aerosols, data = dataTest)

Residuals:
      Min        1Q    Median        3Q       Max
-0.164432 -0.044591 -0.002422  0.039589  0.159291

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.517e+03  1.158e+03   1.310   0.2100
MEI          3.175e-02  4.571e-02   0.695   0.4979
CO2          4.379e-03  1.784e-02   0.245   0.8094
CH4         -2.262e-03  4.250e-03  -0.532   0.6024
N2O         -1.180e-01  1.591e-01  -0.742   0.4697
CFC.11       3.817e-02  1.494e-01   0.255   0.8018
CFC.12      -7.364e-02  1.630e-01  -0.452   0.6578
TSI         -1.060e+00  8.147e-01  -1.301   0.2130
Aerosols     1.484e+02  6.559e+01   2.263   0.0389 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08178 on 15 degrees of freedom
Multiple R-squared:  0.6249,    Adjusted R-squared:  0.4249
F-statistic: 3.124 on 8 and 15 DF,  p-value: 0.02734

>
```