Hw4

MariaRomanova

First of all raw data is needed, and assume that file in current directory

```
weather = readRDS('./weather.rds')
```

Quick look at the data:

1 Measure column looks unnormal (X* variables must be days)

2 All X* columns are of class character

3 There are not complete cases with NA data

```
head(weather, 10)
```

```
##     X year month                  measure    X1    X2   X3    X4    X5
## 1   1 2014    12         Max.TemperatureF    64    42   51    43    42
## 2   2 2014    12        Mean.TemperatureF    52    38   44    37    34
## 3   3 2014    12         Min.TemperatureF    39    33   37    30    26
## 4   4 2014    12           Max.Dew.PointF    46    40   49    24    37
## 5   5 2014    12           MeanDew.PointF    40    27   42    21    25
## 6   6 2014    12            Min.DewpointF    26    17   24    13    12
## 7   7 2014    12              Max.Humidity    74    92  100    69    85
## 8   8 2014    12             Mean.Humidity    63    72   79    54    66
## 9   9 2014    12              Min.Humidity    52    51   57    39    47
## 10 10 2014    12 Max.Sea.Level.PressureIn 30.45 30.71 30.4 30.56 30.68
##       X6    X7    X8    X9   X10   X11   X12   X13   X14   X15   X16   X17
## 1     45    38    29    49    48    39    39    42    45    42    44    49
## 2     42    30    24    39    43    36    35    37    39    37    40    45
## 3     38    21    18    29    38    32    31    32    33    32    35    41
## 4     45    36    28    49    45    37    28    28    29    33    42    46
## 5     40    20    16    41    39    31    27    26    27    29    36    41
## 6     36    -3     3    28    37    27    25    24    25    27    30    32
## 7    100    92    92   100   100    92    85    75    82    89    96   100
## 8     93    61    70    93    95    87    75    65    68    75    85    85
## 9     85    29    47    86    89    82    64    55    53    60    73    70
## 10 30.42 30.69 30.77 30.51 29.58 29.81 29.88 29.86 29.91 30.15 30.17 29.91
##      X18   X19   X20   X21   X22   X23   X24   X25   X26   X27   X28   X29
## 1     44    37    36    36    44    47    46    59    50    52    52    41
## 2     40    33    32    33    39    45    44    52    44    45    46    36
## 3     36    29    27    30    33    42    41    44    37    38    40    30
## 4     34    25    30    30    39    45    46    58    31    34    42    26
## 5     30    22    24    27    34    42    44    43    29    31    35    20
## 6     26    20    20    25    25    37    41    29    28    29    27    10
## 7     89    69    89    85    89   100   100   100    70    70    76    64
## 8     73    63    79    77    79    91    98    75    60    60    65    51
## 9     57    56    69    69    69    82    96    49    49    50    53    37
## 10 29.87 30.15 30.31 30.37 30.4 30.31 30.13 29.96 30.16 30.22 29.99 30.22
```

```
##      X30   X31
## 1     30    30
## 2     26    25
## 3     22    20
## 4     10     8
## 5      4     5
## 6     -6     1
## 7     50    57
## 8     38    44
## 9     26    31
## 10 30.36 30.32
```

```
sum(complete.cases(weather))
```

```
## [1] 151
```

```
summary(weather)
```

```
##        X                year          month          measure
##  Min.   :  1.00   Min.   :2014   Min.   : 1.000   Length:286
##  1st Qu.: 72.25   1st Qu.:2015   1st Qu.: 4.000   Class :character
##  Median :143.50   Median :2015   Median : 7.000   Mode  :character
##  Mean   :143.50   Mean   :2015   Mean   : 6.923
##  3rd Qu.:214.75   3rd Qu.:2015   3rd Qu.:10.000
##  Max.   :286.00   Max.   :2015   Max.   :12.000
##       X1                 X2                 X3
##  Length:286         Length:286         Length:286
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##       X4                 X5                 X6
##  Length:286         Length:286         Length:286
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##       X7                 X8                 X9
##  Length:286         Length:286         Length:286
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##      X10                X11                X12
##  Length:286         Length:286         Length:286
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##      X13                X14                X15
##  Length:286         Length:286         Length:286
```

```
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##      X16                X17                X18
##  Length:286         Length:286         Length:286
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##      X19                X20                X21
##  Length:286         Length:286         Length:286
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##      X22                X23                X24
##  Length:286         Length:286         Length:286
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##      X25                X26                X27
##  Length:286         Length:286         Length:286
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##      X28                X29                X30
##  Length:286         Length:286         Length:286
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##      X31
##  Length:286
##  Class :character
##  Mode  :character
##
##
##
```

```
sum(duplicated(weather))
```

```
## [1] 0
```

Library will be tidyr and stringr

```
library(stringr)
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
```

It seems wrong that months and years are rows and days are colunms, and logicaly it should be columns: days, months, years, and then all measures; rows: measured data

```
weather <- gather(weather, day, data, colnames(weather[5:35]))
```

```
weather <- weather[,-1]
weather <- spread(weather, key = measure, value = data)
```

```
weather <- weather %>%
  mutate(day = extract_numeric(day))
```

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

```
weather <- arrange(weather, year, month,day)
```

Get rid of 'T' in PrecipitationIn collumn and replace with NA

```
weather$PrecipitationIn <- gsub('T', NA, weather$PrecipitationIn)
```

Next changing of class to numeric

```
weather[,-5] <- lapply(weather[,-5], as.numeric)
weather$Events <- as.factor(weather$Events)
```

Piece of code can be used for delete full empty rows we can also use na.omit but not full NA rows will be deleted as well

```
vec = c()
f=1
for(i in 1:length(weather$month)){
  if (sum(is.na(weather[i,4:25])) == 22){
    vec[f] <- i
```

```
    f <- f+1
  }
}
weather <- weather[-c(vec),]
```

And one more code which will replace blanck rows in column, because if later on this raws will needed to be omited that can be usefull, but it is optional

```
for (i in 1:length(weather$Events)){
  if (!grepl('^.', weather$Events[i])){
    weather$Events[i] <- NA
  }
}
```