

Data Science - Practice 5

모든 문제에 대하여 코드만 작성하지 말고 데이터를 해석한 결과를 함께 작성하시오.

For all questions include explanation of the process and the result in your report as well as R code.

loading data into R

아래와 같이 실습을 위한 데이터를 R에 loading 하시오.

Load the weather dataset for this practice as in the code below.

```
weather_df <- readRDS('weather.rds')
```

Data description

weather_df는 미국 보스턴에서 2014년 12월부터 12개월간 측정된 날씨 정보를 담고 있는 data frame이다. 다음 질문에 답하라.

“weather_df” is a data frame that contains Historical weather information from Boston, USA collected for 12 months beginning Dec 2014. Answer the following questions.

```
## Rows: 286
## Columns: 35
## $ X      <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, ...
## $ year   <int> 2014, 2014, 2014, 2014, 2014, 2014, 2014, 2014, 2014, 2014, 2014, ...
## $ month  <int> 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, ...
## $ measure <chr> "Max.TemperatureF", "Mean.TemperatureF", "Min.TemperatureF"...
## $ X1     <chr> "64", "52", "39", "46", "40", "26", "74", "63", "52", "30.4...
## $ X2     <chr> "42", "38", "33", "40", "27", "17", "92", "72", "51", "30.7...
## $ X3     <chr> "51", "44", "37", "49", "42", "24", "100", "79", "57", "30....
## $ X4     <chr> "43", "37", "30", "24", "21", "13", "69", "54", "39", "30.5...
## $ X5     <chr> "42", "34", "26", "37", "25", "12", "85", "66", "47", "30.6...
## $ X6     <chr> "45", "42", "38", "45", "40", "36", "100", "93", "85", "30....
## $ X7     <chr> "38", "30", "21", "36", "20", "-3", "92", "61", "29", "30.6...
## $ X8     <chr> "29", "24", "18", "28", "16", "3", "92", "70", "47", "30.77...
## $ X9     <chr> "49", "39", "29", "49", "41", "28", "100", "93", "86", "30....
## $ X10    <chr> "48", "43", "38", "45", "39", "37", "100", "95", "89", "29....
## $ X11    <chr> "39", "36", "32", "37", "31", "27", "92", "87", "82", "29.8...
## $ X12    <chr> "39", "35", "31", "28", "27", "25", "85", "75", "64", "29.8...
## $ X13    <chr> "42", "37", "32", "28", "26", "24", "75", "65", "55", "29.8...
## $ X14    <chr> "45", "39", "33", "29", "27", "25", "82", "68", "53", "29.9...
## $ X15    <chr> "42", "37", "32", "33", "29", "27", "89", "75", "60", "30.1...
## $ X16    <chr> "44", "40", "35", "42", "36", "30", "96", "85", "73", "30.1...
## $ X17    <chr> "49", "45", "41", "46", "41", "32", "100", "85", "70", "29....
## $ X18    <chr> "44", "40", "36", "34", "30", "26", "89", "73", "57", "29.8...
## $ X19    <chr> "37", "33", "29", "25", "22", "20", "69", "63", "56", "30.1...
## $ X20    <chr> "36", "32", "27", "30", "24", "20", "89", "79", "69", "30.3...
## $ X21    <chr> "36", "33", "30", "30", "27", "25", "85", "77", "69", "30.3...
```

```

## $ X22      <chr> "44", "39", "33", "39", "34", "25", "89", "79", "69", "30.4...
## $ X23      <chr> "47", "45", "42", "45", "42", "37", "100", "91", "82", "30....
## $ X24      <chr> "46", "44", "41", "46", "44", "41", "100", "98", "96", "30....
## $ X25      <chr> "59", "52", "44", "58", "43", "29", "100", "75", "49", "29....
## $ X26      <chr> "50", "44", "37", "31", "29", "28", "70", "60", "49", "30.1...
## $ X27      <chr> "52", "45", "38", "34", "31", "29", "70", "60", "50", "30.2...
## $ X28      <chr> "52", "46", "40", "42", "35", "27", "76", "65", "53", "29.9...
## $ X29      <chr> "41", "36", "30", "26", "20", "10", "64", "51", "37", "30.2...
## $ X30      <chr> "30", "26", "22", "10", "4", "-6", "50", "38", "26", "30.36...
## $ X31      <chr> "30", "25", "20", "8", "5", "1", "57", "44", "31", "30.32",...

```

```

##      X year month      measure      X1      X2      X3      X4      X5      X6
## 1    1 2014     12      Max.TemperatureF      64      42      51      43      42      45
## 2    2 2014     12      Mean.TemperatureF      52      38      44      37      34      42
## 3    3 2014     12      Min.TemperatureF      39      33      37      30      26      38
## 4    4 2014     12      Max.Dew.PointF      46      40      49      24      37      45
## 5    5 2014     12      MeanDew.PointF      40      27      42      21      25      40
## 6    6 2014     12      Min.DewpointF      26      17      24      13      12      36
## 7    7 2014     12      Max.Humidity      74      92     100      69      85     100
## 8    8 2014     12      Mean.Humidity      63      72      79      54      66      93
## 9    9 2014     12      Min.Humidity      52      51      57      39      47      85
## 10   10 2014     12 Max.Sea.Level.PressureIn 30.45 30.71 30.4 30.56 30.68 30.42
##      X7      X8      X9      X10     X11     X12     X13     X14     X15     X16     X17     X18
## 1    38      29      49      48      39      39      42      45      42      44      49      44
## 2    30      24      39      43      36      35      37      39      37      40      45      40
## 3    21      18      29      38      32      31      32      33      32      35      41      36
## 4    36      28      49      45      37      28      28      29      33      42      46      34
## 5    20      16      41      39      31      27      26      27      29      36      41      30
## 6    -3       3      28      37      27      25      24      25      27      30      32      26
## 7    92      92     100     100      92      85      75      82      89      96     100      89
## 8    61      70      93      95      87      75      65      68      75      85      85      73
## 9    29      47      86      89      82      64      55      53      60      73      70      57
## 10   30.69 30.77 30.51 29.58 29.81 29.88 29.86 29.91 30.15 30.17 29.91 29.87
##      X19     X20     X21     X22     X23     X24     X25     X26     X27     X28     X29     X30     X31
## 1    37      36      36      44      47      46      59      50      52      52      41      30      30
## 2    33      32      33      39      45      44      52      44      45      46      36      26      25
## 3    29      27      30      33      42      41      44      37      38      40      30      22      20
## 4    25      30      30      39      45      46      58      31      34      42      26      10      8
## 5    22      24      27      34      42      44      43      29      31      35      20      4      5
## 6    20      20      25      25      37      41      29      28      29      27      10      -6      1
## 7    69      89      85      89     100     100     100      70      70      76      64      50      57
## 8    63      79      77      79      91      98      75      60      60      65      51      38      44
## 9    56      69      69      69      82      96      49      49      50      53      37      26      31
## 10   30.15 30.31 30.37 30.4 30.31 30.13 29.96 30.16 30.22 29.99 30.22 30.36 30.32

```

1. 첫번째 column은 인덱스를 표현한 것이지만 불필요하다.
2. year을 month로 합쳐도 데이터를 classify하기 충분하다.
3. missing value가 많다.
4. 행들 이름이 의미없는 값이 들어있다.

< Question 1 >

*weather_{df}*가 tidy가 아닌 이유를 설명하시오.

Explain why the dataset *weather_{df}* is not tidy.

< Question 2 >

불필요한 column을 제거하시오.

Remove unnecessary columns from data frame.

< Question 3 >

Dataset을 tidy한 형태로 변환하시오.

Transform the dataset into tidy one.

```
head(weather_tidy, 10)
```

```
##      year month dayOfMonth 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
##      Max.Humidity Max.Sea.Level.PressureIn Max.TemperatureF Max.VisibilityMiles
## 1              74              30.45              64              10
## 2             100              29.58              48              10
## 3              92              29.81              39              10
## 4              85              29.88              39              10
## 5              75              29.86              42              10
## 6              82              29.91              45              10
## 7              89              30.15              42              10
## 8              96              30.17              44              10
## 9             100              29.91              49              10
## 10             89              29.87              44              10
##      Max.Wind.SpeedMPH Mean.Humidity Mean.Sea.Level.PressureIn Mean.TemperatureF
## 1              22              63              30.13              52
## 2              23              95              29.5              43
## 3              21              87              29.61              36
## 4              16              75              29.85              35
## 5              17              65              29.82              37
## 6              15              68              29.83              39
## 7              15              75              30.05              37
## 8               8              85              30.09              40
## 9              20              85              29.75              45
## 10             23              73              29.78              40
##      Mean.VisibilityMiles Mean.Wind.SpeedMPH MeanDew.PointF Min.DewpointF
## 1              10              13              40              26
```

```
## 2          3          13          39          37
## 3          7          13          31          27
## 4         10          11          27          25
## 5         10          12          26          24
## 6         10          10          27          25
## 7         10          6          29          27
## 8          9          4          36          30
## 9          6          11          41          32
## 10        10          14          30          26
##   Min.Humidity Min.Sea.Level.PressureIn Min.TemperatureF Min.VisibilityMiles
## 1          52          30.01          39          10
## 2          89          29.43          38           1
## 3          82          29.44          32           1
## 4          64          29.81          31           7
## 5          55          29.78          32          10
## 6          53          29.78          33          10
## 7          60          29.91          32          10
## 8          73          29.92          35           5
## 9          70          29.69          41           1
## 10         57          29.71          36          10
##   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
```

```
str(weather_tidy)
```

```
## 'data.frame':  403 obs. of  25 variables:
## $ year          : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
## $ month         : int  12 12 12 12 12 12 12 12 12 12 ...
## $ dayOfMonth    : chr  "X1" "X10" "X11" "X12" ...
## $ CloudCover    : chr  "6" "8" "8" "7" ...
## $ Events        : chr  "Rain" "Rain" "Rain-Snow" "Snow" ...
## $ Max.Dew.PointF : chr  "46" "45" "37" "28" ...
## $ Max.Gust.SpeedMPH : chr  "29" "29" "28" "21" ...
## $ Max.Humidity   : chr  "74" "100" "92" "85" ...
## $ Max.Sea.Level.PressureIn : chr  "30.45" "29.58" "29.81" "29.88" ...
## $ Max.TemperatureF : chr  "64" "48" "39" "39" ...
## $ Max.VisibilityMiles : chr  "10" "10" "10" "10" ...
## $ Max.Wind.SpeedMPH : chr  "22" "23" "21" "16" ...
## $ Mean.Humidity   : chr  "63" "95" "87" "75" ...
## $ Mean.Sea.Level.PressureIn: chr  "30.13" "29.5" "29.61" "29.85" ...
## $ Mean.TemperatureF : chr  "52" "43" "36" "35" ...
## $ Mean.VisibilityMiles : chr  "10" "3" "7" "10" ...
## $ Mean.Wind.SpeedMPH : chr  "13" "13" "13" "11" ...
## $ MeanDew.PointF  : chr  "40" "39" "31" "27" ...
## $ Min.DewpointF   : chr  "26" "37" "27" "25" ...
## $ Min.Humidity     : chr  "52" "89" "82" "64" ...
```

```
## $ Min.Sea.Level.PressureIn : chr "30.01" "29.43" "29.44" "29.81" ...
## $ Min.TemperatureF         : chr "39" "38" "32" "31" ...
## $ Min.VisibilityMiles      : chr "10" "1" "1" "7" ...
## $ PrecipitationIn          : chr "0.01" "0.28" "0.02" "T" ...
## $ WindDirDegrees           : chr "268" "357" "230" "286" ...
```

< Question 4 >

dayOfMonth 변수를 수치형 변수로 적절하게 변환하여라.

Convert the variable 'dayofMonth' to numeric type.

```
head(weather_tidy, 10)
```

```
##   year month dayOfMonth 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
## 7  2014    12        15         2           33           21
## 8  2014    12        16         8     Rain           42           10
## 9  2014    12        17         8     Rain           46           26
## 10 2014    12        18         7     Rain           34           30
##   Max.Humidity Max.Sea.Level.PressureIn Max.TemperatureF Max.VisibilityMiles
## 1             74             30.45             64             10
## 2             100            29.58             48             10
## 3             92             29.81             39             10
## 4             85             29.88             39             10
## 5             75             29.86             42             10
## 6             82             29.91             45             10
## 7             89             30.15             42             10
## 8             96             30.17             44             10
## 9             100            29.91             49             10
## 10            89             29.87             44             10
##   Max.Wind.SpeedMPH Mean.Humidity Mean.Sea.Level.PressureIn Mean.TemperatureF
## 1                  22             63             30.13             52
## 2                  23             95             29.5             43
## 3                  21             87             29.61             36
## 4                  16             75             29.85             35
## 5                  17             65             29.82             37
## 6                  15             68             29.83             39
## 7                  15             75             30.05             37
## 8                   8             85             30.09             40
## 9                  20             85             29.75             45
## 10                 23             73             29.78             40
##   Mean.VisibilityMiles Mean.Wind.SpeedMPH MeanDew.PointF Min.DewpointF
## 1                   10                  13             40             26
## 2                   3                  13             39             37
## 3                   7                  13             31             27
## 4                  10                  11             27             25
## 5                  10                  12             26             24
## 6                  10                  10             27             25
```

```

## 7          10          6          29          27
## 8           9          4          36          30
## 9           6         11          41          32
## 10         10         14          30          26
##   Min.Humidity Min.Sea.Level.PressureIn Min.TemperatureF Min.VisibilityMiles
## 1           52           30.01           39           10
## 2           89           29.43           38            1
## 3           82           29.44           32            1
## 4           64           29.81           31            7
## 5           55           29.78           32           10
## 6           53           29.78           33           10
## 7           60           29.91           32           10
## 8           73           29.92           35            5
## 9           70           29.69           41            1
## 10          57           29.71           36           10
##   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

```

```
str(weather_tidy)
```

```

## 'data.frame':   403 obs. of  25 variables:
## $ year          : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
## $ month         : int  12 12 12 12 12 12 12 12 12 12 ...
## $ dayOfMonth    : num  1 10 11 12 13 14 15 16 17 18 ...
## $ CloudCover    : chr  "6" "8" "8" "7" ...
## $ Events        : chr  "Rain" "Rain" "Rain-Snow" "Snow" ...
## $ Max.Dew.PointF : chr  "46" "45" "37" "28" ...
## $ Max.Gust.SpeedMPH : chr  "29" "29" "28" "21" ...
## $ Max.Humidity   : chr  "74" "100" "92" "85" ...
## $ Max.Sea.Level.PressureIn : chr  "30.45" "29.58" "29.81" "29.88" ...
## $ Max.TemperatureF : chr  "64" "48" "39" "39" ...
## $ Max.VisibilityMiles : chr  "10" "10" "10" "10" ...
## $ Max.Wind.SpeedMPH : chr  "22" "23" "21" "16" ...
## $ Mean.Humidity   : chr  "63" "95" "87" "75" ...
## $ Mean.Sea.Level.PressureIn: chr  "30.13" "29.5" "29.61" "29.85" ...
## $ Mean.TemperatureF : chr  "52" "43" "36" "35" ...
## $ Mean.VisibilityMiles : chr  "10" "3" "7" "10" ...
## $ Mean.Wind.SpeedMPH : chr  "13" "13" "13" "11" ...
## $ MeanDew.PointF : chr  "40" "39" "31" "27" ...
## $ Min.DewpointF   : chr  "26" "37" "27" "25" ...
## $ Min.Humidity    : chr  "52" "89" "82" "64" ...
## $ Min.Sea.Level.PressureIn : chr  "30.01" "29.43" "29.44" "29.81" ...
## $ Min.TemperatureF : chr  "39" "38" "32" "31" ...
## $ Min.VisibilityMiles : chr  "10" "1" "1" "7" ...
## $ PrecipitationIn : chr  "0.01" "0.28" "0.02" "T" ...
## $ WindDirDegrees  : chr  "268" "357" "230" "286" ...

```

< Question 5 >

데이터에 year month dayOfMonth 세 column이 있는데 이를 하나로 합쳐서 date column을 추가하시오. date column은 Date type 이어야합니다.

그리고 year month dayOfMonth 세 column은 제거하시오.

There are year, month, and dayOfMonth column in the dataset. Combine these three columns into a new column named “date” which is in type of **Date**.

Then remove the columns of year, month, and dayOfMonth from the data frame

```
##      date CloudCover    Events Max.Dew.PointF Max.Gust.SpeedMPH
## 1  2014-12-01         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
## 7  2014-12-15         2              33                21
## 8  2014-12-16         8     Rain              42                10
## 9  2014-12-17         8     Rain              46                26
## 10 2014-12-18         7     Rain              34                30
##      Max.Humidity Max.Sea.Level.PressureIn Max.TemperatureF Max.VisibilityMiles
## 1              74                      30.45              64              10
## 2             100                      29.58              48              10
## 3              92                      29.81              39              10
## 4              85                      29.88              39              10
## 5              75                      29.86              42              10
## 6              82                      29.91              45              10
## 7              89                      30.15              42              10
## 8              96                      30.17              44              10
## 9             100                      29.91              49              10
## 10             89                      29.87              44              10
##      Max.Wind.SpeedMPH Mean.Humidity Mean.Sea.Level.PressureIn Mean.TemperatureF
## 1                  22              63                      30.13              52
## 2                  23              95                      29.5              43
## 3                  21              87                      29.61              36
## 4                  16              75                      29.85              35
## 5                  17              65                      29.82              37
## 6                  15              68                      29.83              39
## 7                  15              75                      30.05              37
## 8                   8              85                      30.09              40
## 9                  20              85                      29.75              45
## 10                 23              73                      29.78              40
##      Mean.VisibilityMiles Mean.Wind.SpeedMPH MeanDew.PointF Min.DewpointF
## 1                   10              13              40              26
## 2                   3              13              39              37
## 3                   7              13              31              27
## 4                  10              11              27              25
## 5                  10              12              26              24
## 6                  10              10              27              25
## 7                  10               6              29              27
## 8                   9               4              36              30
## 9                   6              11              41              32
## 10                 10              14              30              26
```

```
##      Min.Humidity Min.Sea.Level.PressureIn Min.TemperatureF Min.VisibilityMiles
## 1           52           30.01           39           10
## 2           89           29.43           38            1
## 3           82           29.44           32            1
## 4           64           29.81           31            7
## 5           55           29.78           32           10
## 6           53           29.78           33           10
## 7           60           29.91           32           10
## 8           73           29.92           35            5
## 9           70           29.69           41            1
## 10          57           29.71           36           10
##      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
```

< Question 6 >

PrecipitationIn(강수량) 변수를 보면 “T”라는 값이 있는데 이는 Trace 비가 아주 미량왔다는 의미이다. 해당 변수를 숫자형으로 변환할 수 있도록, “T”를 숫자 0으로 변환하시오.

There are some values of **T** in the variable “PrecipitationIn”, meaning a trace amount (i.e. too small to be accurately measured) of precipitation.

To have this variable as numeric one, change all “T” to zero.

```
## [1] "0.01" "0.28" "0.02" "T"    "T"    "0.00" "0.00" "T"    "0.43" "0.01"
## [11] "0.00" "0.10" "T"    "T"    "0.05" "0.25" "0.56" "0.14" "0.00" "0.00"

## [1] "0.01" "0.28" "0.02" "0"    "0"    "0.00" "0.00" "0"    "0.43" "0.01"
## [11] "0.00" "0.10" "0"    "0"    "0.05" "0.25" "0.56" "0.14" "0.00" "0.00"
```

< Question 7 >

각 변수의 data type을 적절한 것으로 변환하시오.

Convert the data type of variables into proper ones.

```
## Rows: 403
## Columns: 23
## $ date          <date> 2014-12-01, 2014-12-10, 2014-12-11, 2014...
## $ CloudCover    <chr> "6", "8", "8", "7", "5", "4", "2", "8", "...
## $ Events        <chr> "Rain", "Rain", "Rain-Snow", "Snow", "", ...
## $ Max.Dew.PointF <chr> "46", "45", "37", "28", "28", "29", "33",...
## $ Max.Gust.SpeedMPH <chr> "29", "29", "28", "21", "23", "20", "21",...
## $ Max.Humidity   <chr> "74", "100", "92", "85", "75", "82", "89"...
## $ Max.Sea.Level.PressureIn <chr> "30.45", "29.58", "29.81", "29.88", "29.8...
## $ Max.TemperatureF <chr> "64", "48", "39", "39", "42", "45", "42",...
```



```

## $ Max.VisibilityMiles      <chr> "10", "10", "10", "10", "10", "10", "10",...
## $ Max.Wind.SpeedMPH       <chr> "22", "23", "21", "16", "17", "15", "15",...
## $ Mean.Humidity           <chr> "63", "95", "87", "75", "65", "68", "75",...
## $ Mean.Sea.Level.PressureIn <chr> "30.13", "29.5", "29.61", "29.85", "29.82...
## $ Mean.TemperatureF       <chr> "52", "43", "36", "35", "37", "39", "37",...
## $ Mean.VisibilityMiles     <chr> "10", "3", "7", "10", "10", "10", "10", "..."
## $ Mean.Wind.SpeedMPH      <chr> "13", "13", "13", "11", "12", "10", "6", "..."
## $ MeanDew.PointF          <chr> "40", "39", "31", "27", "26", "27", "29",...
## $ Min.DewpointF           <chr> "26", "37", "27", "25", "24", "25", "27",...
## $ Min.Humidity            <chr> "52", "89", "82", "64", "55", "53", "60",...
## $ Min.Sea.Level.PressureIn <chr> "30.01", "29.43", "29.44", "29.81", "29.7...
## $ Min.TemperatureF        <chr> "39", "38", "32", "31", "32", "33", "32",...
## $ Min.VisibilityMiles     <chr> "10", "1", "1", "7", "10", "10", "10", "5...
## $ PrecipitationIn         <chr> "0.01", "0.28", "0.02", "0", "0", "0.00",...
## $ WindDirDegrees          <chr> "268", "357", "230", "286", "298", "306",...

## Rows: 403
## Columns: 23
## $ date                    <date> 2014-12-01, 2014-12-10, 2014-12-11, 2014...
## $ CloudCover              <dbl> 6, 8, 8, 7, 5, 4, 2, 8, 8, 7, 4, 7, 6, 8,...
## $ Events                  <fct> Rain, Rain, Rain-Snow, Snow, , , , Rain, ...
## $ Max.Dew.PointF          <dbl> 46, 45, 37, 28, 28, 29, 33, 42, 46, 34, 2...
## $ Max.Gust.SpeedMPH       <dbl> 29, 29, 28, 21, 23, 20, 21, 10, 26, 30, 2...
## $ Max.Humidity            <dbl> 74, 100, 92, 85, 75, 82, 89, 96, 100, 89,...
## $ Max.Sea.Level.PressureIn <dbl> 30.45, 29.58, 29.81, 29.88, 29.86, 29.91,...
## $ Max.TemperatureF        <dbl> 64, 48, 39, 39, 42, 45, 42, 44, 49, 44, 3...
## $ Max.VisibilityMiles     <dbl> 10, 10, 10, 10, 10, 10, 10, 10, 10, 1...
## $ Max.Wind.SpeedMPH       <dbl> 22, 23, 21, 16, 17, 15, 15, 8, 20, 23, 17...
## $ Mean.Humidity           <dbl> 63, 95, 87, 75, 65, 68, 75, 85, 85, 73, 6...
## $ Mean.Sea.Level.PressureIn <dbl> 30.13, 29.50, 29.61, 29.85, 29.82, 29.83,...
## $ Mean.TemperatureF       <dbl> 52, 43, 36, 35, 37, 39, 37, 40, 45, 40, 3...
## $ Mean.VisibilityMiles     <dbl> 10, 3, 7, 10, 10, 10, 10, 9, 6, 10, 10, 8...
## $ Mean.Wind.SpeedMPH      <dbl> 13, 13, 13, 11, 12, 10, 6, 4, 11, 14, 11,...
## $ MeanDew.PointF          <dbl> 40, 39, 31, 27, 26, 27, 29, 36, 41, 30, 2...
## $ Min.DewpointF           <dbl> 26, 37, 27, 25, 24, 25, 27, 30, 32, 26, 2...
## $ Min.Humidity            <dbl> 52, 89, 82, 64, 55, 53, 60, 73, 70, 57, 5...
## $ Min.Sea.Level.PressureIn <dbl> 30.01, 29.43, 29.44, 29.81, 29.78, 29.78,...
## $ Min.TemperatureF        <dbl> 39, 38, 32, 31, 32, 33, 32, 35, 41, 36, 2...
## $ Min.VisibilityMiles     <dbl> 10, 1, 1, 7, 10, 10, 10, 5, 1, 10, 10, 2...
## $ PrecipitationIn         <dbl> 0.01, 0.28, 0.02, 0.00, 0.00, 0.00, 0.00,...
## $ WindDirDegrees          <dbl> 268, 357, 230, 286, 298, 306, 324, 79, 31...

```

< Question 8 >

데이터셋에 missing values가 있나요?

몇 개나 있나요?

각 변수 별로 몇 개씩 있나요?

[Missing Values] Does the dataset contains any missing values?

How may are they in the dataset?

How many missing values are in each variable?

< Question 9 >

Max.Humidity(최대 습도) 변수를 보시오. outlier가 있나요? outlier 값이 실수로 0이 하나 더 붙어 나온 값이라고 합시다. 해당 outlier를 적절한 값으로 고치시오.

[Outliers] Look at the variable Max.Humidity.

Is there any outlier (extreme value) in the variable?

Assuming that one more "0" was added accidently to the outlier, correct the outlier into proper value.

< Question 10 >

Mean.VisibilityMiles(평균시야거리) 변수를 보시오. outlier가 있나요? outlier를 적절한 값으로 고치시오.

[Outliers] Look at the variable Mean.VisibilityMiles.

Is there any outlier (extreme value) in the variable?

Correct the outlier into proper value.

< Question 11 >

Event변수를 보면 공백문자 " "가 포함되어있습니다. 비나 안개 같은 특별한 event가 없는 날이라는 표시인데, 더욱 명백하게 표현하는 것이 좋습니다. 공백문자를 "None"으로 바꾸시오.

The Events variable contains an empty space (" ") for the days of no significant weather events such as rain, fog, a thunderstorm, etc.

However, if it's the first time you're seeing these data, it may not be obvious that this is the case, so it's best for us to be explicit and replace the empty spaces with something more intuitive.

Convert the empty space into "None".

```
## [1] Rain      Rain      Rain-Snow Snow
## [8] Rain      Rain      Rain              Rain-Snow Snow      Snow
## [15] Rain      Rain      Fog-Rain  Rain
## 12 Levels: Fog Fog-Rain Fog-Rain-Hail-Thunderstorm ... Thunderstorm

## [1] Rain      Rain      Rain-Snow Snow      None      None      None
## [8] Rain      Rain      Rain      None      Rain-Snow Snow      Snow
## [15] Rain      Rain      Fog-Rain  Rain      None      None
## 12 Levels: Fog Fog-Rain Fog-Rain-Hail-Thunderstorm ... Thunderstorm
```

< Question 12 >

data frame의 column name은 모두 소문자로 하는 것이 좋습니다. 나중에 대문자인지 소문자인지 기억하지 않아도 되기 때문입니다. data frame에서 column name을 모두 소문자로 바꾸시오.

For the column names of data frame, we prefer to have them in all lower-case letters. So we do not have to remember which letters are uppercase or lowercase.

Convert all column names of data frame into lower case letters

```
## [1] "date" "CloudCover"
## [3] "Events" "Max.Dew.PointF"
## [5] "Max.Gust.SpeedMPH" "Max.Humidity"
## [7] "Max.Sea.Level.PressureIn" "Max.TemperatureF"
## [9] "Max.VisibilityMiles" "Max.Wind.SpeedMPH"
## [11] "Mean.Humidity" "Mean.Sea.Level.PressureIn"
## [13] "Mean.TemperatureF" "Mean.VisibilityMiles"
## [15] "Mean.Wind.SpeedMPH" "MeanDew.PointF"
## [17] "Min.DewpointF" "Min.Humidity"
## [19] "Min.Sea.Level.PressureIn" "Min.TemperatureF"
## [21] "Min.VisibilityMiles" "PrecipitationIn"
## [23] "WindDirDegrees"

## [1] "date" "cloudcover"
## [3] "events" "max.dew.pointf"
## [5] "max.gust.speedmph" "max.humidity"
## [7] "max.sea.level.pressurein" "max.temperaturef"
## [9] "max.visibilitymiles" "max.wind.speedmph"
## [11] "mean.humidity" "mean.sea.level.pressurein"
## [13] "mean.temperaturef" "mean.visibilitymiles"
## [15] "mean.wind.speedmph" "meandew.pointf"
## [17] "min.dewpointf" "min.humidity"
## [19] "min.sea.level.pressurein" "min.temperaturef"
## [21] "min.visibilitymiles" "precipitationin"
## [23] "winddirdegrees"
```

< Question 13 >

결과 데이터 프레임을 RData 파일에 저장하여 보고서와 함께 LMS에 제출하십시오.

Include the result data frame into a RData file with **save** command
and submit the file with your report to LMS.