

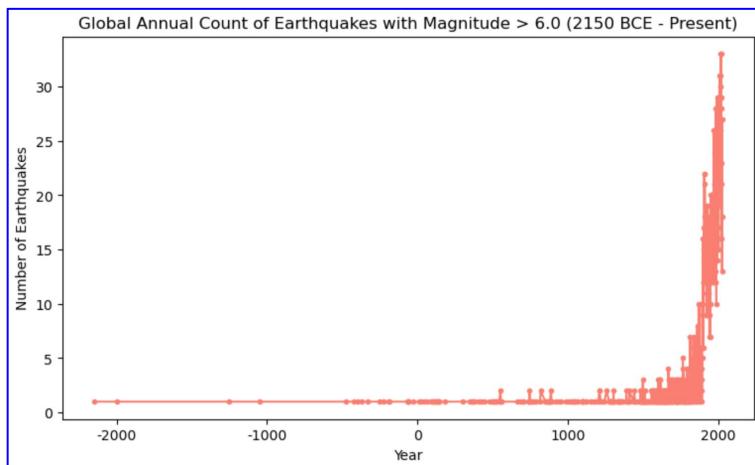
## 1. 公元前 2150 年以来的重大地震

重大地震数据库包含了公元前 2150 年至今破坏性地震的相关信息。在左上角选择所有列，点击“下载 TSV 文件”按钮，以.tsv 格式下载完整的重大地震数据集。点击变量名可查看更多信息。将该文件（例如：earthquakes-2025-10-29\_15-11-32\_+0800.tsv）读取为一个对象，并命名为 Sig\_Eqs。

1.1 (5 分) 计算公元前 2150 年以来每个国家因地震造成的死亡总数，然后输出死亡人数排名前十的国家及其死亡总数。

	Country	Deaths
28	CHINA	2139210.0
142	TURKEY	1199742.0
65	IRAN	1014453.0
69	ITALY	498219.0
132	SYRIA	419226.0
58	HAITI	323484.0
10	AZERBAIJAN	319251.0
71	JAPAN	242445.0
6	ARMENIA	191890.0
103	PAKISTAN	145083.0

1.2 (10 分) 计算全球每年震级大于 6.0（以“Mag”列为震级指标）的地震总数，绘制时间序列图。你是否观察到任何趋势？请解释原因。



通过对全球每年震级大于 6.0 地震数据的分析，观察到阶段性阶梯式增长趋势，该趋势反映人类监测与记录能力的进步：公元前 2150 年 - 公元 1000 年的古代阶段，年均地震记录仅 1.05 次，且年份记录稀疏（仅 84 个年份有数据），因无专业监测设备，大量偏远地区地震被遗漏；公元 1001 年 - 1900 年的近代阶段，年均记录提升至 2.19 次，记录年份增至 343 个，但仍受限于信息传播与记录技术，数据连续性有限；1900 年至今的现代阶段，年均记录激增至 17.90 次，125 个年份均有完整数据，尤其 1960 年后全球地震监测网络建立，可实时捕捉全球强震，数据进入 15-30 次 / 年的稳定区间，且 1900 年后存在微弱逐年增长（年均增 0.0999 次，统计显著），这源于监测精度提升与历史数据回溯补充。所以，“地震数量增长”是人类对地震认知能力提升的结果，地球板块运动活跃度在数千年尺度内相对稳定，1960 年后数据才具备反映地震活动周期性波动的参考价值。

1.3 (10 分) 编写一个函数 CountEq\_LargestEq，返回 (1) 某个国家自公元前 2150 年以来的地震总数；(2) 该国发生过的震级最大地震的日期。将该函数应用于数据集中的所有国家，按地震总数降序排列并报告结果。

```
cc = 'IRELAND'
```

```
print(f'{cc} 的地震总数是 {CountEq_LargestEq(cc)[0]}, 最大震级地震日期是 {CountEq_LargestEq(cc)[1]}。')
```

```
IRELAND 的地震总数是 1, 最大震级地震日期是 1490-00-00。
```

	Country	Earthquake_Count	Largest_Eq_Date
0	CHINA	623	1668-07-25
1	JAPAN	424	2011-03-11
2	INDONESIA	421	2004-12-26
3	IRAN	388	0856-12-22
4	TURKEY	358	1939-12-26
...	...	...	...
153	CENTRAL AFRICAN REPUBLIC	1	1921-09-16
154	GRENADA	1	1822-12-01
155	IRELAND	1	1490-00-00
156	GABON	1	1974-09-23
157	COMOROS	1	2018-05-15

158 rows × 3 columns

## 2. 2010-2020 年深圳风速数据

本习题将分析深圳过去 10 年的风速变化情况，使用宝安国际机场记录的每小时气象数据。该数据集来自美国国家海洋和大气管理局 (NOAA) 综合地面数据集。下载文件 2281305.zip (其中 2281305 为站点 ID)，解压后会得到一个名为 2281305.csv 的文件，将其保存到你的工作目录中。

查阅综合用户指南的第 8-9 页 (位置 61-70)，了解风速数据的详细格式。在报告中说明你如何对数据进行筛选。

(10 分) 绘制月平均风速随观测时间变化的图表。2010-2020 年的月平均风速是否存在趋势？请说明。

数据解析步骤：

提取 WND 列，解析 WND 字符串：

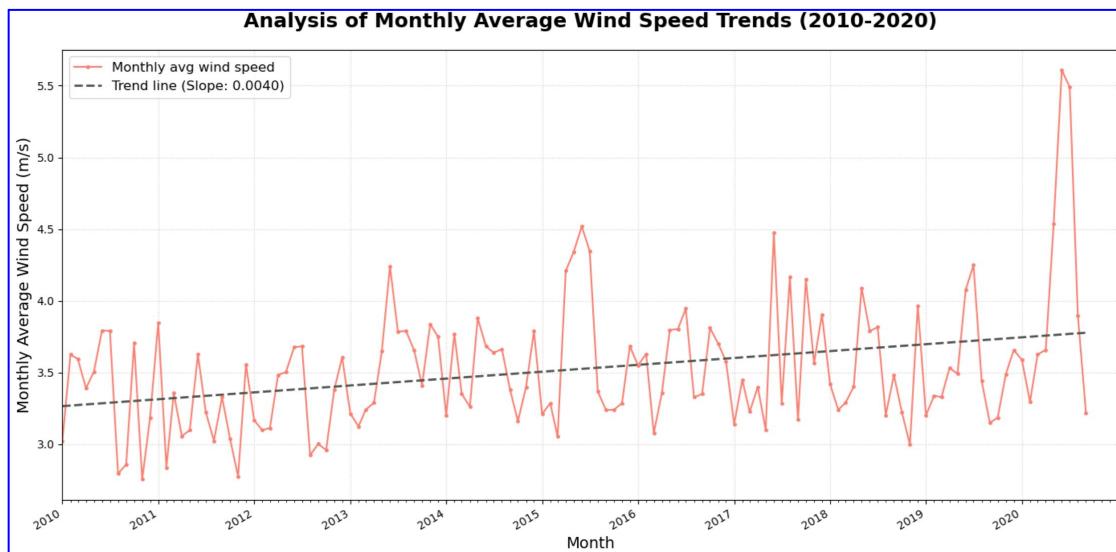
风向角度：从第 0 个字符开始，取 3 个字符 (WND[0:3])。方向质量码：第 3 个字符 (WND[3])。类型码：第 4 个字符 (WND[4])。风速：从第 5 个字符开始，取 4 个字符 (WND[5:9])。风速质量码：第 9 个字符 (WND[9])。

将解析出的字符串转换为数值类型。

过滤无效数据。

将风速值除以 10，得到以米 / 秒 (m/s) 为单位的实际风速。

提取时间信息：从 DATE 列中提取出年份和月份。



2010-2020 年的月平均风速存在显著的短期波动（不同季节的风速差异较大），但从长期趋势来看，年平均风速整体呈现上升态势。这种“波动大但整体上升”的特征，可能源于短期随机因素（如季节性气流、极端天气事件）导致的月度数据起伏。

## 3. 探索一个数据集

浏览 CASEarth (国家地球系统科学数据中心)、美国国家环境信息中心 (NCEI) 或全球大气气体实验先进网络 (AGAGE) 的官网，搜索并下载一个你感兴趣的数据集。也可以使用你所在小组的数据集，但该数据集需为 csv、XLS 或 XLSX 格式，且包含时间信息。

3.1 (5 分) 加载 csv、XLS 或 XLSX 文件，清理存在缺失值或质量异常的数据点。

选取北极点数据集分析，挑出高质量数据列 (年份 下雪天数，降水量，雪深度，平均气温，最高气温，最低气温)，删掉有缺失值的行。

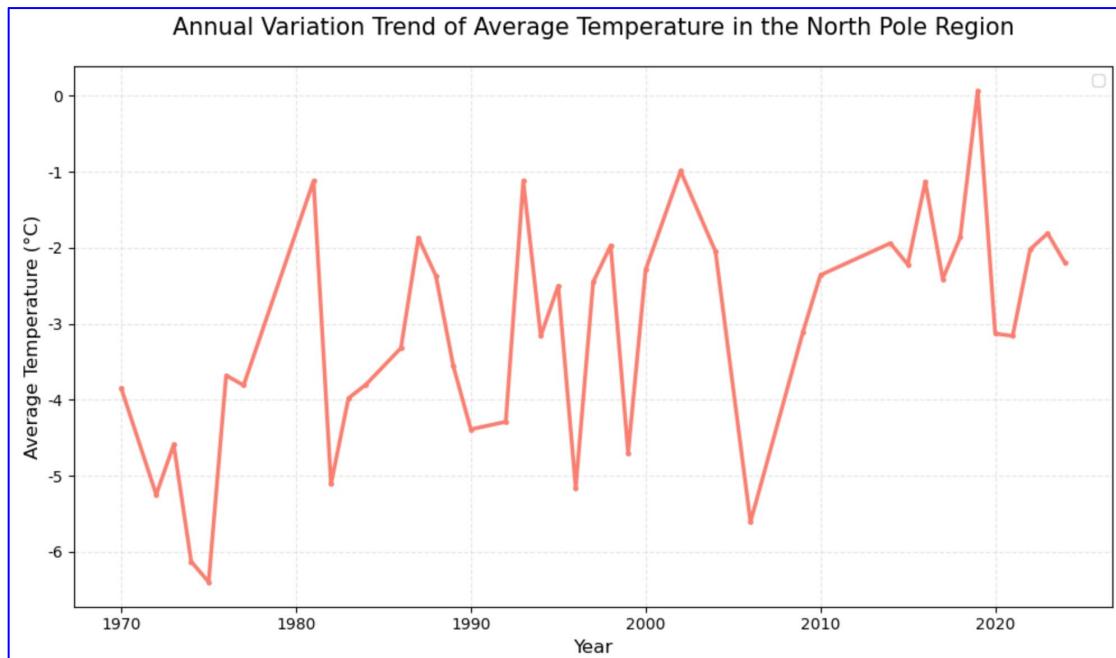
```
NORTH_POLE = NORTH_POLE[['DATE', 'DSNW', 'PRCP', 'SNOW', 'TAVG', 'TMAX', 'TMIN']].rename(columns={'DSNW': 'Snow_Days', 'PRCP': 'Precipitation', 'SNOW': 'Snow_Depth', 'TAVG': 'Average_Temperature', 'TMAX': 'Maximum_Temperature', 'TMIN': 'Minimum_Temperature'})
```

```
NORTH_POLE = NORTH_POLE.dropna(axis=0, how='any')
```

	DATE	Snow_Days	Precipitation	Snow_Depth	Average_Temperature	Maximum_Temperature	Minimum_Temperature
0	1970	27.0	321.2	2475.0	-3.85	2.53	-10.23
2	1972	20.0	241.1	1539.0	-5.25	1.25	-11.75
3	1973	17.0	273.1	862.0	-4.59	1.84	-11.02
4	1974	27.0	247.1	2099.0	-6.13	0.47	-12.73
5	1975	19.0	258.4	1351.0	-6.40	0.29	-13.09

3.2 (5 分) 绘制某一变量的时间序列图。

北极平均温度的时间序列图



3.3 (5 分) 对该变量进行至少 5 项简单统计检验，并报告分析结果。

变量：  
x = NORTH\_POLE['YEAR']  
y = NORTH\_POLE['Average\_Temperature']

### 1. 趋势性检验

斜率：0.0512,  $R^2$ : 0.3193, P值: 0.0001

结论：温度随时间显著上升

### 2. 正态性检验

统计量：0.9730, P值：0.4304

结论：数据符合正态分布

### 3. 相关性检验

Pearson相关系数：0.5650, P值：0.0001

Spearman相关系数：0.5695, P值：0.0001

结论：存在显著线性相关

### 4. 异常值检验

异常值数量：0

### 5. 周期性检验

最强自相关滞后：1 年

主要周期：38.00 年

结论：可能存在 38.00 年的周期性波动