Data Dictionary

Saeid Abolfazli

May 2, 2016

The file "data/climate_change.csv" contains climate data from **May 1983** to **December 2008**. The data and below description are provided by team of MITx: 15.071x The Analytics Edge @ EDX

Features:

The available variables include:

- **Year:** the observation year.
- **Month:** the observation month.
- **Temp:** the difference in degrees Celsius between the average global temperature in that period and a reference value. This data comes from the Climatic Research Unit at the University of East Anglia.
- **CO2, N2O, CH4, CFC.11, CFC.12:** atmospheric concentrations of carbon dioxide (CO2), nitrous oxide (N2O), methane (CH4), trichlorofluoromethane (CCl3F; commonly referred to as CFC-11) and dichlorodifluoromethane (CCl2F2; commonly referred to as CFC-12), respectively. This data comes from the ESRL/NOAA Global Monitoring Division.
- **Aerosols:** the mean stratospheric aerosol optical depth at 550 nm. This variable is linked to volcanoes, as volcanic eruptions result in new particles being added to the atmosphere, which affect how much of the sun's energy is reflected back into space. This data is from the Godard Institute for Space Studies at NASA.
- **TSI:** the total solar irradiance (TSI). Due to sunspots and other solar phenomena, the amount of energy that is given off by the sun varies substantially with time. This data is from the SOLARIS-HEPPA project website.
- **MEI:** multivariate El Nino Southern Oscillation index (MEI), a measure of the strength of the El Nino/La Nina-Southern Oscillation (a weather effect in the Pacific Ocean that affects global temperatures). This data comes from the ESRL/NOAA Physical Sciences Division.

Units

- CO2, N2O and CH4 are expressed in **ppmv** (parts per million by volume – i.e., 397 ppmv of CO2 means that CO2 constitutes 397 millionths of the total volume of the atmosphere)
- CFC.11 and CFC.12 are expressed in **ppbv** (parts per billion by volume).
- TSI is expressed in **W/m2** (the rate at which the sun's energy is deposited per unit area)

Data Structure

```
str(CC)
```

```
## 'data.frame':    308 obs. of  11 variables:
##  $ Year     : int  1983 1983 1983 1983 1983 1983 1983 1983 1984 1984 ...
##  $ Month    : int  5 6 7 8 9 10 11 12 1 2 ...
##  $ MEI      : num  2.556 2.167 1.741 1.13 0.428 ...
##  $ CO2      : num  346 346 344 342 340 ...
##  $ CH4      : num  1639 1634 1633 1631 1648 ...
##  $ N2O      : num  304 304 304 304 304 ...
##  $ CFC.11   : num  191 192 193 194 194 ...
##  $ CFC.12   : num  350 352 354 356 357 ...
##  $ TSI      : num  1366 1366 1366 1366 1366 ...
##  $ Aerosols: num  0.0863 0.0794 0.0731 0.0673 0.0619 0.0569 0.0524 0.0486 0.0451 0.0416 ...
##  $ Temp     : num  0.109 0.118 0.137 0.176 0.149 0.093 0.232 0.078 0.089 0.013 ...
```

Data Summary

```
summary(CC)
```

```
##       Year          Month             MEI               CO2        
##  Min.   :1983   Min.   : 1.000   Min.   :-1.6350   Min.   :340.2  
##  1st Qu.:1989   1st Qu.: 4.000   1st Qu.:-0.3987   1st Qu.:353.0  
##  Median :1996   Median : 7.000   Median : 0.2375   Median :361.7  
##  Mean   :1996   Mean   : 6.552   Mean   : 0.2756   Mean   :363.2  
##  3rd Qu.:2002   3rd Qu.:10.000   3rd Qu.: 0.8305   3rd Qu.:373.5  
##  Max.   :2008   Max.   :12.000   Max.   : 3.0010   Max.   :388.5  
##       CH4            N2O            CFC.11          CFC.12      
##  Min.   :1630   Min.   :303.7   Min.   :191.3   Min.   :350.1  
##  1st Qu.:1722   1st Qu.:308.1   1st Qu.:246.3   1st Qu.:472.4  
##  Median :1764   Median :311.5   Median :258.3   Median :528.4  
##  Mean   :1750   Mean   :312.4   Mean   :252.0   Mean   :497.5  
##  3rd Qu.:1787   3rd Qu.:317.0   3rd Qu.:267.0   3rd Qu.:540.5  
##  Max.   :1814   Max.   :322.2   Max.   :271.5   Max.   :543.8  
##       TSI          Aerosols            Temp        
##  Min.   :1365   Min.   :0.00160   Min.   :-0.2820  
##  1st Qu.:1366   1st Qu.:0.00280   1st Qu.: 0.1217  
##  Median :1366   Median :0.00575   Median : 0.2480  
##  Mean   :1366   Mean   :0.01666   Mean   : 0.2568  
##  3rd Qu.:1366   3rd Qu.:0.01260   3rd Qu.: 0.4073  
##  Max.   :1367   Max.   :0.14940   Max.   : 0.7390
```

As you can see dataset includes `dim(CC)[1]` observations and `{r dim(CC)[2]}` features, namely `paste0(names(CC))`