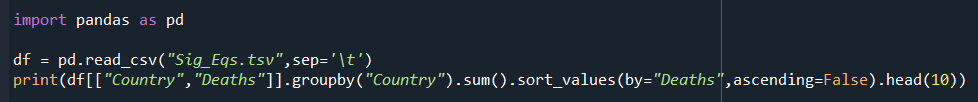
PS2 Report

Name:左小幸 SID:12132243

**PS2\_1: Significant earthquakes since 2150 B.C.**

* 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.

Code:



Result:

文本

描述已自动生成

* 1. **[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?

Code:

文本

描述已自动生成

Result:

图表, 直方图

描述已自动生成

The trend I have observed is that the number of earthquakes greater than magnitude 6 globally has increased significantly over the last 500 years. I think the reason for this trend may not be that the number of high-intensity earthquakes has really increased, but that with the development of technology, people are observing and recording a lot of earthquakes that were not observed and recorded before.

* 1. **[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.

Code:

屏幕的截图

描述已自动生成

Result:

日历

描述已自动生成

The way I think about this problem is. Store **(1)**(the total number**)** and **(2)**(the date of the maximum earthquake) in the same Data frame and print it.

**PS2:** **Wind speed in Shenzhen during the past 10 years**

**[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?

**How to filter data:**

图形用户界面, 文本, 应用程序

描述已自动生成

I filtered the data according to the description of wind speed data quality on the user guide. I selected the data which **speed quality code** = 1 as valid data.

Code:

文本

描述已自动生成

Result:

图表

描述已自动生成

I found that the monthly average wind speed of Bao An International Airport has been fluctuating for nearly ten years, and there seems to be no obvious trend. If one had to give a trend, it would be a slight increase in average wind speeds over the decade.

**PS3: Explore a data set**

Browse the CASEarth, NOAA Land-Based Datasets and Products, 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 chose a piece of data from my research, which contains information on more than 6,000 DAMS around the world. GRanD\_dams\_v1\_3.xlsx

I wanted to count DAMS built since the 19th century, so I ran the following filter



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

I plotted the number of DAMS built each year

文本

描述已自动生成

图形用户界面, 图表, 折线图

描述已自动生成

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

1. #Top 10 countries with the most DAMS

表格

描述已自动生成

2. #The name and country of the longest dam

图片包含 图示

描述已自动生成

3. #The name and country of the deepest dam

图形用户界面, 文本

中度可信度描述已自动生成

4. #The main use of these dams

图片包含 表格

描述已自动生成

5. #The highest dam

图片包含 图示

描述已自动生成

**Reference:**

1. [Getting started — pandas 1.3.4 documentation (pydata.org)](https://pandas.pydata.org/pandas-docs/stable/getting_started/index.html#getting-started)helped me solve many problems related to PANDAS. **In in problem set 1 2 and 3.**

2. [时间序列与日期用法 | Pandas 中文 (pypandas.cn)](https://www.pypandas.cn/docs/user_guide/timeseries.html#%E7%BA%B5%E8%A7%88) Help me understand how to use the datetime method in Pandas, **in problem set2**.