Homework 2　12332288 胡蓓慧

1. Significant earthquakes since 2150 B.C.

The Significant Earthquake Database contains information on destructive earthquakes from 2150 B.C. to the present. On the top left corner, select all columns and download the entire significant earthquake data file in .tsv format by clicking the Download TSV File button. Click the variable name for more information. Read the file (e.g., earthquakes-2023-10-24_16-20-01_+0800.tsv) as an object and name it Sig_Eqs.

1.1 [5 points] Compute the total number of deaths caused by earthquakes since 2150 B.C. in each country, and then print the top ten countries along with the total number of deaths.

```
Country
CHINA         2075045.0
TURKEY        1188881.0
IRAN          1011449.0
ITALY          498478.0
SYRIA          439224.0
HAITI          323478.0
AZERBAIJAN     317219.0
JAPAN          279085.0
ARMENIA        191890.0
PAKISTAN       145083.0
```

Figure 1 top ten countries along with the total number of deaths (the right colomn)

1.2 [10 points] Compute the total number of earthquakes with magnitude larger than 6.0 (use column Mag as the magnitude) worldwide each year, and then plot the time series. Do you observe any trend? Explain why or why not?

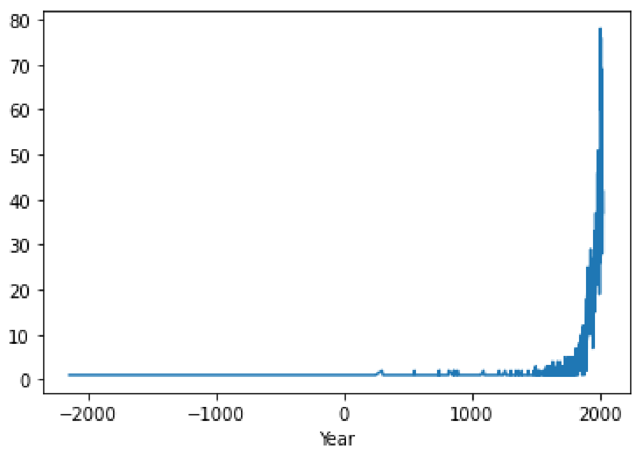

Figure 2 the time series of the total number of earthquakes with magnitude larger than 6.0

I found that the total number of earthquakes with magnitude larger than 6.0 has been increasing volatility from BC to the present. The reason might be due to the fact that the way people record the occurrence of earthquakes and their magnitude has become more comprehensive and accurate over time, another possibility is that the Earth is now in a period of seismic activity.

1.3 [10 points] Write a function CountEq_LargestEq that returns both (1) the total number of earthquakes since 2150 B.C. in a given country AND (2) the date of the largest earthquake ever happened in this country. Apply CountEq_LargestEq to every country in the file, report your results in a descending order.

I got inspired by reading :

https://blog.csdn.net/weixin_46969441/article/details/119859417

https://blog.csdn.net/sunmingyang1987/article/details/106354619

I write a function to solve this problem. In my function, I transform data information into a List since that there might be several of the largest earthquakes tied for the most. The result is as follows:

```
            Country Count       Date
0             CHINA   589  1668-7-25
0             JAPAN   351  2011-3-11
0         INDONESIA   331 2004-12-26
0              IRAN   259  856-12-22
0               USA   223  1964-3-28
..              ...   ...        ...
0     FRENCH GUIANA     1   1885-8-4
0  FRENCH POLYNESIA     1  1848-7-12
0              TOGO     1       1788
0      SIERRA LEONE     1  1795-5-20
0           COMOROS     1  2018-5-15

[169 rows x 3 columns]
```

Figure 3 result of problem 1.3

2. Wind speed in Shenzhen during the past 10 years

In this problem set, we will examine how wind speed changes in Shenzhen during the past 10 years, we will take a look at the hourly weather data measured at the BaoAn International Airport. The data set is from NOAA Integrated Surface Dataset. Download the file 2281305.zip, where the number 2281305 is the site ID. Extract the zip file, you should see a file named 2281305.csv. Save the .csv file to your working directory.

Read page 8-9 (POS 65-69 and POS 70-70) of the comprehensive user guide for the detailed format of the wind data. Explain how you filter the data in your report.
[10 points] Plot monthly averaged wind speed as a function of the observation time. Is there a trend in monthly averaged wind speed within the past 10 years?

I write a function named "splictWND" to filter the data. This function split the value of 'WND' colomn by ','. If the substring which means speed is equal to '9999' (means missing value), the function will return np.nan . My plot result is Figure 4, from this figure, I find that the average wind speed fluctuates particularly much over the course of a year, the average wind speeds of June and July are higher than other months. Within the past 10 years, the trend on the interannual scale is less pronounced, however, the range of fluctuations in average wind speed has become larger in recent years.

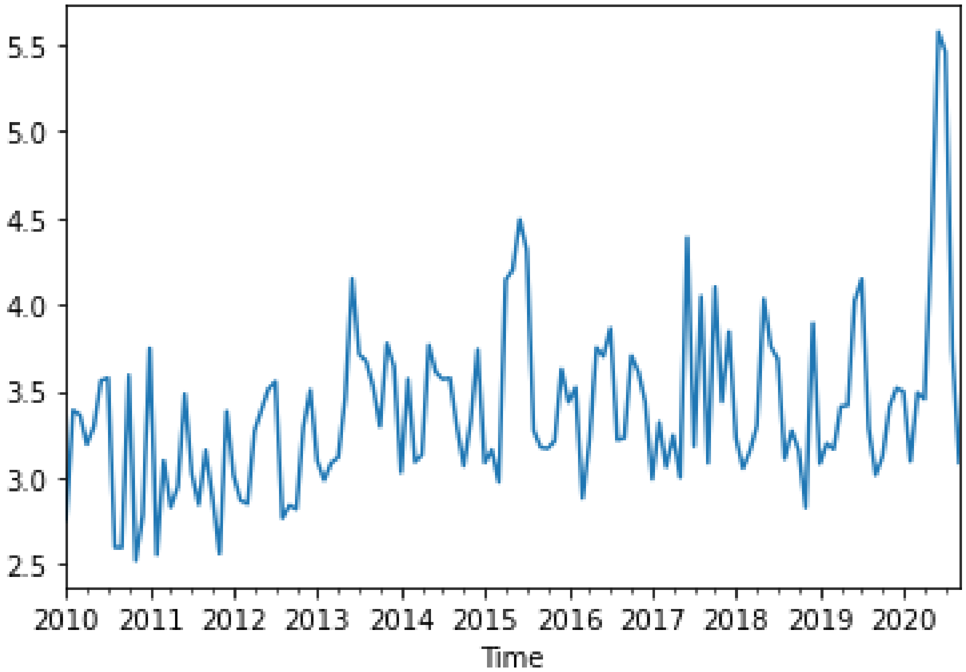

Figure 4 the time series of monthly averaged wind speed

3. Explore a data set

Browse the CASEarth, National Centers for Environmental Information (NCEI), or Advanced Global Atmospheric Gases Experiment (AGAGE) website. Search and download a data set you are interested in. You are also welcome to use data from your group in this problem set. But the data set should be in csv, XLS, or XLSX format, and have temporal information.

3.1 [5 points] Load the csv, XLS, or XLSX file, and clean possible data points with missing values or bad quality.

I use **pd.read_csv()** to load the csv data, and create functions named "splitSLP" and "splitTMP" to found the missing values or bad quality. The principle is like function "splictWND" in problem2. Then I use function **dataframe.dropna()** to delete rows with missing values or bad quality.

3.2 [5 points] Plot the time series of a certain variable.

I plot the time series of daily mean value of sea level pressure, the result is as follows:

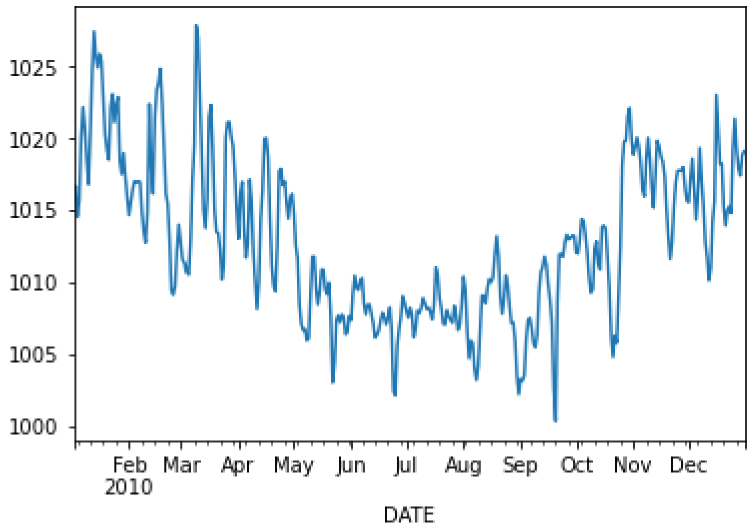

Figure 5 daily mean value of sea level pressure of 2010

3.3 [5 points] Conduct at least 5 simple statistical checks with the variable, and report your findings.

From Figure 6, I find that the sea level pressure is low in summer and high in winter.

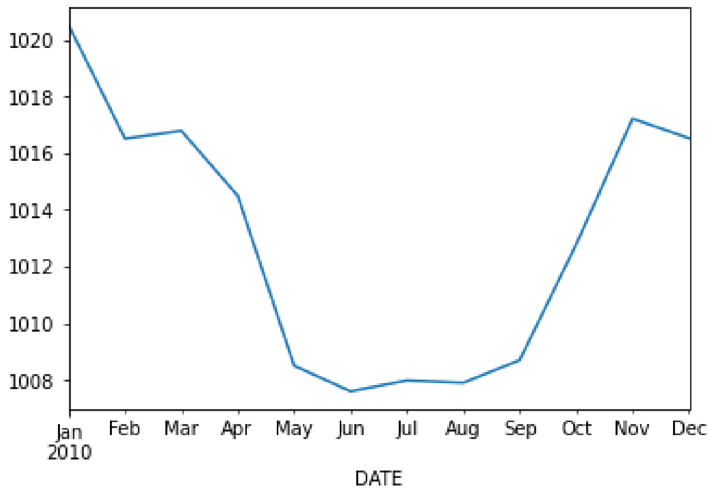

Figure 6 monthly mean value of sea level pressure of 2010

From Figure 7 and Figure 8, I find that there is a cyclical change of monthly standard deviation and variance of sea level pressure in 2010.

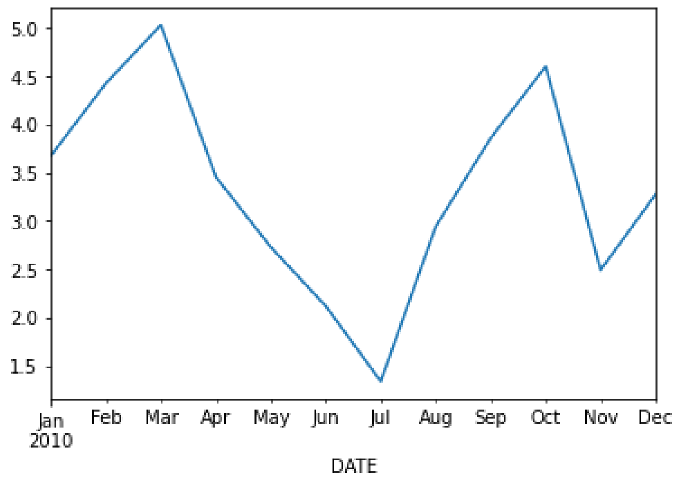

Figure 7 monthly standard deviation value of sea level pressure of 2010

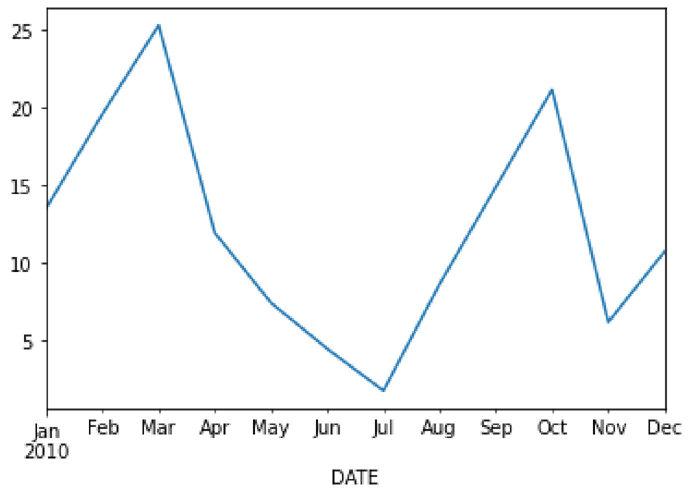

Figure 8 monthly variance value of sea level pressure of 2010

From Figure 9 and Figure 10, I find that the monthly maxinum and minium value have the same trend within 2010.

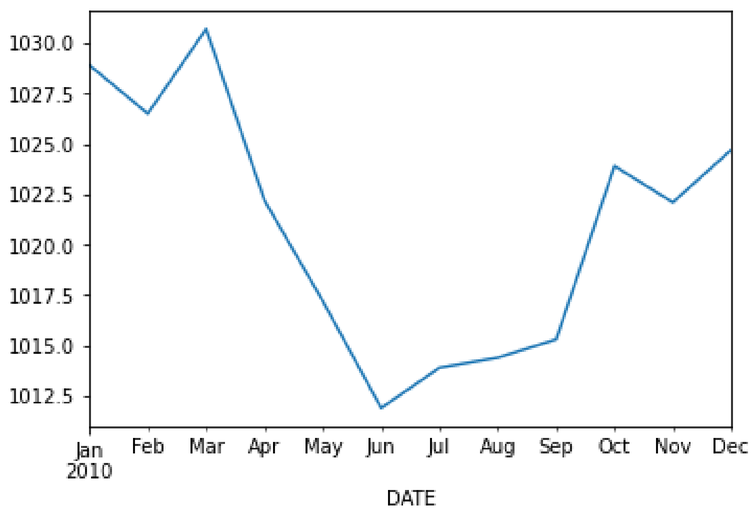

Figure 9 monthly max value of sea level pressure of 2010

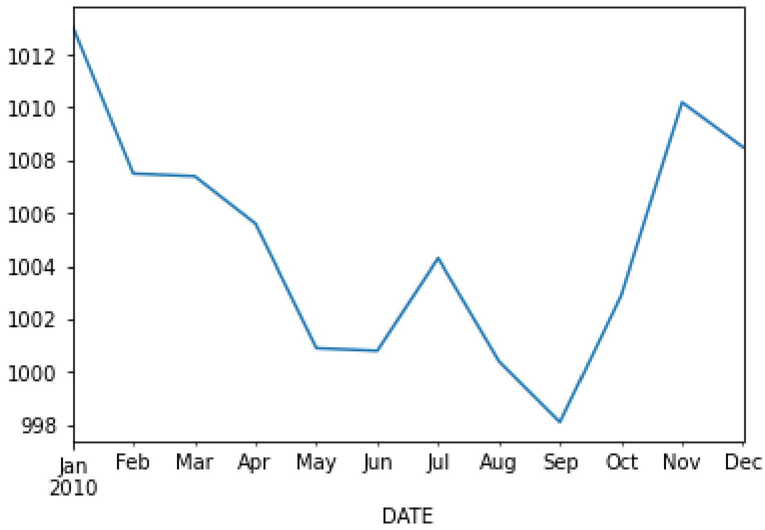

Figure 10 monthly min value of sea level pressure of 2010