Datacamp_Cleaning Data in R_Putting it all together

dizhen

2019/4/5

Get a feel for the data

```
# Data
weather <- readRDS("data/weather.rds")

# Verify that weather is a data.frame
class(weather)
```

```
## [1] "data.frame"
```

```
# Check the dimensions
dim(weather)
```

```
## [1] 286  35
```

```
# View the column names
names(weather)
```

```
##  [1] "X"       "year"    "month"   "measure" "X1"      "X2"      "X3"
##  [8] "X4"      "X5"      "X6"      "X7"      "X8"      "X9"      "X10"
## [15] "X11"     "X12"     "X13"     "X14"     "X15"     "X16"     "X17"
## [22] "X18"     "X19"     "X20"     "X21"     "X22"     "X23"     "X24"
## [29] "X25"     "X26"     "X27"     "X28"     "X29"     "X30"     "X31"
```

Summarize the data

```
# View the structure of the data
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
# Load dplyr package
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
```

```r
# Look at the structure using dplyr's glimpse()
glimpse(weather)
```

```
## Observations: 286
## Variables: 35
## $ X       <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,...
## $ year    <int> 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...
## $ measure <chr> "Max.TemperatureF", "Mean.TemperatureF", "Min.Temperat...
## $ X1      <chr> "64", "52", "39", "46", "40", "26", "74", "63", "52", ...
## $ X2      <chr> "42", "38", "33", "40", "27", "17", "92", "72", "51", ...
## $ X3      <chr> "51", "44", "37", "49", "42", "24", "100", "79", "57",...
## $ X4      <chr> "43", "37", "30", "24", "21", "13", "69", "54", "39", ...
## $ X5      <chr> "42", "34", "26", "37", "25", "12", "85", "66", "47", ...
## $ X6      <chr> "45", "42", "38", "45", "40", "36", "100", "93", "85",...
## $ X7      <chr> "38", "30", "21", "36", "20", "-3", "92", "61", "29", ...
## $ X8      <chr> "29", "24", "18", "28", "16", "3", "92", "70", "47", "...
## $ X9      <chr> "49", "39", "29", "49", "41", "28", "100", "93", "86",...
## $ X10     <chr> "48", "43", "38", "45", "39", "37", "100", "95", "89",...
## $ X11     <chr> "39", "36", "32", "37", "31", "27", "92", "87", "82", ...
## $ X12     <chr> "39", "35", "31", "28", "27", "25", "85", "75", "64", ...
## $ X13     <chr> "42", "37", "32", "28", "26", "24", "75", "65", "55", ...
## $ X14     <chr> "45", "39", "33", "29", "27", "25", "82", "68", "53", ...
## $ X15     <chr> "42", "37", "32", "33", "29", "27", "89", "75", "60", ...
## $ X16     <chr> "44", "40", "35", "42", "36", "30", "96", "85", "73", ...
## $ X17     <chr> "49", "45", "41", "46", "41", "32", "100", "85", "70",...
## $ X18     <chr> "44", "40", "36", "34", "30", "26", "89", "73", "57", ...
## $ X19     <chr> "37", "33", "29", "25", "22", "20", "69", "63", "56", ...
## $ X20     <chr> "36", "32", "27", "30", "24", "20", "89", "79", "69", ...
## $ X21     <chr> "36", "33", "30", "30", "27", "25", "85", "77", "69", ...
## $ X22     <chr> "44", "39", "33", "39", "34", "25", "89", "79", "69", ...
## $ X23     <chr> "47", "45", "42", "45", "42", "37", "100", "91", "82",...
## $ X24     <chr> "46", "44", "41", "46", "44", "41", "100", "98", "96",...
## $ X25     <chr> "59", "52", "44", "58", "43", "29", "100", "75", "49",...
## $ X26     <chr> "50", "44", "37", "31", "29", "28", "70", "60", "49", ...
## $ X27     <chr> "52", "45", "38", "34", "31", "29", "70", "60", "50", ...
## $ X28     <chr> "52", "46", "40", "42", "35", "27", "76", "65", "53", ...
## $ X29     <chr> "41", "36", "30", "26", "20", "10", "64", "51", "37", ...
## $ X30     <chr> "30", "26", "22", "10", "4", "-6", "50", "38", "26", "...
## $ X31     <chr> "30", "25", "20", "8", "5", "1", "57", "44", "31", "30...
```

```
# View a summary of the data
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
```

```
##
##
##
```

Take a closer look

```
# View first 6 rows
head(weather, n = 6)
```

```
##   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
```

```
# View first 15 rows
head(weather, n = 15)
```

```
##     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
## 11 11 2014    12 Mean.Sea.Level.PressureIn 30.13 30.59 30.07 30.33 30.59
## 12 12 2014    12  Min.Sea.Level.PressureIn 30.01  30.4 29.87 30.09 30.45
## 13 13 2014    12        Max.VisibilityMiles    10    10    10    10    10
## 14 14 2014    12       Mean.VisibilityMiles    10     8     5    10    10
## 15 15 2014    12        Min.VisibilityMiles    10     2     1    10     5
##       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
## 11 30.24 30.46 30.67 30.04  29.5 29.61 29.85 29.82 29.83 30.05 30.09 29.75
## 12 30.16 30.24 30.51 29.49 29.43 29.44 29.81 29.78 29.78 29.91 29.92 29.69
## 13    10    10    10    10    10    10    10    10    10    10    10    10
## 14     4    10     8     2     3     7    10    10    10    10     9     6
## 15     0     5     2     1     1     1     7    10    10    10     5     1
##      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
## 11 29.78 29.98 30.26 30.32 30.35 30.23  29.9 29.63 30.11 30.14 29.87 30.12
## 12 29.71 29.86 30.17 30.28  30.3 30.16 29.55 29.47 29.99 30.03 29.77    30
## 13    10    10    10    10    10    10     2    10    10    10    10    10
## 14    10    10    10     9    10     5     1     8    10    10    10    10
## 15    10    10     7     6     4     1     0     1    10    10    10    10
##      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
## 11 30.32 30.25
## 12 30.23 30.13
## 13    10    10
## 14    10    10
## 15    10    10
```

```r
# View the last 6 rows
tail(weather, n=6)
```

```
##       X year month            measure   X1   X2   X3   X4   X5   X6   X7
## 281 281 2015    12 Mean.Wind.SpeedMPH    6 <NA> <NA> <NA> <NA> <NA> <NA>
## 282 282 2015    12  Max.Gust.SpeedMPH   17 <NA> <NA> <NA> <NA> <NA> <NA>
## 283 283 2015    12    PrecipitationIn 0.14 <NA> <NA> <NA> <NA> <NA> <NA>
## 284 284 2015    12         CloudCover    7 <NA> <NA> <NA> <NA> <NA> <NA>
## 285 285 2015    12             Events Rain <NA> <NA> <NA> <NA> <NA> <NA>
## 286 286 2015    12     WindDirDegrees  109 <NA> <NA> <NA> <NA> <NA> <NA>
##       X8   X9  X10  X11  X12  X13  X14  X15  X16  X17  X18  X19  X20  X21
## 281 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 282 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 283 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 284 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 285 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 286 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
##      X22  X23  X24  X25  X26  X27  X28  X29  X30  X31
## 281 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 282 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 283 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 284 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 285 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 286 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
```

```r
# View the last 10 rows
tail(weather, n=10)
```

```
##       X year month              measure   X1   X2   X3   X4   X5   X6   X7
## 277 277 2015    12  Max.VisibilityMiles   10 <NA> <NA> <NA> <NA> <NA> <NA>
## 278 278 2015    12 Mean.VisibilityMiles    8 <NA> <NA> <NA> <NA> <NA> <NA>
## 279 279 2015    12  Min.VisibilityMiles    1 <NA> <NA> <NA> <NA> <NA> <NA>
## 280 280 2015    12    Max.Wind.SpeedMPH   15 <NA> <NA> <NA> <NA> <NA> <NA>
## 281 281 2015    12   Mean.Wind.SpeedMPH    6 <NA> <NA> <NA> <NA> <NA> <NA>
## 282 282 2015    12    Max.Gust.SpeedMPH   17 <NA> <NA> <NA> <NA> <NA> <NA>
## 283 283 2015    12      PrecipitationIn 0.14 <NA> <NA> <NA> <NA> <NA> <NA>
## 284 284 2015    12           CloudCover    7 <NA> <NA> <NA> <NA> <NA> <NA>
## 285 285 2015    12               Events Rain <NA> <NA> <NA> <NA> <NA> <NA>
## 286 286 2015    12       WindDirDegrees  109 <NA> <NA> <NA> <NA> <NA> <NA>
##       X8   X9  X10  X11  X12  X13  X14  X15  X16  X17  X18  X19  X20  X21
## 277 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 278 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 279 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 280 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 281 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 282 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 283 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 284 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 285 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 286 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
##      X22  X23  X24  X25  X26  X27  X28  X29  X30  X31
## 277 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 278 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 279 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 280 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 281 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 282 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 283 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 284 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 285 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 286 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
```

Column names are values

```r
# Load the tidyr package
library(tidyr)

# Gather the columns
weather2 <- gather(weather, day, value,X1:X31,na.rm = TRUE)

# View the head
head(weather2)
```

```
##   X year month           measure day value
## 1 1 2014    12  Max.TemperatureF  X1    64
## 2 2 2014    12 Mean.TemperatureF  X1    52
## 3 3 2014    12  Min.TemperatureF  X1    39
## 4 4 2014    12    Max.Dew.PointF  X1    46
## 5 5 2014    12    MeanDew.PointF  X1    40
## 6 6 2014    12     Min.DewpointF  X1    26
```

Values are variable names

```r
# First remove column of row names
without_x <- weather2[, -1]

# Spread the data
weather3 <- spread(without_x, measure, value)

# View the head
head(weather3)
```

```
##   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
##   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               T            286
## 5               T            298
## 6            0.00            306
```

Clean up dates

```
# Load the stringr and lubridate packages
library(stringr)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date
```

```
# Remove X's from day column
weather3$day <- str_replace(weather3$day,"X","")

# Unite the year, month, and day columns
weather4 <- unite(weather3, date, year, month, day, sep = "-")

# Convert date column to proper date format using lubridates's ymd()
weather4$date <- ymd(weather4$date)

# Rearrange columns using dplyr's select()
weather5 <- select(weather4, date, Events, CloudCover:WindDirDegrees)

# View the head of weather5
head(weather5)
```

```
##         date    Events CloudCover Max.Dew.PointF Max.Gust.SpeedMPH
## 1 2014-12-01      Rain          6             46                29
## 2 2014-12-10      Rain          8             45                29
## 3 2014-12-11 Rain-Snow          8             37                28
## 4 2014-12-12      Snow          7             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               T            286
## 5               T            298
## 6            0.00            306
```

A closer look at column types

```r
# View the structure of weather5
str(weather5)
```

```
## 'data.frame':    366 obs. of  23 variables:
##  $ date                     : Date, format: "2014-12-01" "2014-12-10" ...
##  $ Events                   : chr  "Rain" "Rain" "Rain-Snow" "Snow" ...
##  $ CloudCover               : chr  "6" "8" "8" "7" ...
##  $ 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" ...
```

```r
# Examine the first 8 rows of weather5. Are most of the characters numeric?
head(weather5,n=8)
```

```
##         date    Events CloudCover Max.Dew.PointF Max.Gust.SpeedMPH
## 1 2014-12-01      Rain          6             46                29
## 2 2014-12-10      Rain          8             45                29
## 3 2014-12-11 Rain-Snow          8             37                28
## 4 2014-12-12      Snow          7             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      Rain          8             42                10
##   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
##   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
##   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
##   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
##   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
##   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
```

```
# See what happens if we try to convert PrecipitationIn to numeric.
# as.numeric(weather5$PrecipitationIn)
```

Column type conversions

```
# Replace "T" with "0" (T = trace)
weather5$PrecipitationIn <-str_replace(weather5$PrecipitationIn, "T","0")

# Convert characters to numerics
weather6 <- mutate_at(weather5, vars(CloudCover:WindDirDegrees), funs(as.numeric))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

```
# Look at result
str(weather6)
```

```
## 'data.frame':    366 obs. of  23 variables:
##  $ date                     : Date, format: "2014-12-01" "2014-12-10" ...
##  $ Events                   : chr  "Rain" "Rain" "Rain-Snow" "Snow" ...
##  $ CloudCover               : num  6 8 8 7 5 4 2 8 8 7 ...
##  $ 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 0 0 0 0 0 0.43 0.01 ...
##  $ WindDirDegrees           : num  268 357 230 286 298 306 324 79 311 281 ...
```

Find missing values

```
# Count missing values
sum(is.na(weather6))
```

```
## [1] 6
```

```
# Find missing values
summary(weather6)
```

```
##       date                Events            CloudCover    Max.Dew.PointF
##  Min.   :2014-12-01   Length:366         Min.   :0.000   Min.   :-6.00
##  1st Qu.:2015-03-02   Class :character   1st Qu.:3.000   1st Qu.:32.00
##  Median :2015-06-01   Mode  :character   Median :5.000   Median :47.50
##  Mean   :2015-06-01                      Mean   :4.708   Mean   :45.48
##  3rd Qu.:2015-08-31                      3rd Qu.:7.000   3rd Qu.:61.00
##  Max.   :2015-12-01                      Max.   :8.000   Max.   :75.00
##
##  Max.Gust.SpeedMPH Max.Humidity      Max.Sea.Level.PressureIn
##  Min.   : 0.00     Min.   :  39.00   Min.   :29.58
##  1st Qu.:21.00     1st Qu.:  73.25   1st Qu.:30.00
##  Median :25.50     Median :  86.00   Median :30.14
##  Mean   :26.99     Mean   :  85.69   Mean   :30.16
##  3rd Qu.:31.25     3rd Qu.:  93.00   3rd Qu.:30.31
##  Max.   :94.00     Max.   :1000.00   Max.   :30.88
##  NA's   :6
##  Max.TemperatureF Max.VisibilityMiles Max.Wind.SpeedMPH Mean.Humidity
##  Min.   :18.00    Min.   : 2.000      Min.   : 8.00     Min.   :28.00
##  1st Qu.:42.00    1st Qu.:10.000      1st Qu.:16.00     1st Qu.:56.00
##  Median :60.00    Median :10.000      Median :20.00     Median :66.00
##  Mean   :58.93    Mean   : 9.907      Mean   :20.62     Mean   :66.02
##  3rd Qu.:76.00    3rd Qu.:10.000      3rd Qu.:24.00     3rd Qu.:76.75
##  Max.   :96.00    Max.   :10.000      Max.   :38.00     Max.   :98.00
##
##  Mean.Sea.Level.PressureIn Mean.TemperatureF Mean.VisibilityMiles
##  Min.   :29.49             Min.   : 8.00     Min.   :-1.000
##  1st Qu.:29.87             1st Qu.:36.25     1st Qu.: 8.000
##  Median :30.03             Median :53.50     Median :10.000
##  Mean   :30.04             Mean   :51.40     Mean   : 8.861
##  3rd Qu.:30.19             3rd Qu.:68.00     3rd Qu.:10.000
##  Max.   :30.77             Max.   :84.00     Max.   :10.000
##
##  Mean.Wind.SpeedMPH MeanDew.PointF   Min.DewpointF     Min.Humidity
##  Min.   : 4.00      Min.   :-11.00   Min.   :-18.00   Min.   :16.00
##  1st Qu.: 8.00      1st Qu.: 24.00   1st Qu.: 16.25   1st Qu.:35.00
##  Median :10.00      Median : 41.00   Median : 35.00   Median :46.00
##  Mean   :10.68      Mean   : 38.96   Mean   : 32.25   Mean   :48.31
##  3rd Qu.:13.00      3rd Qu.: 56.00   3rd Qu.: 51.00   3rd Qu.:60.00
##  Max.   :22.00      Max.   : 71.00   Max.   : 68.00   Max.   :96.00
##
##  Min.Sea.Level.PressureIn Min.TemperatureF Min.VisibilityMiles
##  Min.   :29.16            Min.   :-3.00    Min.   : 0.000
##  1st Qu.:29.76            1st Qu.:30.00    1st Qu.: 2.000
##  Median :29.94            Median :46.00    Median :10.000
##  Mean   :29.93            Mean   :43.33    Mean   : 6.716
##  3rd Qu.:30.09            3rd Qu.:60.00    3rd Qu.:10.000
##  Max.   :30.64            Max.   :74.00    Max.   :10.000
##
##  PrecipitationIn  WindDirDegrees
##  Min.   :0.0000   Min.   :  1.0
##  1st Qu.:0.0000   1st Qu.:113.0
##  Median :0.0000   Median :222.0
##  Mean   :0.1016   Mean   :200.1
##  3rd Qu.:0.0400   3rd Qu.:275.0
##  Max.   :2.9000   Max.   :360.0
##
```

```r
# Find indices of NAs in Max.Gust.SpeedMPH
ind <- which(is.na(weather6$Max.Gust.SpeedMPH))

# Look at the full rows for records missing Max.Gust.SpeedMPH
weather6[ind, ]
```

```
##           date Events CloudCover Max.Dew.PointF Max.Gust.SpeedMPH
## 161 2015-05-18    Fog          6             52                NA
## 205 2015-06-03                 7             48                NA
## 273 2015-08-08                 4             61                NA
## 275 2015-09-01                 1             63                NA
## 308 2015-10-12                 0             56                NA
## 358 2015-11-03                 1             44                NA
##     Max.Humidity Max.Sea.Level.PressureIn Max.TemperatureF
## 161          100                    30.30               58
## 205           93                    30.31               56
## 273           87                    30.02               76
## 275           78                    30.06               79
## 308           89                    29.86               76
## 358           82                    30.25               73
##     Max.VisibilityMiles Max.Wind.SpeedMPH Mean.Humidity
## 161                  10                16            79
## 205                  10                14            82
## 273                  10                14            68
## 275                  10                15            65
## 308                  10                15            65
## 358                  10                16            57
##     Mean.Sea.Level.PressureIn Mean.TemperatureF Mean.VisibilityMiles
## 161                     30.23                54                    8
## 205                     30.24                52                   10
## 273                     29.99                69                   10
## 275                     30.02                74                   10
## 308                     29.80                64                   10
## 358                     30.13                60                   10
##     Mean.Wind.SpeedMPH MeanDew.PointF Min.DewpointF Min.Humidity
## 161                 10             48            43           57
## 205                  7             45            43           71
## 273                  6             57            54           49
## 275                  9             62            59           52
## 308                  8             51            48           41
## 358                  8             42            40           31
##     Min.Sea.Level.PressureIn Min.TemperatureF Min.VisibilityMiles
## 161                    30.12               49                   0
## 205                    30.19               47                  10
## 273                    29.95               61                  10
## 275                    29.96               69                  10
## 308                    29.74               51                  10
## 358                    30.06               47                  10
##     PrecipitationIn WindDirDegrees
## 161               0             72
## 205               0             90
## 273               0             45
## 275               0             54
## 308                0            199
## 358                0            281
```

An obvious error

```
# Review distributions for all variables
summary(weather6)
```

```
##       date               Events            CloudCover    Max.Dew.PointF
##  Min.   :2014-12-01   Length:366          Min.   :0.000   Min.   :-6.00
##  1st Qu.:2015-03-02   Class :character   1st Qu.:3.000   1st Qu.:32.00
##  Median :2015-06-01   Mode  :character   Median :5.000   Median :47.50
##  Mean   :2015-06-01                      Mean   :4.708   Mean   :45.48
##  3rd Qu.:2015-08-31                      3rd Qu.:7.000   3rd Qu.:61.00
##  Max.   :2015-12-01                      Max.   :8.000   Max.   :75.00
##
##  Max.Gust.SpeedMPH  Max.Humidity     Max.Sea.Level.PressureIn
##  Min.   : 0.00     Min.   :  39.00   Min.   :29.58
##  1st Qu.:21.00     1st Qu.:  73.25   1st Qu.:30.00
##  Median :25.50     Median :  86.00   Median :30.14
##  Mean   :26.99     Mean   :  85.69   Mean   :30.16
##  3rd Qu.:31.25     3rd Qu.:  93.00   3rd Qu.:30.31
##  Max.   :94.00     Max.   :1000.00   Max.   :30.88
##  NA's   :6
##  Max.TemperatureF Max.VisibilityMiles Max.Wind.SpeedMPH Mean.Humidity
##  Min.   :18.00    Min.   : 2.000      Min.   : 8.00     Min.   :28.00
##  1st Qu.:42.00    1st Qu.:10.000      1st Qu.:16.00     1st Qu.:56.00
##  Median :60.00    Median :10.000      Median :20.00     Median :66.00
##  Mean   :58.93    Mean   : 9.907      Mean   :20.62     Mean   :66.02
##  3rd Qu.:76.00    3rd Qu.:10.000      3rd Qu.:24.00     3rd Qu.:76.75
##  Max.   :96.00    Max.   :10.000      Max.   :38.00     Max.   :98.00
##
##  Mean.Sea.Level.PressureIn Mean.TemperatureF Mean.VisibilityMiles
##  Min.   :29.49             Min.   : 8.00     Min.   :-1.000
##  1st Qu.:29.87             1st Qu.:36.25     1st Qu.: 8.000
##  Median :30.03             Median :53.50     Median :10.000
##  Mean   :30.04             Mean   :51.40     Mean   : 8.861
##  3rd Qu.:30.19             3rd Qu.:68.00     3rd Qu.:10.000
##  Max.   :30.77             Max.   :84.00     Max.   :10.000
##
##  Mean.Wind.SpeedMPH MeanDew.PointF   Min.DewpointF     Min.Humidity
##  Min.   : 4.00      Min.   :-11.00   Min.   :-18.00   Min.   :16.00
##  1st Qu.: 8.00      1st Qu.: 24.00   1st Qu.: 16.25   1st Qu.:35.00
##  Median :10.00      Median : 41.00   Median : 35.00   Median :46.00
##  Mean   :10.68      Mean   : 38.96   Mean   : 32.25   Mean   :48.31
##  3rd Qu.:13.00      3rd Qu.: 56.00   3rd Qu.: 51.00   3rd Qu.:60.00
##  Max.   :22.00      Max.   : 71.00   Max.   : 68.00   Max.   :96.00
##
##  Min.Sea.Level.PressureIn Min.TemperatureF Min.VisibilityMiles
##  Min.   :29.16            Min.   :-3.00    Min.   : 0.000
##  1st Qu.:29.76            1st Qu.:30.00    1st Qu.: 2.000
##  Median :29.94            Median :46.00    Median :10.000
##  Mean   :29.93            Mean   :43.33    Mean   : 6.716
##  3rd Qu.:30.09            3rd Qu.:60.00    3rd Qu.:10.000
##  Max.   :30.64            Max.   :74.00    Max.   :10.000
##
##  PrecipitationIn  WindDirDegrees
##  Min.   :0.0000   Min.   :  1.0
##  1st Qu.:0.0000   1st Qu.:113.0
##  Median :0.0000   Median :222.0
##  Mean   :0.1016   Mean   :200.1
##  3rd Qu.:0.0400   3rd Qu.:275.0
##  Max.   :2.9000   Max.   :360.0
##
```

```
# Find row with Max.Humidity of 1000
ind <- which(weather6$Max.Humidity == 1000)

# Look at the data for that day
weather6[ind, ]
```

```
##           date                Events CloudCover Max.Dew.PointF
## 135 2015-04-21 Fog-Rain-Thunderstorm          6             57
##     Max.Gust.SpeedMPH Max.Humidity Max.Sea.Level.PressureIn
## 135                94         1000                    29.75
##     Max.TemperatureF Max.VisibilityMiles Max.Wind.SpeedMPH Mean.Humidity
## 135               65                  10                20            71
##     Mean.Sea.Level.PressureIn Mean.TemperatureF Mean.VisibilityMiles
## 135                      29.6                56                    5
##     Mean.Wind.SpeedMPH MeanDew.PointF Min.DewpointF Min.Humidity
## 135                 10             49            36           42
##     Min.Sea.Level.PressureIn Min.TemperatureF Min.VisibilityMiles
## 135                    29.53               46                   0
##     PrecipitationIn WindDirDegrees
## 135            0.54            184
```

```
# Change 1000 to 100
weather6$Max.Humidity[ind] <- 100
```

Another obvious error

```
# Look at summary of Mean.VisibilityMiles
summary(weather6$Mean.VisibilityMiles)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -1.000   8.000  10.000   8.861  10.000  10.000
```

```
# Get index of row with -1 value
ind <- which(weather6$Mean.VisibilityMiles == -1)

# Look at full row
weather6[ind,]
```

```
##           date Events CloudCover Max.Dew.PointF Max.Gust.SpeedMPH
## 192 2015-06-18                 5             54                23
##     Max.Humidity Max.Sea.Level.PressureIn Max.TemperatureF
## 192           72                    30.14               76
##     Max.VisibilityMiles Max.Wind.SpeedMPH Mean.Humidity
## 192                  10                17            59
##     Mean.Sea.Level.PressureIn Mean.TemperatureF Mean.VisibilityMiles
## 192                     30.04                67                   -1
##     Mean.Wind.SpeedMPH MeanDew.PointF Min.DewpointF Min.Humidity
## 192                 10             49            45           46
##     Min.Sea.Level.PressureIn Min.TemperatureF Min.VisibilityMiles
## 192                    29.93               57                  10
##     PrecipitationIn WindDirDegrees
## 192               0            189
```

```
# Set Mean.VisibilityMiles to the appropriate value
weather6$Mean.VisibilityMiles[ind] <- 10
```

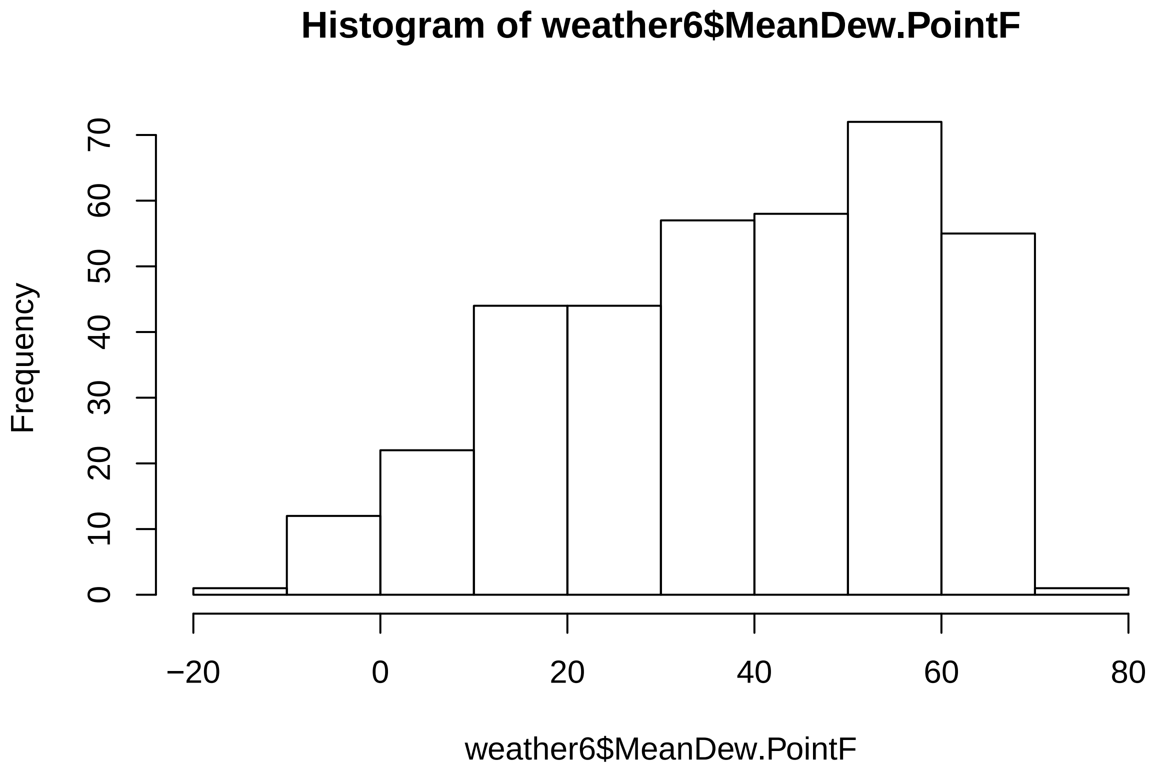
Check other extreme values

```
# Review summary of full data once more
summary(weather6)
```

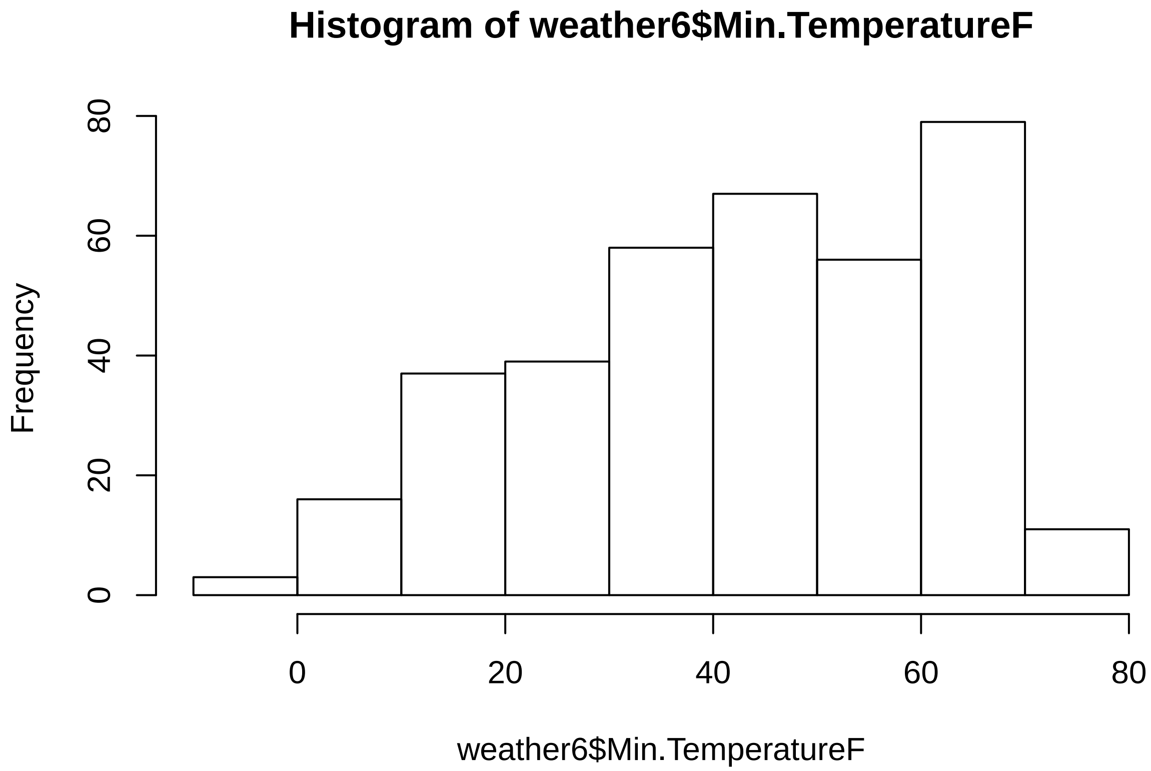
```
##       date               Events            CloudCover    Max.Dew.PointF
##  Min.   :2014-12-01   Length:366         Min.   :0.000   Min.   :-6.00
##  1st Qu.:2015-03-02   Class :character   1st Qu.:3.000   1st Qu.:32.00
##  Median :2015-06-01   Mode  :character   Median :5.000   Median :47.50
##  Mean   :2015-06-01                      Mean   :4.708   Mean   :45.48
##  3rd Qu.:2015-08-31                      3rd Qu.:7.000   3rd Qu.:61.00
##  Max.   :2015-12-01                      Max.   :8.000   Max.   :75.00
##
##  Max.Gust.SpeedMPH  Max.Humidity    Max.Sea.Level.PressureIn
##  Min.   : 0.00     Min.   : 39.00   Min.   :29.58
##  1st Qu.:21.00     1st Qu.: 73.25   1st Qu.:30.00
##  Median :25.50     Median : 86.00   Median :30.14
##  Mean   :26.99     Mean   : 83.23   Mean   :30.16
##  3rd Qu.:31.25     3rd Qu.: 93.00   3rd Qu.:30.31
##  Max.   :94.00     Max.   :100.00   Max.   :30.88
##  NA's   :6
##  Max.TemperatureF Max.VisibilityMiles Max.Wind.SpeedMPH Mean.Humidity
##  Min.   :18.00    Min.   : 2.000      Min.   : 8.00     Min.   :28.00
##  1st Qu.:42.00    1st Qu.:10.000      1st Qu.:16.00     1st Qu.:56.00
##  Median :60.00    Median :10.000      Median :20.00     Median :66.00
##  Mean   :58.93    Mean   : 9.907      Mean   :20.62     Mean   :66.02
##  3rd Qu.:76.00    3rd Qu.:10.000      3rd Qu.:24.00     3rd Qu.:76.75
##  Max.   :96.00    Max.   :10.000      Max.   :38.00     Max.   :98.00
##
##  Mean.Sea.Level.PressureIn Mean.TemperatureF Mean.VisibilityMiles
##  Min.   :29.49             Min.   : 8.00     Min.   : 1.000
##  1st Qu.:29.87             1st Qu.:36.25     1st Qu.: 8.000
##  Median :30.03             Median :53.50     Median :10.000
##  Mean   :30.04             Mean   :51.40     Mean   : 8.891
##  3rd Qu.:30.19             3rd Qu.:68.00     3rd Qu.:10.000
##  Max.   :30.77             Max.   :84.00     Max.   :10.000
##
```

```
##  Mean.Wind.SpeedMPH MeanDew.PointF   Min.DewpointF     Min.Humidity
##  Min.   : 4.00      Min.   :-11.00   Min.   :-18.00   Min.   :16.00
##  1st Qu.: 8.00      1st Qu.: 24.00   1st Qu.: 16.25   1st Qu.:35.00
##  Median :10.00      Median : 41.00   Median : 35.00   Median :46.00
##  Mean   :10.68      Mean   : 38.96   Mean   : 32.25   Mean   :48.31
##  3rd Qu.:13.00      3rd Qu.: 56.00   3rd Qu.: 51.00   3rd Qu.:60.00
##  Max.   :22.00      Max.   : 71.00   Max.   : 68.00   Max.   :96.00
##
##  Min.Sea.Level.PressureIn Min.TemperatureF Min.VisibilityMiles
##  Min.   :29.16            Min.   :-3.00    Min.   : 0.000
##  1st Qu.:29.76            1st Qu.:30.00    1st Qu.: 2.000
##  Median :29.94            Median :46.00    Median :10.000
##  Mean   :29.93            Mean   :43.33    Mean   : 6.716
##  3rd Qu.:30.09            3rd Qu.:60.00    3rd Qu.:10.000
##  Max.   :30.64            Max.   :74.00    Max.   :10.000
##
##  PrecipitationIn  WindDirDegrees
##  Min.   :0.0000   Min.   :  1.0
##  1st Qu.:0.0000   1st Qu.:113.0
##  Median :0.0000   Median :222.0
##  Mean   :0.1016   Mean   :200.1
##  3rd Qu.:0.0400   3rd Qu.:275.0
##  Max.   :2.9000   Max.   :360.0
##
```

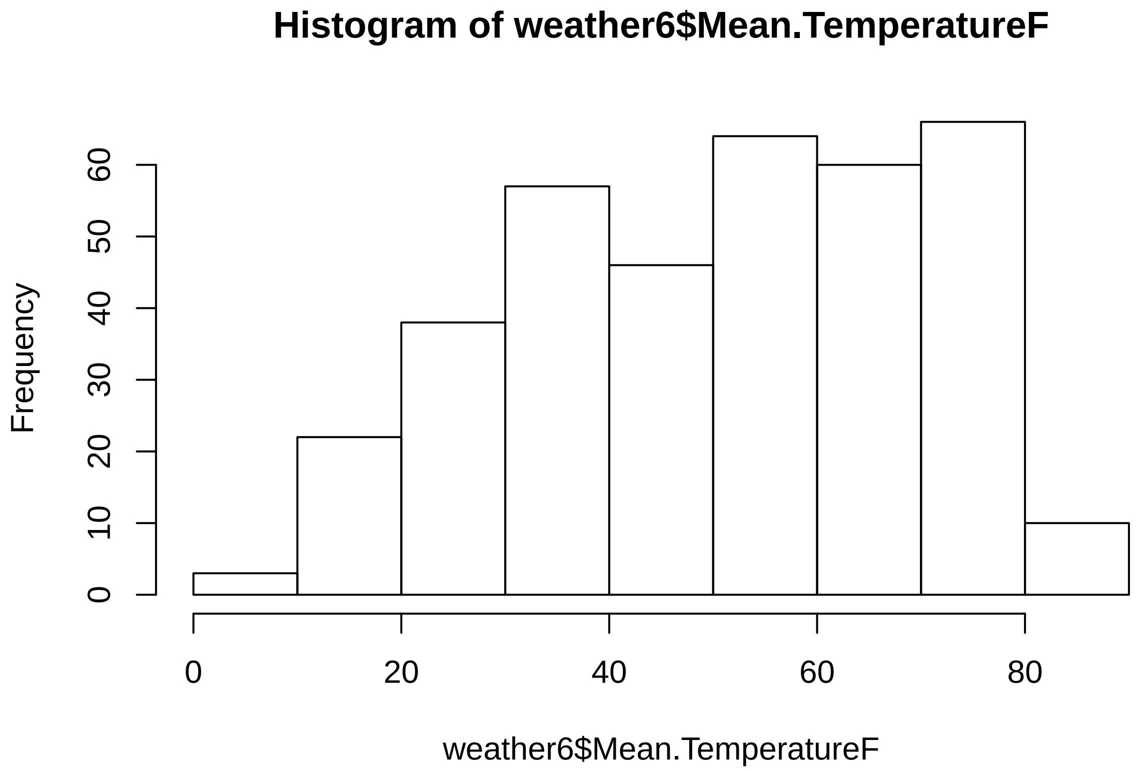
```r
# Look at histogram for MeanDew.PointF
hist(weather6$MeanDew.PointF)
```



```r
# Look at histogram for Min.TemperatureF
hist(weather6$Min.TemperatureF)
```



```r
# Compare to histogram for Mean.TemperatureF
hist(weather6$Mean.TemperatureF)
```



Finishing touches

```
# Clean up column names
new_colnames <- tolower(colnames(weather6))
names(weather6) <- new_colnames

# Replace empty cells in events column
weather6$events[weather6$events == ""] <- "None"

# Print the first 6 rows of weather6
head(weather6, n = 6)
```

```
##         date    events cloudcover max.dew.pointf max.gust.speedmph
## 1 2014-12-01      Rain          6             46                29
## 2 2014-12-10      Rain          8             45                29
## 3 2014-12-11 Rain-Snow          8             37                28
## 4 2014-12-12      Snow          7             28                21
## 5 2014-12-13      None          5             28                23
## 6 2014-12-14      None          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                     29.50                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            0.00            286
## 5            0.00            298
## 6            0.00            306
```