

## Lab Report #2

### 1. Significant earthquakes since 2150 B.C.

#### 1.1. 实验目的

通过对重大地震数据库(Significant Earthquake Database)的分析，掌握地震数据的读取、筛选、分组统计与时间序列可视化等基本数据分析方法，并通过函数化编程实现对不同国家地震特征的自动化统计。

#### 1.2. 实验原理与思路

**数据读取:** 使用 `pandas.read_csv()` 读取制表符分隔 (TSV) 文件，得到完整地震记录表。

**统计分析:**

- 利用 `groupby()` 和 `sum()` 计算各国地震造成的死亡总数，并输出前十名。
- 根据震级 (Mag) 和年份 (Year) 筛选出震级大于 6.0 的记录，统计每年的发生次数并绘制时间序列，并对该趋势进行分析。

**函数编写:**

- 编写 `CountEq_LargestEq()` 函数，实现输入国家名后输出该国地震总数及最大震级地震日期。
- 对所有国家循环调用该函数并按地震次数降序排序。

#### 1.3. 实验结果

**注:** 本次实验涉及到的地震死亡人数是基于表格列“Deaths”进行计算。

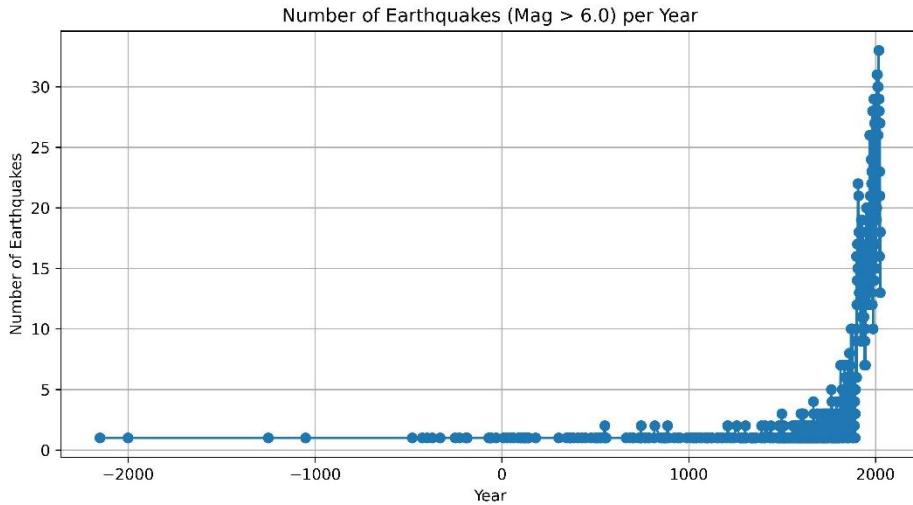
a) 各国地震造成的死亡总数前十名为：

Top 10 countries by total deaths from earthquakes since 2150 B.C.:

Country

CHINA	2139210.0
TURKEY	1199742.0
IRAN	1014453.0
ITALY	498219.0
SYRIA	419226.0
HAITI	323484.0
AZERBAIJAN	319251.0
JAPAN	242445.0
ARMENIA	191890.0
PAKISTAN	145083.0

b) 统计每年震级大于 6.0 的地震次数如图所示：



20世纪后地震记录数量显著上升，反映了观测与记录体系的完善，而非真实地震活动增强。

c) 各国地震统计示例（前十行）

	Total_Earthquakes	Largest_Eq_Date
CHINA	623.0	1668-07-25
JAPAN	424.0	2011-03-11
INDONESIA	421.0	2004-12-26
IRAN	388.0	0856-12-22
TURKEY	358.0	1939-12-26
ITALY	333.0	1915-01-13
GREECE	289.0	0365-07-21
USA	280.0	1964-03-28
PHILIPPINES	230.0	1897-09-21
MEXICO	214.0	1787-03-28

## 2. Wind speed in Shenzhen from 2010 to 2020

### 2.1. 实验目的

利用 NOAA Integrated Surface Dataset (ISD) 提供的深圳宝安国际机场逐小时气象观测数据，提取并分析 2010–2020 年的风速变化特征，绘制月平均风速随时间变化的趋势，并通过线性回归定量检验是否存在显著趋势。

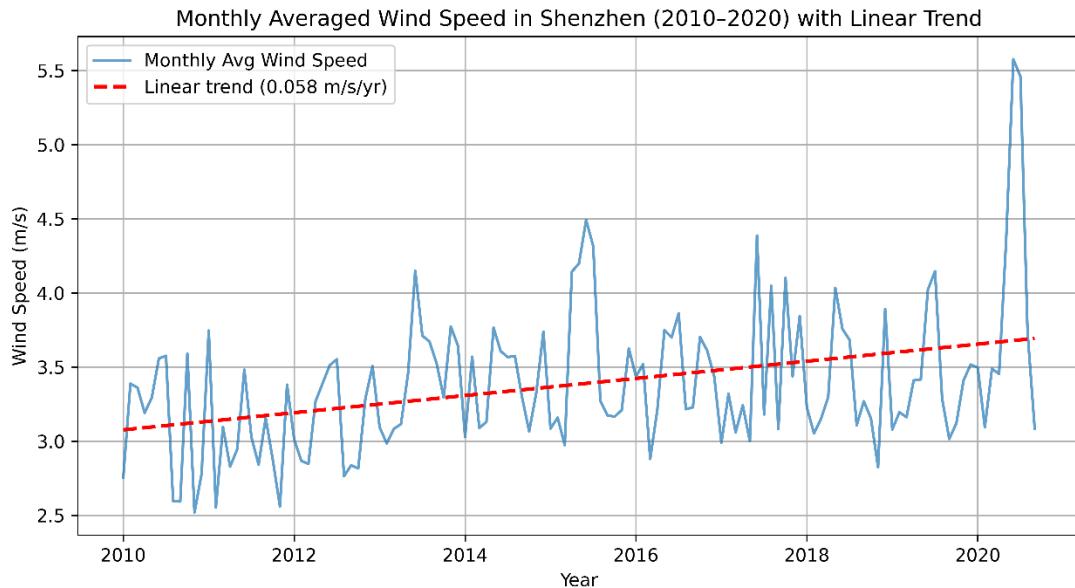
### 2.2. 实验原理与思路

本实验使用 NOAA ISD 的地面观测数据（站点编号 2281305）。根据官方用户手册（第 8–9 页，POS 61–70），风速信息位于原始记录的 61–65 位，对应的质量控制码位于第 66 位。在导出的 .csv 文件中，风速数据集中于 WND 字段。拆分 WND 字段，提取风速子字段（第 4 项），并除以 10 得到真实风速 (m/s)。筛除风速为空值且质量码不为 1 的记录，确

保数据可靠。将时间字段 DATE 转换为 datetime 类型，并筛选 2010–2020 年间的观测，最后以年月为分组单位计算月平均风速。

### 2.3. 实验结果

2010-2020 年宝安机场月均风速时序变化如下图所示：



统计参数为：

```
===== Quantitative Trend Analysis =====
Slope: 0.0578 m/s per year
Intercept: -113.13
R-squared: 0.135
P-value: 1.7854e-05
Standard Error: 0.0130
→ The trend is statistically significant (p < 0.05)
```

上述数据说明，过去十年深圳宝安机场地区月平均风速总体显著上升。

## 3. Explore a data set

### 3.1. 实验目的

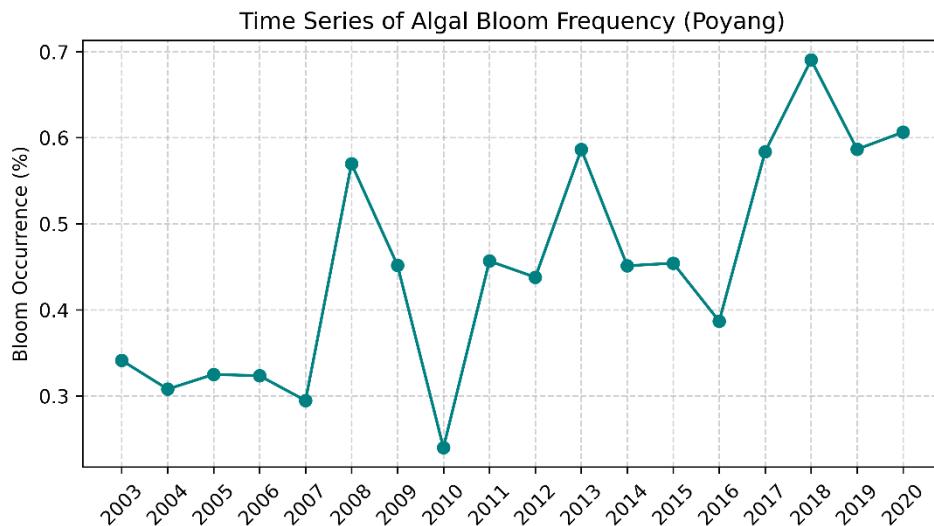
本实验旨在通过对我国湖泊 2003–2020 年藻华发生频率数据的探索性分析，掌握环境时序数据的读取、清洗、可视化与统计分析方法。

### 3.2. 实验原理与思路

- (1) 数据读取与清洗
- (2) 绘制单个湖泊藻华发生频率时间序列图
- (3) 统计特征分析
- (4) 趋势分析

### 3.3. 实验结果

- 1) 面积最大的湖泊（鄱阳湖）在 2003–2020 年间藻华发生频率总体呈上升趋势，部分年份存在明显高值，说明湖泊藻华问题具有显著的年际波动性。



- 2) 全国范围的藻华发生率均值约为 0.99%，标准差 1.33%，最小值接近 0%，最大值 11.84%，中位数为 0.47%，表明不同湖泊之间的藻华发生程度差异较大。

#### 3) 趋势分析结果

线性回归结果显示，2003 – 2020 年全国湖泊藻华发生率呈显著上升趋势。趋势斜率约为 +0.028%/年，决定系数  $R^2 = 0.694$ ，对应的年均变化率约为 2.79%/年。说明过去近二十年间，中国湖泊藻华总体呈加剧趋势，反映出富营养化程度的持续上升。

