Assignment1

Weather dataset downloaded from: https://vincentarelbundock.github.io/Rdatasets/articles/data.html

Documentation: https://vincentarelbundock.github.io/Rdatasets/doc/mosaicData/Weather.html

Importing the dataset into R:

```
weather <- read.csv("C:/Users/lmszr/Documents/School/Fundamentals of Machine Learning/Weather.csv")
```

Descriptive summary

```
summary(weather)
```

```
##        X              city               date                year     
##  Min.   :   1.0   Length:3655        Length:3655        Min.   :2016  
##  1st Qu.: 914.5   Class :character   Class :character   1st Qu.:2016  
##  Median :1828.0   Mode  :character   Mode  :character   Median :2016  
##  Mean   :1828.0                                        Mean   :2016  
##  3rd Qu.:2741.5                                        3rd Qu.:2017  
##  Max.   :3655.0                                        Max.   :2017  
##                                                                      
##      month             day           high_temp         avg_temp    
##  Min.   : 1.00   Min.   : 1.00   Min.   :  5.00   Min.   : 1.00  
##  1st Qu.: 4.00   1st Qu.: 8.00   1st Qu.: 62.00   1st Qu.:56.00  
##  Median : 7.00   Median :16.00   Median : 73.00   Median :66.00  
##  Mean   : 6.52   Mean   :15.74   Mean   : 71.32   Mean   :63.26  
##  3rd Qu.:10.00   3rd Qu.:23.00   3rd Qu.: 86.00   3rd Qu.:76.00  
##  Max.   :12.00   Max.   :31.00   Max.   :102.00   Max.   :90.00  
##                                                                  
##     low_temp        high_dewpt       avg_dewpt        low_dewpt     
##  Min.   :-13.00   Min.   :-17.00   Min.   :-24.00   Min.   :-29.00  
##  1st Qu.: 47.00   1st Qu.: 49.00   1st Qu.: 43.00   1st Qu.: 36.00  
##  Median : 59.00   Median : 59.00   Median : 55.00   Median : 50.00  
##  Mean   : 54.95   Mean   : 55.97   Mean   : 50.99   Mean   : 45.41  
##  3rd Qu.: 66.00   3rd Qu.: 67.00   3rd Qu.: 63.00   3rd Qu.: 59.00  
##  Max.   : 84.00   Max.   : 86.00   Max.   : 80.00   Max.   : 79.00  
##                                                                     
##  high_humidity     avg_humidity     low_humidity      high_hg      
##  Min.   : 24.00   Min.   : 12.00   Min.   : 4.00   Min.   :29.30  
##  1st Qu.: 78.00   1st Qu.: 57.00   1st Qu.:30.00   1st Qu.:29.89  
##  Median : 88.00   Median : 70.00   Median :51.00   Median :30.01  
##  Mean   : 85.23   Mean   : 66.46   Mean   :46.34   Mean   :30.04  
##  3rd Qu.: 94.00   3rd Qu.: 79.00   3rd Qu.:63.00   3rd Qu.:30.16  
##  Max.   :100.00   Max.   :100.00   Max.   :94.00   Max.   :30.91  
##                                                                   
##      avg_hg          low_hg         high_vis         avg_vis     
##  Min.   :29.20   Min.   :29.06   Min.   : 1.000   Min.   : 0.000  
```

```
##  1st Qu.:29.82   1st Qu.:29.74   1st Qu.: 6.000   1st Qu.: 4.000
##  Median :29.95   Median :29.88   Median :10.000   Median : 6.000
##  Mean   :29.96   Mean   :29.89   Mean   : 8.226   Mean   : 6.712
##  3rd Qu.:30.09   3rd Qu.:30.03   3rd Qu.:10.000   3rd Qu.:10.000
##  Max.   :30.84   Max.   :30.80   Max.   :19.000   Max.   :19.000
##
##     low_vis         high_wind         avg_wind         low_wind
##  Min.   : 0.000   Min.   :  4.00   Min.   : 1.000   Min.   :  8.00
##  1st Qu.: 1.000   1st Qu.: 13.00   1st Qu.: 5.000   1st Qu.: 18.00
##  Median : 4.000   Median : 15.00   Median : 6.000   Median : 22.00
##  Mean   : 4.598   Mean   : 16.29   Mean   : 7.464   Mean   : 24.38
##  3rd Qu.: 8.000   3rd Qu.: 20.00   3rd Qu.: 9.000   3rd Qu.: 29.00
##  Max.   :19.000   Max.   :128.00   Max.   :28.000   Max.   :140.00
##                                                     NA's   :1326
##     precip             events
##  Length:3655        Length:3655
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

Mean of average temperatures:

```
mean(weather$avg_temp)
```

```
## [1] 63.25554
```

Standard deviation of average temperatures:

```
sd(weather$avg_temp)
```

```
## [1] 16.67659
```

IQR of average temperatures:

```
IQR(weather$avg_temp)
```

```
## [1] 20
```

Count of different weather events:

```
table(weather$events)
```

```
##
##                         Fog                   Fog , Rain
##                         115                           44
##          Fog , Rain , Snow  Fog , Rain , Thunderstorm
##                           5                           10
##                Fog , Snow  Fog , Snow , Thunderstorm
```

```
##                          12                           1
##         Fog , Thunderstorm                        Rain
##                           1                         910
##                Rain , Hail Rain , Hail , Thunderstorm
##                           7                           5
##                Rain , Snow        Rain , Thunderstorm
##                          15                         229
##                        Snow               Thunderstorm
##                          66                          18
```

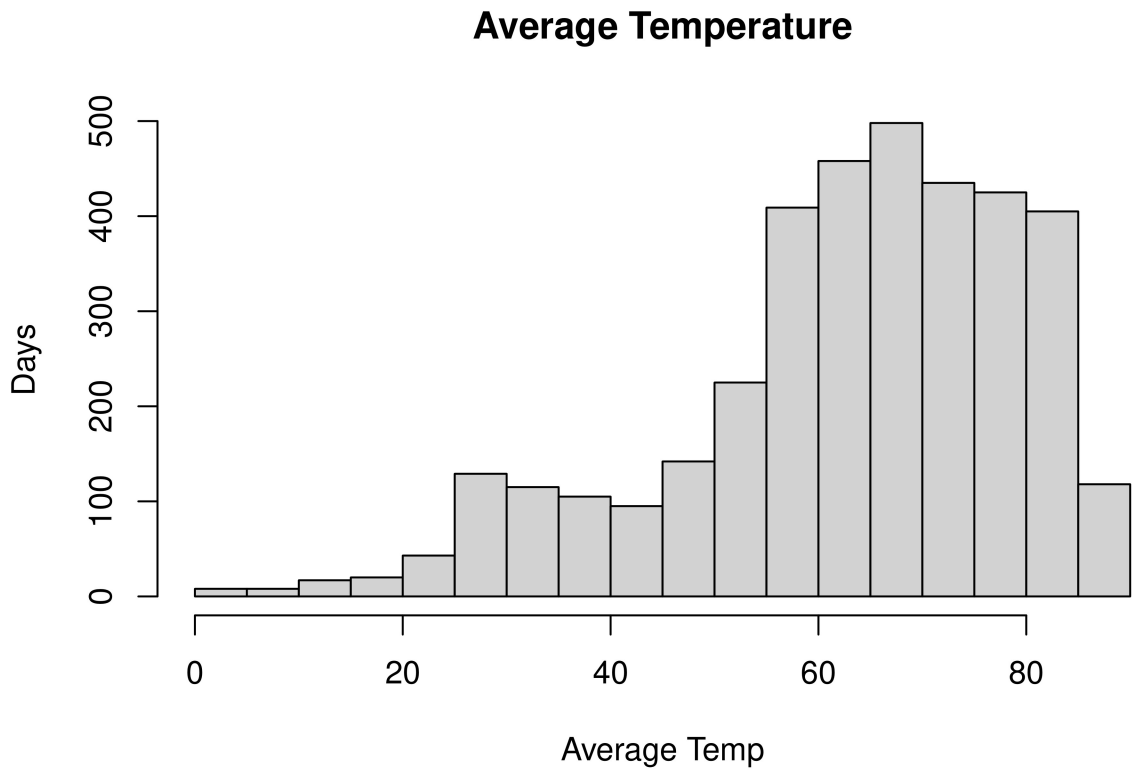
Log transform average temperature and print short sample:

```
avg_temp_log <- log(weather$avg_temp)
avg_temp_log[1:10]
```

```
##  [1] 4.174387 4.189655 4.276666 4.189655 4.127134 4.143135 4.189655 4.248495
##  [9] 4.189655 4.189655
```

Average temperature histogram:

```
hist(weather$avg_temp, xlab = 'Average Temp', ylab = 'Days', main = 'Average Temperature')
```



Scatterplot of average temperature and average humidity:

```
plot(weather$avg_temp, weather$avg_humidity, xlab = 'Average Temp', ylab = 'Average Humidity')
```