Task4

Darya Nemirich

13 May 2019

Upload the dataframe.

```
weather <- readRDS("D:/Bioinformatics and System Biology/2nd term/R/R_classwork/Task4/weather.rds")
```

Let's look on the table and on the information about its variables.

```
head(weather)
```

```
##   X year month           measure X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12
## 1 1 2014    12  Max.TemperatureF 64 42 51 43 42 45 38 29 49  48  39  39
## 2 2 2014    12 Mean.TemperatureF 52 38 44 37 34 42 30 24 39  43  36  35
## 3 3 2014    12  Min.TemperatureF 39 33 37 30 26 38 21 18 29  38  32  31
## 4 4 2014    12    Max.Dew.PointF 46 40 49 24 37 45 36 28 49  45  37  28
## 5 5 2014    12    MeanDew.PointF 40 27 42 21 25 40 20 16 41  39  31  27
## 6 6 2014    12     Min.DewpointF 26 17 24 13 12 36 -3  3 28  37  27  25
##   X13 X14 X15 X16 X17 X18 X19 X20 X21 X22 X23 X24 X25 X26 X27 X28 X29 X30
## 1  42  45  42  44  49  44  37  36  36  44  47  46  59  50  52  52  41  30
## 2  37  39  37  40  45  40  33  32  33  39  45  44  52  44  45  46  36  26
## 3  32  33  32  35  41  36  29  27  30  33  42  41  44  37  38  40  30  22
## 4  28  29  33  42  46  34  25  30  30  39  45  46  58  31  34  42  26  10
## 5  26  27  29  36  41  30  22  24  27  34  42  44  43  29  31  35  20   4
## 6  24  25  27  30  32  26  20  20  25  25  37  41  29  28  29  27  10  -6
##   X31
## 1  30
## 2  25
## 3  20
## 4   8
## 5   5
## 6   1
```

```
str(weather)
```

```
## 'data.frame':    286 obs. of  35 variables:
##  $ X      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ year   : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
##  $ month  : int  12 12 12 12 12 12 12 12 12 12 ...
##  $ measure: chr  "Max.TemperatureF" "Mean.TemperatureF" "Min.TemperatureF" "Max.Dew.PointF" ...
##  $ X1     : chr  "64" "52" "39" "46" ...
##  $ X2     : chr  "42" "38" "33" "40" ...
##  $ X3     : chr  "51" "44" "37" "49" ...
##  $ X4     : chr  "43" "37" "30" "24" ...
##  $ X5     : chr  "42" "34" "26" "37" ...
##  $ X6     : chr  "45" "42" "38" "45" ...
##  $ X7     : chr  "38" "30" "21" "36" ...
##  $ X8     : chr  "29" "24" "18" "28" ...
##  $ X9     : chr  "49" "39" "29" "49" ...
##  $ X10    : chr  "48" "43" "38" "45" ...
##  $ X11    : chr  "39" "36" "32" "37" ...
##  $ X12    : chr  "39" "35" "31" "28" ...
##  $ X13    : chr  "42" "37" "32" "28" ...
##  $ X14    : chr  "45" "39" "33" "29" ...
##  $ X15    : chr  "42" "37" "32" "33" ...
##  $ X16    : chr  "44" "40" "35" "42" ...
##  $ X17    : chr  "49" "45" "41" "46" ...
##  $ X18    : chr  "44" "40" "36" "34" ...
##  $ X19    : chr  "37" "33" "29" "25" ...
##  $ X20    : chr  "36" "32" "27" "30" ...
##  $ X21    : chr  "36" "33" "30" "30" ...
##  $ X22    : chr  "44" "39" "33" "39" ...
##  $ X23    : chr  "47" "45" "42" "45" ...
##  $ X24    : chr  "46" "44" "41" "46" ...
##  $ X25    : chr  "59" "52" "44" "58" ...
##  $ X26    : chr  "50" "44" "37" "31" ...
##  $ X27    : chr  "52" "45" "38" "34" ...
##  $ X28    : chr  "52" "46" "40" "42" ...
##  $ X29    : chr  "41" "36" "30" "26" ...
##  $ X30    : chr  "30" "26" "22" "10" ...
##  $ X31    : chr  "30" "25" "20" "8" ...
```

```r
complete.cases(weather)
```

```
##   [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [12]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [23]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [34]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [45] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [56] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [67]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [78]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [89] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [100] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [111]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [122]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [144] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [155]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [166]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [177]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [188]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
## [199] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [210] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [221]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [232]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
## [243] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [254] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [265] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [276] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

There are several NAs in the table. There is an additional column named X, which is, as I suppose, represents a day of the measurement. Variables are rows, not columns. Numeric observations are not numeric.

Let's get rid of X varible.

```r
weather$X <- NULL
```

Uploading necessary libraries

```
library(tidyr)
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
```

Transform the data from wide to long and then vice versa.

```
weather_tidy <- gather(weather, day, value, X1:X31)
weather_tidy <- spread(weather_tidy, measure, value)
head(weather_tidy, 15)
```

```
##    year month day CloudCover    Events Max.Dew.PointF Max.Gust.SpeedMPH
## 1  2014    12  X1          6      Rain             46                29
## 2  2014    12 X10          8      Rain             45                29
## 3  2014    12 X11          8 Rain-Snow             37                28
## 4  2014    12 X12          7      Snow             28                21
## 5  2014    12 X13          5                       28                23
## 6  2014    12 X14          4                       29                20
## 7  2014    12 X15          2                       33                21
## 8  2014    12 X16          8      Rain             42                10
## 9  2014    12 X17          8      Rain             46                26
## 10 2014    12 X18          7      Rain             34                30
## 11 2014    12 X19          4                       25                23
## 12 2014    12  X2          7 Rain-Snow             40                29
## 13 2014    12 X20          6      Snow             30                26
## 14 2014    12 X21          8      Snow             30                20
## 15 2014    12 X22          7      Rain             39                22
##    Max.Humidity Max.Sea.Level.PressureIn Max.TemperatureF
## 1            74                    30.45               64
## 2           100                    29.58               48
## 3            92                    29.81               39
## 4            85                    29.88               39
## 5            75                    29.86               42
## 6            82                    29.91               45
## 7            89                    30.15               42
## 8            96                    30.17               44
## 9           100                    29.91               49
## 10           89                    29.87               44
## 11           69                    30.15               37
## 12           92                    30.71               42
## 13           89                    30.31               36
## 14           85                    30.37               36
## 15           89                     30.4               44
##    Max.VisibilityMiles Max.Wind.SpeedMPH Mean.Humidity
## 1                   10                22            63
## 2                   10                23            95
## 3                   10                21            87
## 4                   10                16            75
## 5                   10                17            65
## 6                   10                15            68
## 7                   10                15            75
## 8                   10                 8            85
## 9                   10                20            85
## 10                  10                23            73
## 11                  10                17            63
## 12                  10                24            72
## 13                  10                21            79
## 14                  10                16            77
## 15                  10                18            79
##    Mean.Sea.Level.PressureIn Mean.TemperatureF Mean.VisibilityMiles
## 1                      30.13                52                   10
## 2                       29.5                43                    3
## 3                      29.61                36                    7
## 4                      29.85                35                   10
## 5                      29.82                37                   10
## 6                      29.83                39                   10
## 7                      30.05                37                   10
## 8                      30.09                40                    9
## 9                      29.75                45                    6
## 10                     29.78                40                   10
## 11                     29.98                33                   10
## 12                     30.59                38                    8
## 13                     30.26                32                   10
## 14                     30.32                33                    9
## 15                     30.35                39                   10
##    Mean.Wind.SpeedMPH MeanDew.PointF Min.DewpointF Min.Humidity
## 1                  13             40            26           52
## 2                  13             39            37           89
## 3                  13             31            27           82
## 4                  11             27            25           64
## 5                  12             26            24           55
## 6                  10             27            25           53
## 7                   6             29            27           60
## 8                   4             36            30           73
## 9                  11             41            32           70
## 10                 14             30            26           57
## 11                 11             22            20           56
## 12                 15             27            17           51
## 13                 10             24            20           69
## 14                  9             27            25           69
## 15                  8             34            25           69
##    Min.Sea.Level.PressureIn Min.TemperatureF Min.VisibilityMiles
## 1                     30.01               39                  10
## 2                     29.43               38                   1
## 3                     29.44               32                   1
## 4                     29.81               31                   7
## 5                     29.78               32                  10
## 6                     29.78               33                  10
## 7                     29.91               32                  10
## 8                     29.92               35                   5
## 9                     29.69               41                   1
## 10                    29.71               36                  10
## 11                    29.86               29                  10
## 12                     30.4               33                   2
## 13                    30.17               27                   7
## 14                    30.28               30                   6
## 15                     30.3               33                   4
##    PrecipitationIn WindDirDegrees
## 1             0.01            268
## 2             0.28            357
## 3             0.02            230
## 4                T            286
## 5                T            298
## 6             0.00            306
## 7             0.00            324
## 8                T             79
## 9             0.43            311
## 10            0.01            281
## 11            0.00            305
## 12            0.10             62
## 13               T            350
## 14               T              2
## 15            0.05             24
```

Now our variables on the right places. Replace T letters in precipitation column with NA's. Get rid of the x letter in day number.

```
weather_tidy$PrecipitationIn <- gsub('T', NA, weather_tidy$PrecipitationIn, ignore.case = TRUE)
weather_tidy <- mutate(weather_tidy, day=extract_numeric(day))
```

```
## extract_numeric() is deprecated: please use readr::parse_number() instead
```

```
head(weather_tidy)
```

```
##   year month day CloudCover    Events Max.Dew.PointF Max.Gust.SpeedMPH
## 1 2014    12   1          6      Rain             46                29
## 2 2014    12  10          8      Rain             45                29
## 3 2014    12  11          8 Rain-Snow             37                28
## 4 2014    12  12          7      Snow             28                21
## 5 2014    12  13          5                       28                23
## 6 2014    12  14          4                       29                20
##   Max.Humidity Max.Sea.Level.PressureIn Max.TemperatureF
## 1           74                    30.45               64
## 2          100                    29.58               48
## 3           92                    29.81               39
## 4           85                    29.88               39
## 5           75                    29.86               42
## 6           82                    29.91               45
##   Max.VisibilityMiles Max.Wind.SpeedMPH Mean.Humidity
## 1                  10                22            63
## 2                  10                23            95
## 3                  10                21            87
## 4                  10                16            75
## 5                  10                17            65
## 6                  10                15            68
##   Mean.Sea.Level.PressureIn Mean.TemperatureF Mean.VisibilityMiles
## 1                     30.13              52                    10
## 2                      29.5              43                     3
## 3                     29.61              36                     7
## 4                     29.85              35                    10
## 5                     29.82              37                    10
## 6                     29.83              39                    10
##   Mean.Wind.SpeedMPH MeanDew.PointF Min.DewpointF Min.Humidity
## 1                 13             40            26           52
## 2                 13             39            37           89
## 3                 13             31            27           82
## 4                 11             27            25           64
## 5                 12             26            24           55
## 6                 10             27            25           53
##   Min.Sea.Level.PressureIn Min.TemperatureF Min.VisibilityMiles
## 1                    30.01               39                  10
## 2                    29.43               38                   1
## 3                    29.44               32                   1
## 4                    29.81               31                   7
## 5                    29.78               32                  10
## 6                    29.78               33                  10
##   PrecipitationIn WindDirDegrees
## 1            0.01            268
## 2            0.28            357
## 3            0.02            230
## 4            <NA>            286
## 5            <NA>            298
## 6            0.00            306
```

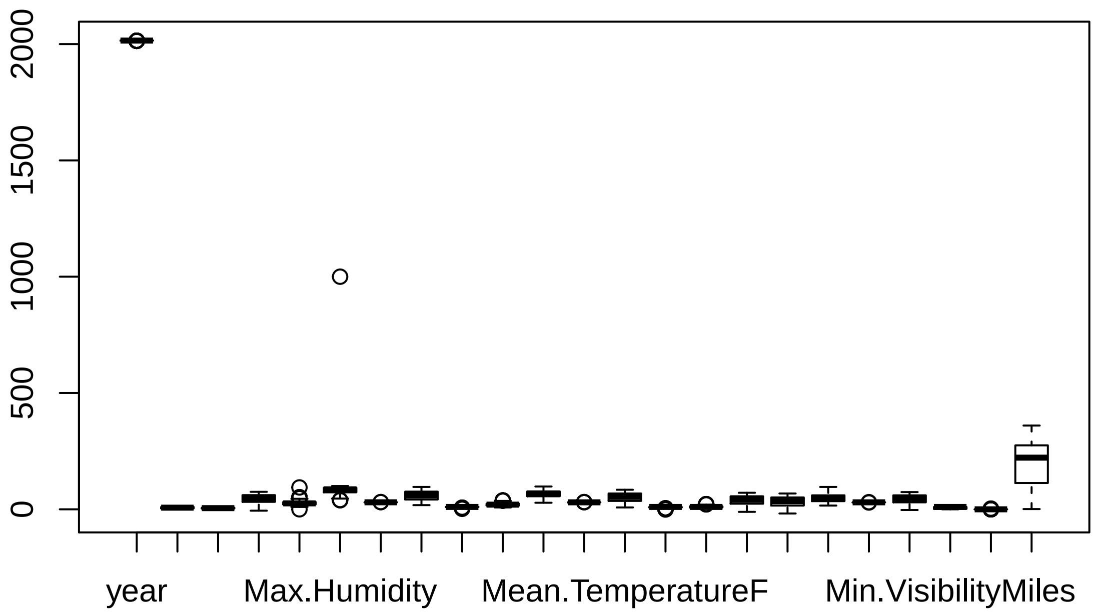
Now I will make our columns with numeric data real numeric.

```
weather_tidy[,c(1,2,4,6:25)] <- lapply(weather_tidy[,c(1,2,4,6:25)], as.numeric)
str(weather_tidy)
```

```
## 'data.frame':    403 obs. of  25 variables:
##  $ year                     : num  2014 2014 2014 2014 2014 ...
##  $ month                    : num  12 12 12 12 12 12 12 12 12 12 ...
##  $ day                      : num  1 10 11 12 13 14 15 16 17 18 ...
##  $ CloudCover               : num  6 8 8 7 5 4 2 8 8 7 ...
##  $ Events                   : chr  "Rain" "Rain" "Rain-Snow" "Snow" ...
##  $ Max.Dew.PointF           : num  46 45 37 28 28 29 33 42 46 34 ...
##  $ Max.Gust.SpeedMPH        : num  29 29 28 21 23 20 21 10 26 30 ...
##  $ Max.Humidity             : num  74 100 92 85 75 82 89 96 100 89 ...
##  $ Max.Sea.Level.PressureIn : num  30.4 29.6 29.8 29.9 29.9 ...
##  $ Max.TemperatureF         : num  64 48 39 39 42 45 42 44 49 44 ...
##  $ Max.VisibilityMiles      : num  10 10 10 10 10 10 10 10 10 10 ...
##  $ Max.Wind.SpeedMPH        : num  22 23 21 16 17 15 15 8 20 23 ...
##  $ Mean.Humidity            : num  63 95 87 75 65 68 75 85 85 73 ...
##  $ Mean.Sea.Level.PressureIn: num  30.1 29.5 29.6 29.9 29.8 ...
##  $ Mean.TemperatureF        : num  52 43 36 35 37 39 37 40 45 40 ...
##  $ Mean.VisibilityMiles     : num  10 3 7 10 10 10 10 9 6 10 ...
##  $ Mean.Wind.SpeedMPH       : num  13 13 13 11 12 10 6 4 11 14 ...
##  $ MeanDew.PointF           : num  40 39 31 27 26 27 29 36 41 30 ...
##  $ Min.DewpointF            : num  26 37 27 25 24 25 27 30 32 26 ...
##  $ Min.Humidity             : num  52 89 82 64 55 53 60 73 70 57 ...
##  $ Min.Sea.Level.PressureIn : num  30 29.4 29.4 29.8 29.8 ...
##  $ Min.TemperatureF          : num  39 38 32 31 32 33 32 35 41 36 ...
##  $ Min.VisibilityMiles       : num  10 1 1 7 10 10 10 5 1 10 ...
##  $ PrecipitationIn           : num  0.01 0.28 0.02 NA NA 0 0 NA 0.43 0.01 ...
##  $ WindDirDegrees            : num  268 357 230 286 298 306 324 79 311 281 ...
```

Let's plot our numeric values. As we can see, there is the outlier in Max. Humidity column. There is an additional zero in 138 observation. Let's get rid of it.

```
boxplot(weather_tidy[,c(1,2,4,6:25)])
```



```
weather_tidy$Max.Humidity
```

```
##   [1]   74  100   92   85   75   82   89   96  100   89   69   92   89   85
##  [15]   89  100  100  100   70   70   76   64  100   50   57   69   85  100
##  [29]   92   92  100   53   62   63  100   75   88   92   71   67   93   86
##  [43]   53   59   75   78   68  100   78   92   92   80   75  100   92   83
##  [57]  100   65   80   88   56   88   67   88   77   92   77  100   92   56
##  [71]   84   81   92  100   55   92  100   72   80   92   84   52   58   NA
##  [85]   73   NA   NA   75   89   60   92  100   92   92  100  100   59   69
##  [99]  100  100   63  100   46   39   92  100   96   54   49   72   85  100
## [113]   96   96   59   92   69   70   92   76   62   67   78   81   40  100
## [127]   89   59   66   80   39   67   89   89   76   61  100 1000   89   71
## [141]   64   60   71   63   76   77  100   89   NA  100   76   92   92   92
## [155]  100   92  100   93  100   59   54   64   90  100  100  100   92   93
## [169]   61   64   50   59   73   73   87   87  100   76   93  100   80   93
## [183]   89   47   86   97  100   78   78   79   84   73  100  100   73   72
## [197]   84  100   78  100  100  100   73   76   84   93  100   93   93   97
## [211]   NA   86   93  100   60   89   93  100  100   61   67   93   93  100
## [225]   78   84   93   93   78   93   79   84   68   81   93   93   93   97
## [239]   90   73   93   93   87   90   84   93   93   93   66   87  100  100
## [253]   78   78   87   93   84   90   93   73  100  100  100  100  100  100
## [267]  100   68   73   78  100   78   79   90   81   67   87   87   93   78
## [281]   93  100  100  100   90   67   78   84   87  100   87   90   72   93
## [295]   93   78   86   77   93   93  100   84  100   NA   93   93   90   73
## [309]   84   84   77   74   86   89  100   84   71   74   83   64   69   80
## [323]   65  100  100   60   80   89   70   83  100  100   80   53   71   77
## [337]   96   89   77   83   93   80   80   93  100   86   62   64   66   76
## [351]   85   89   83  100   70   89   85   50   76  100  100   93   79   82
## [365]   75   NA   83  100   93   57   65   70   96   NA   NA   NA   NA   NA
## [379]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## [393]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
```

```
weather_tidy[138, 8] <- 100
summary(weather_tidy$Max.Humidity)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   39.00   73.25   86.00   83.23   93.00  100.00      37
```