Our code is made of three main files: get\_data.R, bootstrap\_tmy.R , visualization.R

1. In get\_data.R, we set a function named *get\_data()*, the default input is all the TMY files in the folder *data*, and the output is a data frame contained all rows in 22 TMY flies.

Run the function:



The output data frame:



表格

描述已自动生成

1. The second file named bootstrap\_tmy.R is the core of our procedure, consist of function *bootstrap\_tmy()* and function *repeat\_bootstrap()*, the first function is applied in the second function.

The input of bootstrap\_tmy() is a data frame as the source to bootstrap, and in this project the source is the output df1 from former function df1 = get\_data(), and it will output a new data frame produced by df1 as the result of bootstrap, that is a one-year new TMY file. This function is the basics for the second fuction repeat\_bootsrap(n, df)

And in the function repeat\_bootsrap(n, df), n is the repeat times of running bootstrap, if user input n=30, it will get 30 new one-year TMY files, and df is the data frame as the source of bootstrap. The output of this function will be n csv files, which are all one-year TMY files using bootstrap.

When n = 5:



It will generate 5 csv files in folder automatically:

图片包含 应用程序

描述已自动生成

1. We used visualization.R to validate and visualize the outcome.

In my example I used one of the outcomes of bootstrap *tmy\_1.csv* to do the comparison with the first one-year TMY file in original data. Tmy\_1 has 8760 observations(hours), which is equal to 24\*365, and this value may have slight difference because of the randomness of bootstrap.

And I choose the Temperature as factor to plot figure shows the comparison of two data frame, this factor can be any one variable exists in TMY file.

图表, 折线图, 直方图

描述已自动生成

In this figure, we can see the general trend of two data frames is similar.