12332294 李普双

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.

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?

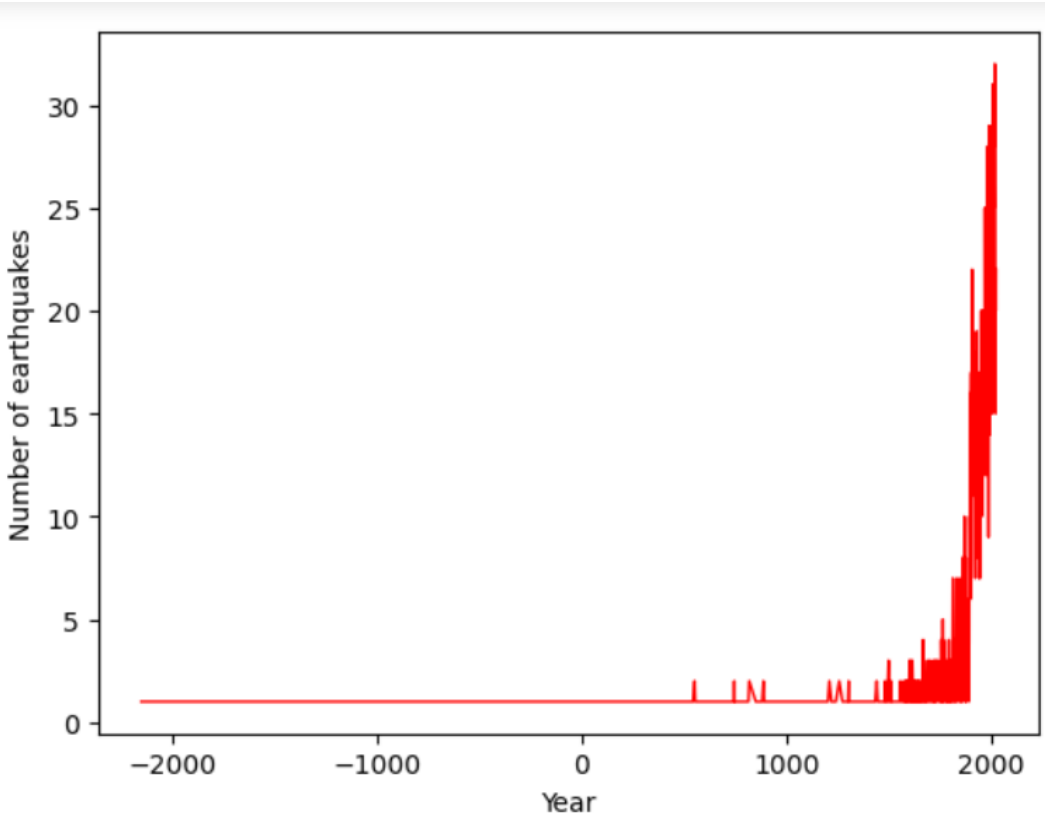
1.3 [10 points] Write a function `CountEq_LargestEq` that returns both (1) the total number of earthquakes since 2150 B.C. in a given country AND (2) the date of the largest earthquake ever happened in this country. Apply `CountEq_LargestEq` to every country in the file, report your results in a descending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

Result

1.1

```
Top Ten Countries by Total Deaths Since 2150 B.C.:
Country
CHINA         2041929.0
TURKEY         995648.0
IRAN           758650.0
SYRIA          437700.0
ITALY          422679.0
JAPAN          356083.0
HAITI          323776.0
AZERBAIJAN     310119.0
INDONESIA      282819.0
ARMENIA        189000.0
Name: Total Deaths, dtype: float64
```

1.2



From the graph above, it is obvious that the number of earthquakes with magnitude larger than 6.0 gets higher rapidly in recent years.
That may be because the data collection is gradually completed in recent years.

1.3

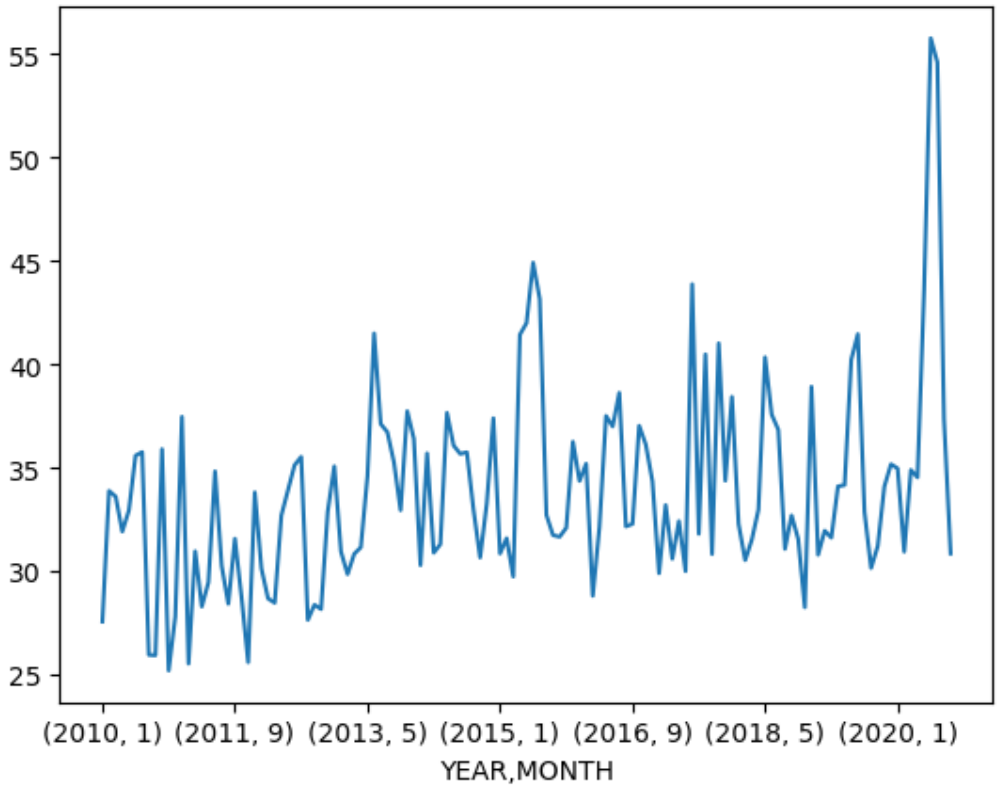
```
          Country  Total Earthquakes  \
0           CHINA                620
1           JAPAN                414
2       INDONESIA                411
3            IRAN                384
4          TURKEY                335
..            ...                ...
132       GRENADA                  1
131        NORWAY                  1
130  SIERRA LEONE                  1
129       IRELAND                  1
156           NaN                  0

                                Largest Earthquake Info
0                Year   Mo    Dy
982  1668.0  7.0  25.0
1                Year   Mo    Dy
5742  2011.0  3.0  11.0
2               Year    Mo    Dy
5340  2004.0  12.0  26.0
3                Year    Mo    Dy
238  856.0  12.0  22.0
4               Year    Mo    Dy
3412  1939.0  12.0  2...
..                                                  ...
132  Empty DataFrame
Columns: [Year, Mo, Dy]
Index: []
131              Year   Mo    Dy
1569  1819.0  8.0  31.0
130              Year   Mo    Dy
1444  1795.0  5.0  20.0
129  Empty DataFrame
Columns: [Year, Mo, Dy]
Index: []
156  Empty DataFrame
Columns: [Year, Mo, Dy]
```

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 dat 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?



总体风速呈逐年上升趋势。

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.

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

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

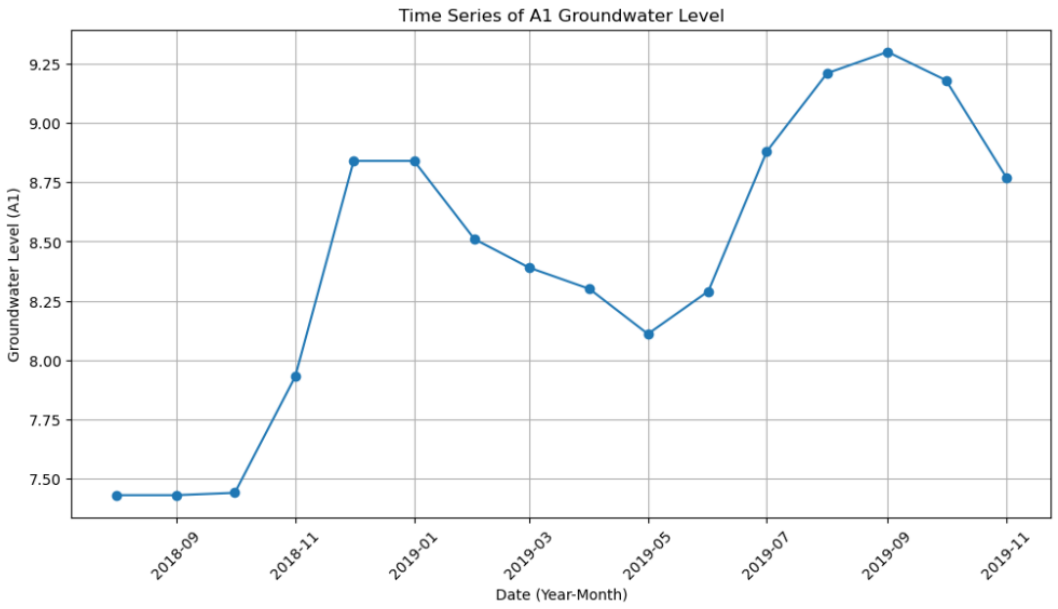
3.1

	年	月	A1	A2	A3	A4	A5	D1	D2	D3	...	W9	\
0	2018	8	7.43	7.99	7.63	8.39	8.47	15.18	8.17	6.13	...	10.76	
1	2018	9	7.43	8.00	7.62	8.38	8.47	15.14	8.06	6.12	...	10.72	
2	2018	10	7.44	8.01	7.62	8.36	8.46	15.13	7.94	6.10	...	10.65	
3	2018	11	7.93	7.60	7.80	8.32	8.42	15.15	7.94	6.12	...	10.66	
4	2018	12	8.84	8.44	8.51	8.38	8.48	15.15	7.93	6.13	...	10.69	
5	2019	1	8.84	8.48	8.60	8.44	8.53	15.09	7.91	6.16	...	10.72	
6	2019	2	8.51	8.26	8.49	8.40	8.53	15.04	7.89	6.19	...	10.84	
7	2019	3	8.39	8.17	8.50	8.50	8.66	15.02	7.90	6.20	...	10.92	
8	2019	4	8.30	8.06	8.37	8.36	8.52	15.11	7.97	6.14	...	10.83	
9	2019	5	8.11	7.87	8.21	8.21	8.37	15.08	8.12	6.15	...	10.76	
10	2019	6	8.29	7.92	8.12	8.05	8.19	15.05	8.27	6.17	...	10.73	
11	2019	7	8.88	8.35	8.30	8.09	8.18	15.00	8.19	6.15	...	10.74	
12	2019	8	9.21	8.69	8.56	8.28	8.31	14.88	7.90	6.11	...	10.71	
13	2019	9	9.30	8.84	8.80	8.52	8.54	14.80	7.79	6.10	...	10.72	
14	2019	10	9.18	8.75	8.79	8.58	8.60	15.06	8.00	6.14	...	10.70	
15	2019	11	8.77	8.46	8.66	8.55	8.63	15.15	7.94	6.12	...	10.70	

	Y1	Y2	Y3	Y4	Y5	H1	H2	H3	H4
0	5.86	2.60	5.99	5.92	7.14	8.09	7.91	10.73	9.27
1	6.13	2.60	5.65	5.64	7.07	8.03	7.86	10.69	9.26
2	6.22	2.61	5.94	5.76	7.05	8.02	7.85	10.68	9.25
3	6.27	5.84	6.23	6.19	6.01	8.07	7.91	10.69	5.01
4	6.18	6.06	6.17	6.16	5.96	8.08	7.91	10.70	7.13
5	6.14	6.02	6.14	6.13	5.94	8.09	7.92	10.72	7.98
6	6.14	6.01	6.14	6.15	5.96	8.11	7.94	10.74	7.98
7	6.15	6.03	6.16	6.16	5.98	8.13	7.96	10.75	8.65
8	6.13	6.00	6.12	6.01	5.95	8.08	7.91	10.71	9.31
9	6.00	5.94	6.07	6.02	5.92	8.09	7.94	10.72	9.32
10	6.14	5.82	5.96	6.14	5.82	8.13	7.96	10.76	9.31
11	6.11	5.66	5.81	6.12	5.69	8.14	7.97	10.76	9.29
12	6.05	5.46	5.61	6.08	5.47	8.14	7.96	10.76	9.27
13	5.94	5.70	5.65	5.99	5.36	8.10	7.93	10.74	9.26
14	6.06	5.90	5.84	6.08	5.49	8.13	7.92	10.75	9.25
15	6.21	6.00	5.94	5.83	5.57	8.03	7.86	10.67	9.29

[16 rows x 46 columns]

3.2



3.3

1. Maximum Value for A1: 9.3
2. Minimum Value for A1: 7.43
3. Mean Value for A1: 8.428125
4. Variance for A1: 0.39726958333333356

5. Box Plot for A1: