```
          > airquality
    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
7      23     299  8.6   65     5   7
8      19      99 13.8   59     5   8
9       8      19 20.1   61     5   9
10     NA     194  8.6   69     5  10
> summary(airquality)
     Ozone            Solar.R           Wind             Temp
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00
 1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00
 Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88
 3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00
 NA's   :37       NA's   :7
     Month            Day
 Min.   :5.000   Min.   : 1.0
 1st Qu.:6.000   1st Qu.: 8.0
 Median :7.000   Median :16.0
 Mean   :6.993   Mean   :15.8
 3rd Qu.:8.000   3rd Qu.:23.0
 Max.   :9.000   Max.   :31.0
```

Data Cleaning

```
> airquality$Ozone=ifelse(is.na(airquality$Ozone),median(airquality$Ozone,na.
rm=TRUE),airquality$Ozone)
> summary(airquality)
     Ozone            Solar.R           Wind             Temp
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00
 1st Qu.: 21.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00
 Mean   : 39.56   Mean   :185.9   Mean   : 9.958   Mean   :77.88
 3rd Qu.: 46.00   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00
                  NA's   :7
     Month            Day
 Min.   :5.000   Min.   : 1.0
 1st Qu.:6.000   1st Qu.: 8.0
 Median :7.000   Median :16.0
 Mean   :6.993   Mean   :15.8
 3rd Qu.:8.000   3rd Qu.:23.0
 Max.   :9.000   Max.   :31.0


> airquality$Solar.R=ifelse(is.na(airquality$Solar.R),median(airquality$Solar
.R,na.rm=TRUE),airquality$Solar.R)
> summary(airquality)
     Ozone            Solar.R           Wind             Temp
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00
```

```
 1st Qu.: 21.00    1st Qu.:120.0    1st Qu.: 7.400    1st Qu.:72.00
 Median : 31.50    Median :205.0    Median : 9.700    Median :79.00
 Mean   : 39.56    Mean   :186.8    Mean   : 9.958    Mean   :77.88
 3rd Qu.: 46.00    3rd Qu.:256.0    3rd Qu.:11.500    3rd Qu.:85.00
 Max.   :168.00    Max.   :334.0    Max.   :20.700    Max.   :97.00
      Month              Day
 Min.   :5.000    Min.   : 1.0
 1st Qu.:6.000    1st Qu.: 8.0
 Median :7.000    Median :16.0
 Mean   :6.993    Mean   :15.8
 3rd Qu.:8.000    3rd Qu.:23.0
 Max.   :9.000    Max.   :31.0
```

```
> head(airquality)
  Ozone Solar.R Wind Temp Month Day
1  41.0     190  7.4   67     5   1
2  36.0     118  8.0   72     5   2
3  12.0     149 12.6   74     5   3
4  18.0     313 11.5   62     5   4
5  31.5     205 14.3   56     5   5
6  28.0     205 14.9   66     5   6
```

Data Transformation

```
> breaks=c(0,50,100,150,200,250,300,350)
> breaks
[1]   0  50 100 150 200 250 300 350
> airquality$Solar.R=cut(airquality$Solar.R,breaks=breaks,include.lowest=TRUE
)
> head(airquality)
  Ozone   Solar.R Wind Temp Month Day
1  41.0 (150,200]  7.4   67     5   1
2  36.0 (100,150]  8.0   72     5   2
3  12.0 (100,150] 12.6   74     5   3
4  18.0 (300,350] 11.5   62     5   4
5  31.5 (200,250] 14.3   56     5   5
6  28.0 (200,250] 14.9   66     5   6
```

```
air=airquality
> air$Month=gsub(13,"May",air$Month)
> air$Month=gsub(14,"June",air$Month)
> air$Month=gsub(15,"July",air$Month)
> air$Month=gsub(16,"August",air$Month)
> air$Month=gsub(17,"Sept",air$Month)
> head(air)
  Ozone   Solar.R Wind Temp Month Day
1  41.0 (150,200]  7.4   67     5   1
2  36.0 (100,150]  8.0   72     5   2
3  12.0 (100,150] 12.6   74     5   3
4  18.0 (300,350] 11.5   62     5   4
5  31.5 (200,250] 14.3   56     5   5
6  28.0 (200,250] 14.9   66     5   6
```

Data model Building

```
> data("airquality")
> str(airquality)
'data.frame':   153 obs. of  6 variables:
 $ Ozone  : int  41 36 12 18 NA 28 23 19 8 NA ...
 $ Solar.R: int  190 118 149 313 NA NA 299 99 19 194 ...
 $ Wind   : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
 $ Temp   : int  67 72 74 62 56 66 65 59 61 69 ...
 $ Month  : int  5 5 5 5 5 5 5 5 5 5 ...
 $ Day    : int  1 2 3 4 5 6 7 8 9 10 ...
> head(airquality)
  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
> air=airquality
> head(air)
  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
> summary(air)
     Ozone           Solar.R           Wind             Temp
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00
 1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00
 Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88
 3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00
 NA's   :37       NA's   :7
     Month            Day
 Min.   :5.000   Min.   : 1.0
 1st Qu.:6.000   1st Qu.: 8.0
 Median :7.000   Median :16.0
 Mean   :6.993   Mean   :15.8
 3rd Qu.:8.000   3rd Qu.:23.0
 Max.   :9.000   Max.   :31.0


> for(i in 1:nrow(air)){
 if(is.na(air[i,"Ozone"])){
 air[i,"Ozone"]<-mean(air[which(air[,"Month"]==air[i,"month"]),"Ozone"],na.rm
=TRUE)
 }
 #input monthly mean in Solar.R
 if(is.na(air[i,"Solar.R"])){
 air[i,"Solar.R"]<-mean(air[which(air[,"Month"]==air[i,"Month"]),"Solar.R"],n
a.rm=TRUE)
 }
 }
> summary(air)
     Ozone           Solar.R           Wind             Temp
 Min.   :  1.00   Min.   :144.6   Min.   : 1.700   Min.   :56.00
 1st Qu.: 18.00   1st Qu.:163.4   1st Qu.: 7.400   1st Qu.:72.00
 Median : 31.50   Median :179.5   Median : 9.700   Median :79.00
 Mean   : 42.13   Mean   :179.8   Mean   : 9.958   Mean   :77.88
 3rd Qu.: 63.25   3rd Qu.:188.0   3rd Qu.:11.500   3rd Qu.:85.00
```

```
 Max.   :168.00   Max.   :218.0   Max.   :20.700   Max.   :97.00
 NA's   :37
      Month             Day
 Min.   :5.000   Min.   : 1.0
 1st Qu.:6.000   1st Qu.: 8.0
 Median :7.000   Median :16.0
 Mean   :6.993   Mean   :15.8
 3rd Qu.:8.000   3rd Qu.:23.0
 Max.   :9.000   Max.   :31.0

> normalize<-function(x){
+ return(x-min(x)/(max(x)-min(x)))
+ }
> air<-normalize(air)
> summary(air)
     Ozone         Solar.R          Wind            Temp           Month
 Min.   : NA   Min.   : NA   Min.   : NA   Min.   : NA   Min.   : NA
 1st Qu.: NA   1st Qu.: NA   1st Qu.: NA   1st Qu.: NA   1st Qu.: NA
 Median : NA   Median : NA   Median : NA   Median : NA   Median : NA
 Mean   :NaN   Mean   :NaN   Mean   :NaN   Mean   :NaN   Mean   :NaN
 3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA
 Max.   : NA   Max.   : NA   Max.   : NA   Max.   : NA   Max.   : NA
 NA's   :153   NA's   :153   NA's   :153   NA's   :153   NA's   :153
      Day
 Min.   : NA
 1st Qu.: NA
 Median : NA
 Mean   :NaN
 3rd Qu.: NA
 Max.   : NA
 NA's   :153
> str(air)
'data.frame':   153 obs. of  6 variables:
 $ Ozone  : num  NA NA NA NA NA NA NA NA NA NA ...
 $ Solar.R: num  NA NA NA NA NA NA NA NA NA NA ...
 $ Wind   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ Temp   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ Month  : num  NA NA NA NA NA NA NA NA NA NA ...
 $ Day    : num  NA NA NA NA NA NA NA NA NA NA ...


> x=air[,"Ozone"]
> y=air[,"Solar.R"]
> plot(x~y)
>
> plot(y~x)
> model=lm(y~x)
> x=airquality[,"Ozone"]
> y=airquality[,"Solar.R"]
> plot(x~y)
> model1=lm(y~x)
> model1

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
   144.6306       0.9542


>
> abline(model1,col="pink")
> abline(model1,col="Blue")
```

```
> plot(y~x)
> model1=lm(y~x)
> model1

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)              x
   144.6306         0.9542

> abline(model1,col="Red")
> abline(model1,row="Purple")
> abline(model1,row="Green")
> abline(model1,col="Red")
> abline(model1,co="Blue")
> x=airquality[,"Ozone"]
> y=airquality[,"Wind"]
> model1=lm(y~x)
> abline(model1,col="red")
> plot(y~x)
> model1=lm(y~x)
> model1

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)              x
   12.60843      -0.06519
```