Report of Assignment #2

Name:HUANG Pizhu

SID:12332298

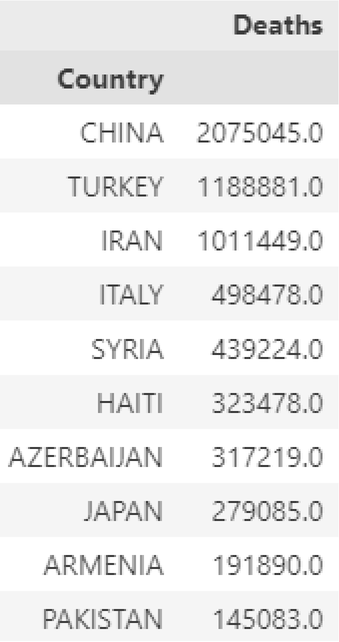
Date:2023/11/6

1. Significant earthquakes since 2150 B.C.

Read data file Sig_Eqs.tsv, and do some preprocess.Name it as df.

1.1

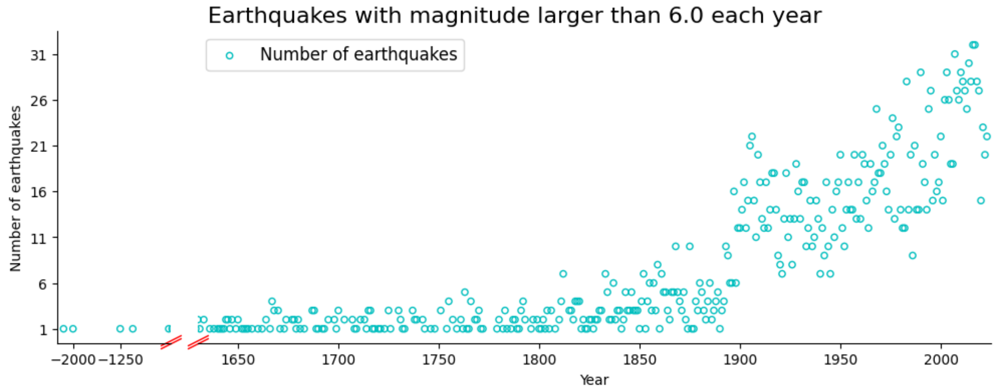
Compute the total number of deaths caused by earthquakes since 2150 B.C. in each country.
<u>The top ten countries along with the total number of deaths.</u>



	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

1.2

<u>The total number of earthquakes with magnitude larger than 6.0 is as follows.</u> ***(I learn the ticks in plotting Broken X-axis from https://zhuanlan.zhihu.com/p/205263612)***



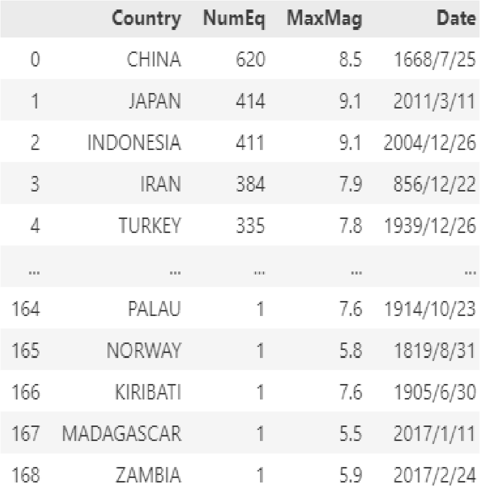
The Ring of Fire, located at the boundary between the Pacific Plate, EuThe Ring of Fire, located at the

boundary between the Pacific Plate, Eurasian Plate, Indian Plate, Antarctic Plate, and American Plate, is characterized by intense crustal activity and is the most extensive seismic zone in the world. This seismic zone hosts 80% of the world's earthquakes and is the primary location for most catastrophic earthquakes and large-scale (magnitude 8 or higher) global earthquakes.
Since the year 1600, the number of earthquakes with a magnitude greater than 6 has gradually increased, and this trend has further accelerated after 1850. This phenomenon can be attributed to the gradual activation of the corresponding tectonic plates within this seismic zone.

1.3

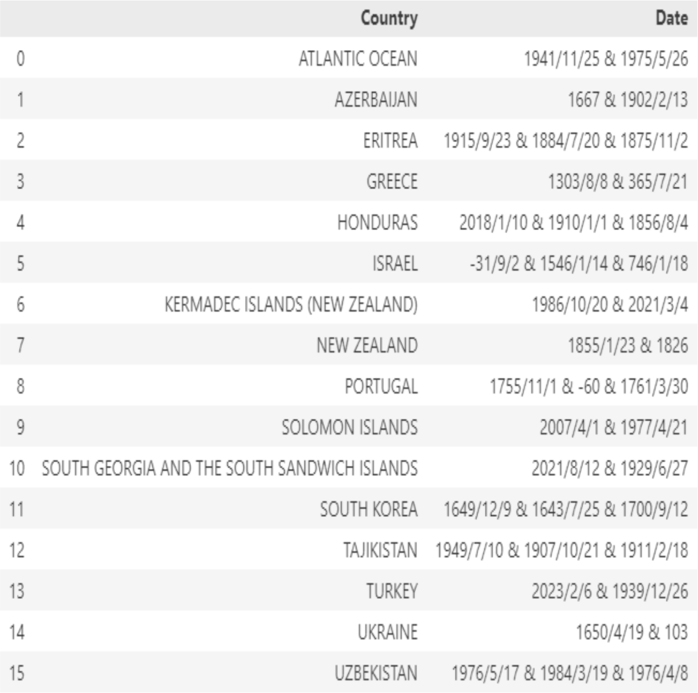
Thw function `CountEq_LargestEq()` returns both (1) the total number of earthquakes since 2150 B.C. in a given country AND (2) the date of the largest earthquake ever happened in this country. *(I get information of some useful function access online https://zhuanlan.zhihu.com/p/340770847, https://blog.csdn.net/PY0312/article/details/88956795 and https://zhuanlan.zhihu.com/p/370851569)*

Output of 146 country



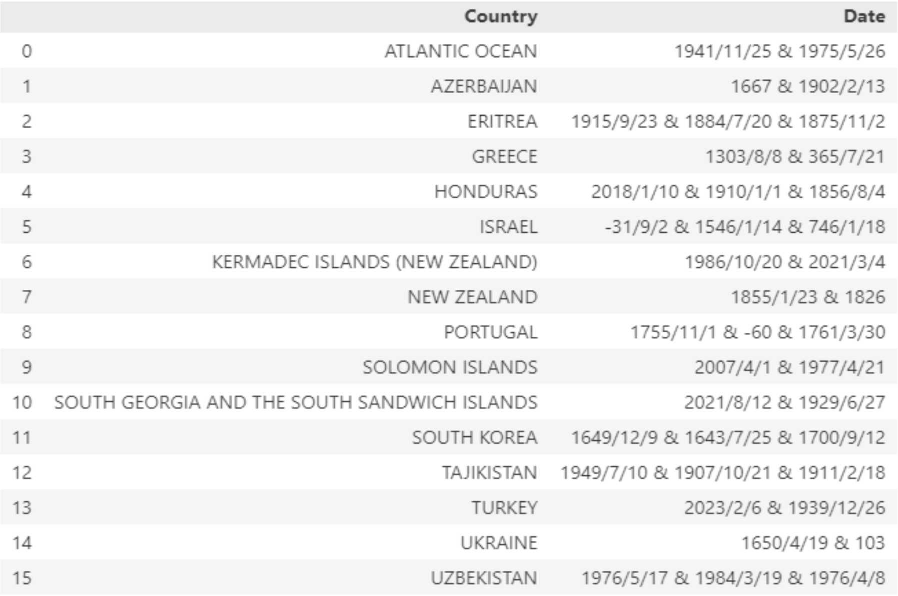
	Country	NumEq	MaxMag	Date
0	CHINA	620	8.5	1668/7/25
1	JAPAN	414	9.1	2011/3/11
2	INDONESIA	411	9.1	2004/12/26
3	IRAN	384	7.9	856/12/22
4	TURKEY	335	7.8	1939/12/26
...	...	...	...	...
164	PALAU	1	7.6	1914/10/23
165	NORWAY	1	5.8	1819/8/31
166	KIRIBATI	1	7.6	1905/6/30
167	MADAGASCAR	1	5.5	2017/1/11
168	ZAMBIA	1	5.9	2017/2/24

Duplication of country



	Country	Date
0	ATLANTIC OCEAN	1941/11/25 & 1975/5/26
1	AZERBAIJAN	1667 & 1902/2/13
2	ERITREA	1915/9/23 & 1884/7/20 & 1875/11/2
3	GREECE	1303/8/8 & 365/7/21
4	HONDURAS	2018/1/10 & 1910/1/1 & 1856/8/4
5	ISRAEL	-31/9/2 & 1546/1/14 & 746/1/18
6	KERMADEC ISLANDS (NEW ZEALAND)	1986/10/20 & 2021/3/4
7	NEW ZEALAND	1855/1/23 & 1826
8	PORTUGAL	1755/11/1 & -60 & 1761/3/30
9	SOLOMON ISLANDS	2007/4/1 & 1977/4/21
10	SOUTH GEORGIA AND THE SOUTH SANDWICH ISLANDS	2021/8/12 & 1929/6/27
11	SOUTH KOREA	1649/12/9 & 1643/7/25 & 1700/9/12
12	TAJIKISTAN	1949/7/10 & 1907/10/21 & 1911/2/18
13	TURKEY	2023/2/6 & 1939/12/26
14	UKRAINE	1650/4/19 & 103
15	UZBEKISTAN	1976/5/17 & 1984/3/19 & 1976/4/8

Handle the duplication and concat. The last Dataframe tdf is as follows.



	Country	Date
0	ATLANTIC OCEAN	1941/11/25 & 1975/5/26
1	AZERBAIJAN	1667 & 1902/2/13
2	ERITREA	1915/9/23 & 1884/7/20 & 1875/11/2
3	GREECE	1303/8/8 & 365/7/21
4	HONDURAS	2018/1/10 & 1910/1/1 & 1856/8/4
5	ISRAEL	-31/9/2 & 1546/1/14 & 746/1/18
6	KERMADEC ISLANDS (NEW ZEALAND)	1986/10/20 & 2021/3/4
7	NEW ZEALAND	1855/1/23 & 1826
8	PORTUGAL	1755/11/1 & -60 & 1761/3/30
9	SOLOMON ISLANDS	2007/4/1 & 1977/4/21
10	SOUTH GEORGIA AND THE SOUTH SANDWICH ISLANDS	2021/8/12 & 1929/6/27
11	SOUTH KOREA	1649/12/9 & 1643/7/25 & 1700/9/12
12	TAJIKISTAN	1949/7/10 & 1907/10/21 & 1911/2/18
13	TURKEY	2023/2/6 & 1939/12/26
14	UKRAINE	1650/4/19 & 103
15	UZBEKISTAN	1976/5/17 & 1984/3/19 & 1976/4/8

2.Wind speed in Shenzhen during the past 10 years

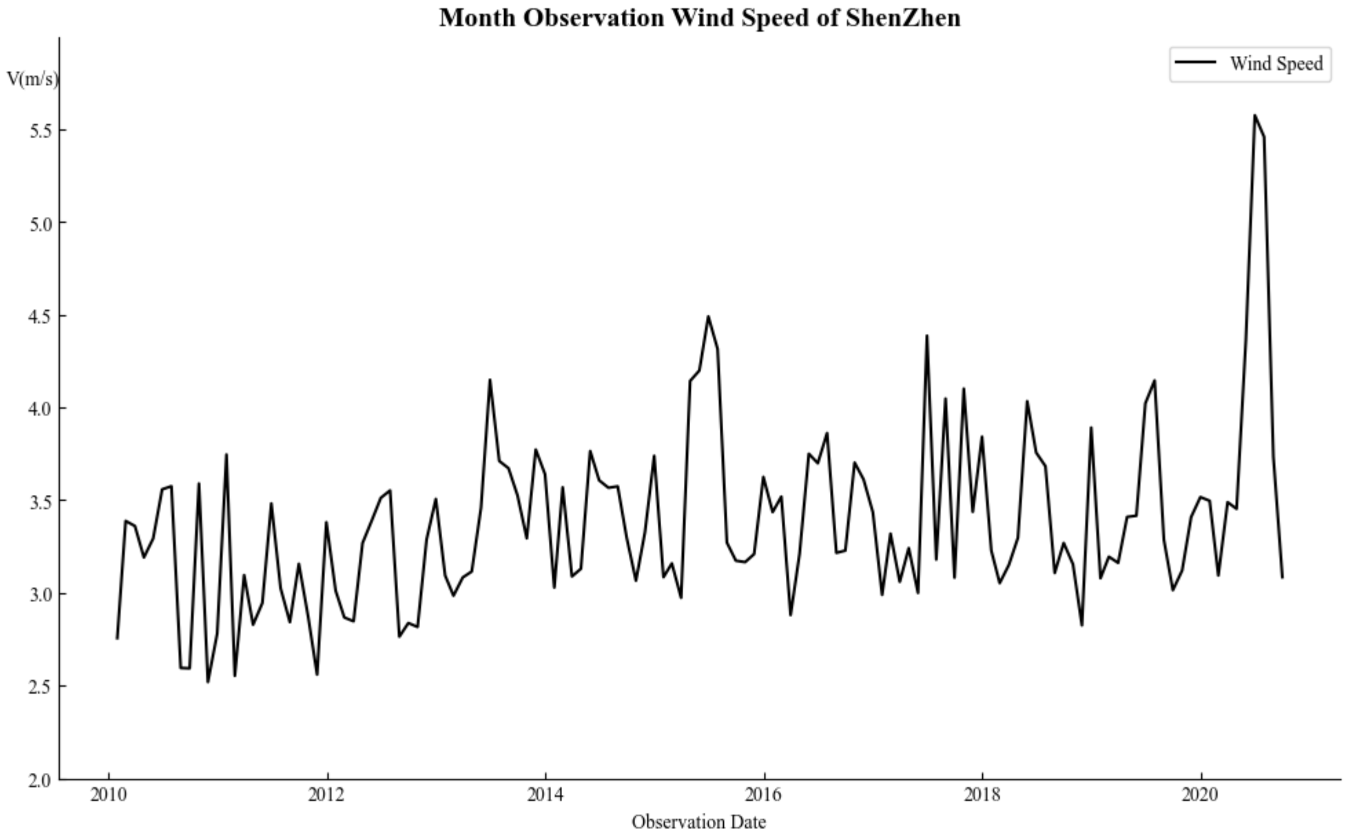
By reading the user guild, the information of wind is in the last column. the 4th part of WND is wind speed.Filter it and clean data with set missing values as average of the numbers before and after.By the way, it has a scale factor 10.

```
df['v'] = df['WND'].str.split(',').str[3].astype(int)
df.drop(columns=['WND'],inplace=True)
df = df[df['v'] != 9999]
df['v'] = df['v'].replace(9999,np.nan)
df['v']=df['v'].interpolate()/10
```

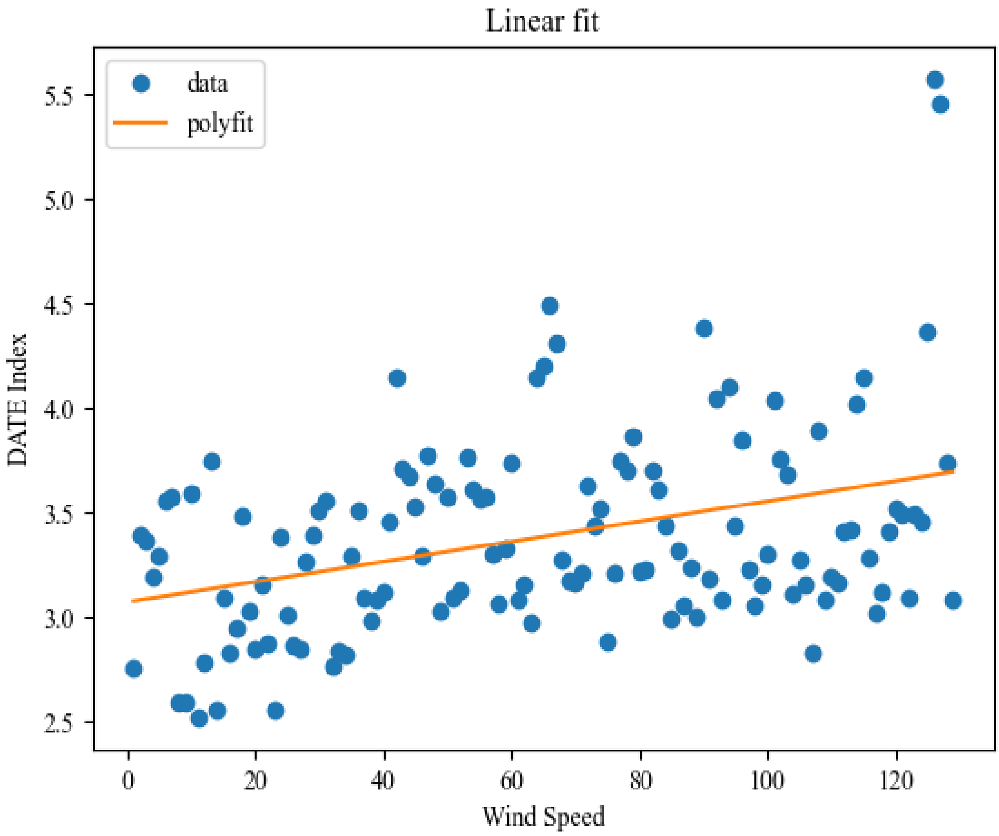
Resample the hourly data to monthly data

```
df['DATE'] = pd.to_datetime(df['DATE'])
df = df.set_index('DATE')
mw = df.resample('M').mean()
```

Plot Monthly Observation Wind Speed



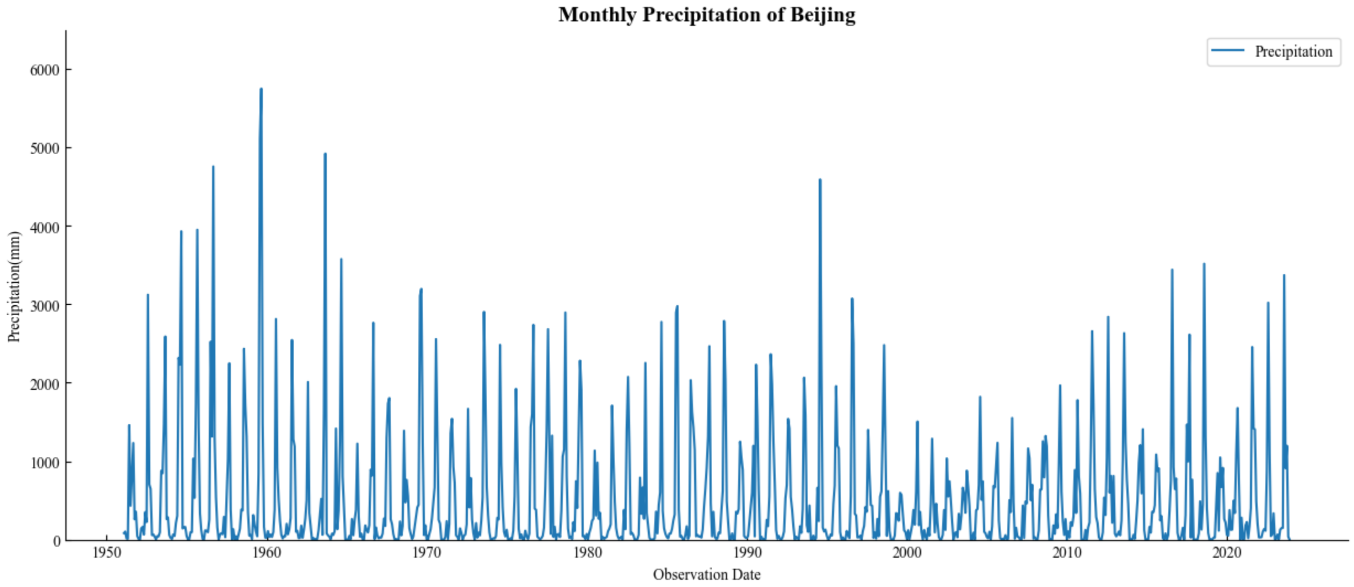
Linear fit. The trend of wind is ascending.*(I use the code from this website: https://blog.csdn.net/u013066730/article/details/103297380.)*



3.Explore a data set

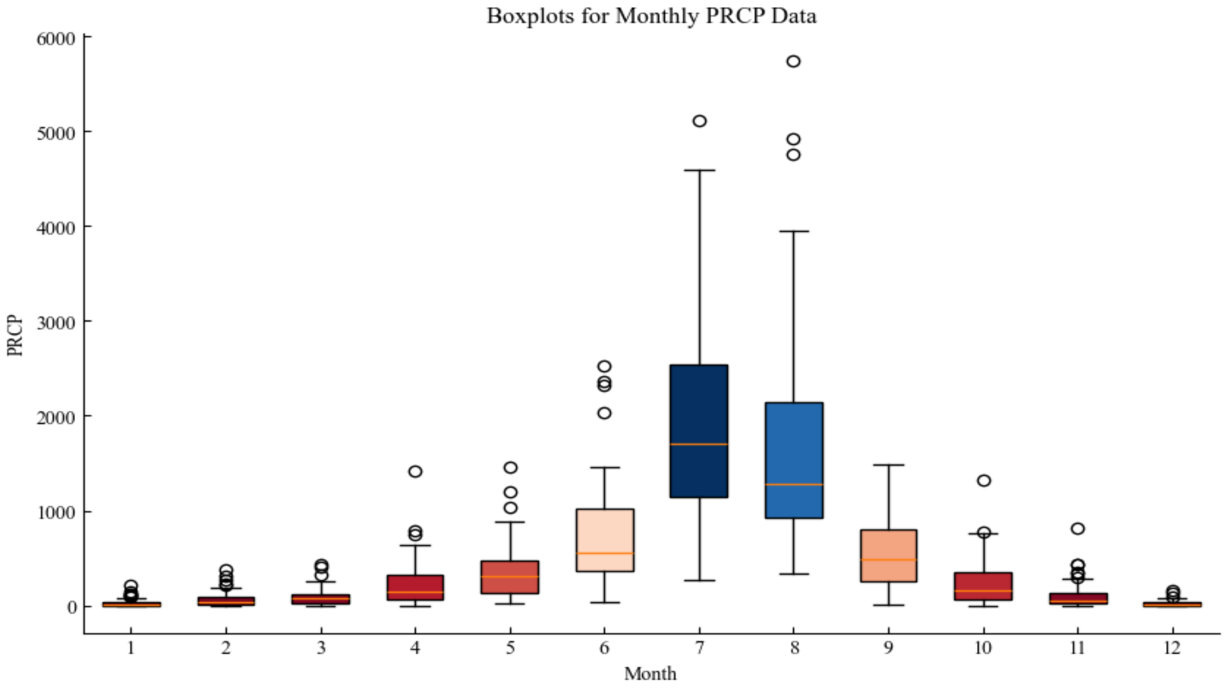
3.1 & 3.2

The Data set, precipitation and temperature of BeiJing, is comes from NCEI.For the precipitation the NaN are subtituded as value 0. And resample hourly data as monthly data and show.



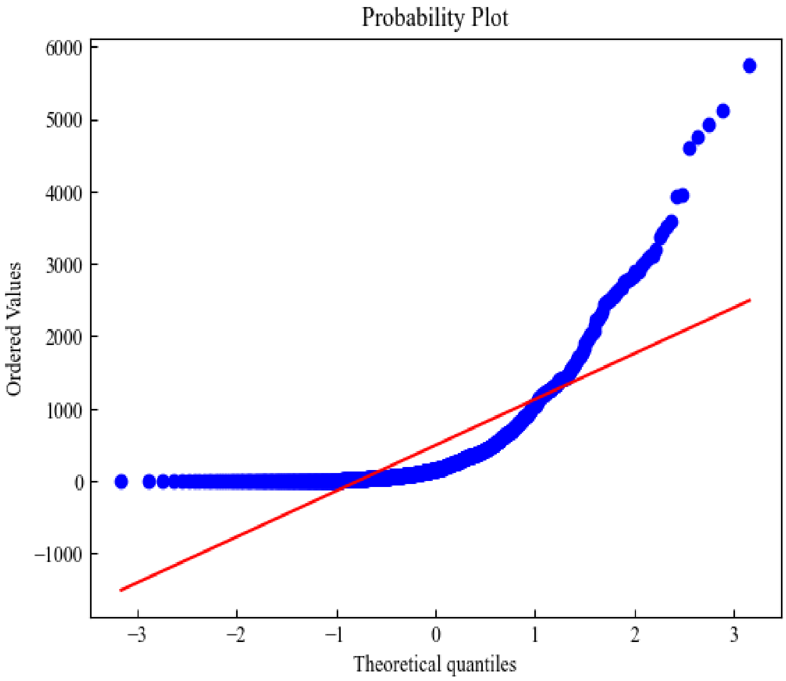
3.3

Use function `groupby()` to calculate the mean, variance, standard deviation,min,max of 12 month. The description is in PS2.ipynb.
A boxplot is used to describe the data. Every month has anormal point.The summer has most precipitation which also max varince in the past 70 year.



A probability plot is used to Normal Test. p-value is equal to 4.83, so that data is not normally distributed. ***(I learn konwledge of normally distribution form https://www.biaodianfu.com/python-***

normal-distribution-test.html)


Probability Plot
Ordered Values
Theoretical quantiles