

Result1

1.1 前 10 个国家的地震死亡人数

```
=====
CHINA: 2,106,524 人
TURKEY: 1,008,863 人
IRAN: 761,654 人
SYRIA: 487,726 人
ITALY: 423,280 人
HAITI: 323,782 人
JAPAN: 319,443 人
AZERBAIJAN: 319,251 人
INDONESIA: 282,838 人
ARMENIA: 189,000 人
```

1.2 地震时间序列分析

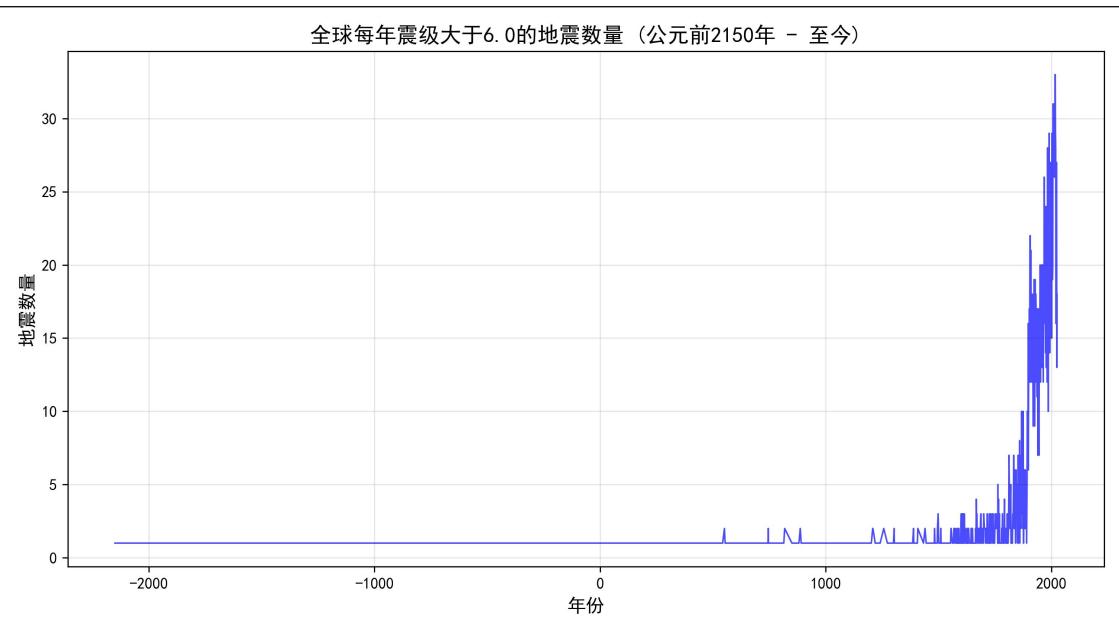
```
=====
数据时间范围: -2150 - 2025
```

数据列表已保存至:

C:\Users\Administrator\Desktop\all\homework\code\yearly_earthquake_counts.csv

最近 20 年的地震数量示例:

```
-----
年份: 2006, 地震数量: 20
年份: 2007, 地震数量: 31
年份: 2008, 地震数量: 27
年份: 2009, 地震数量: 27
年份: 2010, 地震数量: 31
年份: 2011, 地震数量: 30
年份: 2012, 地震数量: 29
年份: 2013, 地震数量: 26
年份: 2014, 地震数量: 30
年份: 2015, 地震数量: 28
年份: 2016, 地震数量: 33
年份: 2017, 地震数量: 33
年份: 2018, 地震数量: 29
年份: 2019, 地震数量: 28
年份: 2020, 地震数量: 16
年份: 2021, 地震数量: 23
年份: 2022, 地震数量: 21
年份: 2023, 地震数量: 27
年份: 2024, 地震数量: 13
年份: 2025, 地震数量: 18
```



趋势情况：近代地震记录数量明显增加，这主要是由于监测技术的进步和记录体系的完善吧，不一定是真实发生频率升高。

1.3 各国地震统计（按地震总数降序排列 - 前 20 名）

国家	地震总数	最大地震日期	最大震级
<hr/>			
CHINA	623	1668-07-25	8.5
JAPAN	424	2011-03-11	9.1
INDONESIA	421	2004-12-26	9.1
IRAN	388	856-12-22	7.9
TURKEY	358	1939-12-26	7.8
ITALY	333	1915-01-13	7.5
GREECE	289	365-07-21	8.0
USA	280	1964-03-28	9.2
PHILIPPINES	230	1897-09-21	8.7
MEXICO	214	1787-03-28	8.6
CHILE	200	1960-05-22	9.5
PERU	194	1716-02-06	8.8
RUSSIA	158	1952-11-04	9.0
PAPUA NEW GUINEA	107	1919-05-06	8.2
INDIA	102	1950-08-15	8.6
TAIWAN	101	1920-06-05	8.2
COLOMBIA	82	1826-06-18	8.2
NEW ZEALAND	72	1826-01-01	8.0
ECUADOR	69	1906-01-31	8.8
AFGHANISTAN	68	1909-07-07	8.1

完整结果已保存至:

C:\Users\Administrator\Desktop\all\homework\code\country_earthquake_statistics.csv

Result2

POS 61-70 核心质量筛选条件总结

1. 质量代码筛选

风向质量码 (POS 64) 和风速质量码 (POS 70) 的合格标准相同:

0 = 通过粗限检查

1 = 通过所有质量控制检查

4 = 通过粗限检查, 数据源自 NCEI 数据源

5 = 通过所有质量控制检查, 数据源自 NCEI 数据源

9 = 如果元素存在则通过粗限检查

应排除的质量码:

2 = 可疑

3 = 错误

6 = 可疑 (NCEI 源)

7 = 错误 (NCEI 源)

2. 缺失值标识

风向角度 = 999 表示缺失

风速速率 = 9999 表示缺失

风类型码 = 9 表示缺失

3. 特殊情况的处理逻辑

关于风向 999 的特殊规则:

如果风类型码 = V (Variable), 那么风向=999 是合理的, 表示风向可变

如果风类型码 ≠ V, 那么风向=999 应视为缺失数据

关于静风的特殊规则:

如果风类型码 = C (Calm) 且风速 = 0000, 这是有效的静风记录

如果风类型码 = 9 且风速 = 0000, 这也表示静风 (根据 NOTE 说明)

4. 数据合理性范围

风向角度: 001-360 度 (有效范围)

风速速率: 0000-0900 (换算后为 0-90 m/s)

筛选流程

1: 质量码检查 - 只保留质量码为 0, 1, 4, 5, 9 的数据

2: 风类型码检查 - 处理类型码为 9 的特殊情况 (如果同时风速为 0000 则保留)

3: 风向合理性检查 - 正确处理 C/V 类型下风向 999 的合理性

4: 风速缺失值检查 - 排除风速为 9999 的记录

5: 数据转换和范围验证 - 转换风速并验证合理性

结果

数据读取完成, 共 111984 条记录

有效风速数据记录: 111346/111984 (99.43%)

计算完成, 共 129 个月的月平均数据

时间范围: 2010-01-01 00:00:00 到 2020-09-01 00:00:00

==== 统计摘要 ===

总体月平均风速: 3.385 m/s

最大月平均风速: 5.576 m/s (出现在 2020-06)

最小月平均风速: 2.520 m/s (出现在 2010-11)

数据时间跨度: 129 个月

==== 趋势分析 ===

线性趋势斜率: 0.004818 m/s/月

10 年变化量: 0.5782 m/s

相关系数 R: 0.3680

P 值: 0.000018

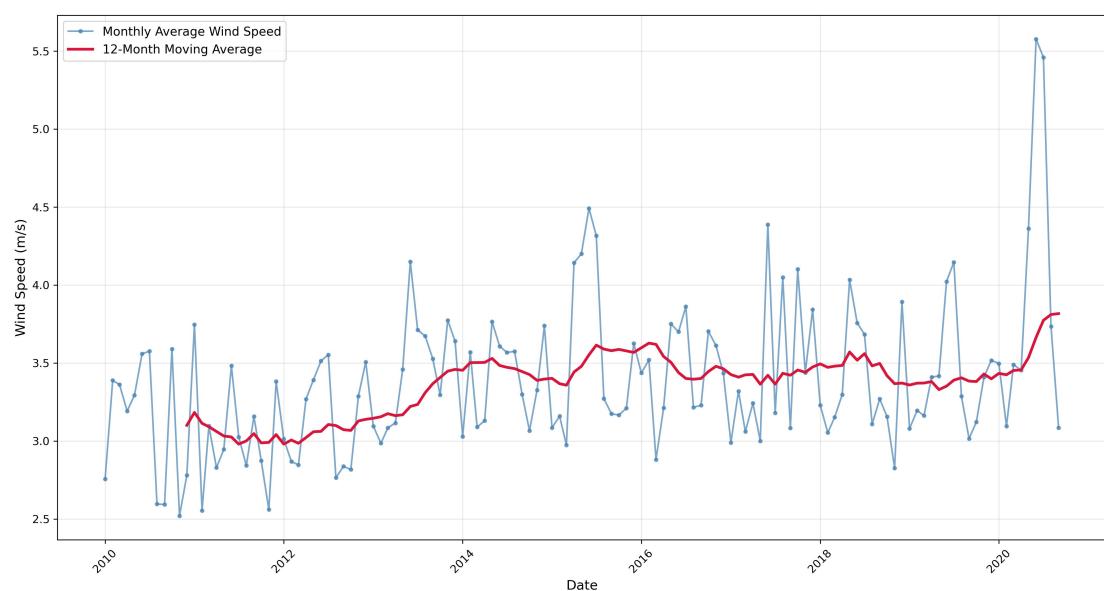
趋势显著性: 在 95% 置信水平下具有统计显著性 (上升趋势)

图表已保存到: E:\homework_files\wind_speed_trend.png

月平均数据已保存到:

E:\homework_files\monthly_wind_speed_shenzhen_2010_2020.csv

Monthly Average Wind Speed Trend at Shenzhen BaoAn International Airport (2010-2020)



Result3

下载数据及主要工作叙述:

数据信息:

US Geological Survey

retrieved: 2022-11-22 01:32:22 EST

URL: <https://nwis.waterdata.usgs.gov/nwis/measurements>

#

The Site File stores location and general information about groundwater,
surface water, and meteorological sites

for sites in USA.

示例数据见文件: 1.xlsx

3.1 数据加载与清理

首先对 USGS 水文监测数据进行全面的质量检查和预处理。系统读取原始数据文件后，详细分析了数据的整体结构和完整性，包括记录总数、字段信息和缺失值分布。针对数据质量问题，执行了严格的清理流程：移除了关键字段存在缺失值的记录，确保分析数据的完整性；同时采用统计方法识别并标记了可能的异常数值，为后续分析提供了高质量的数据基础。整个过程还生成了详细的数据质量评估报告。

3.2 时间序列可视化

基于清理后的数据，构建了多层次的时间序列可视化分析。主要包含两个维度的图表展示：上层绘制了详细的日度流量变化曲线，并叠加移动平均线以凸显长期趋势 patterns；下层则提供了年度聚合视图，通过柱状图展示各年份的平均流量水平，清晰呈现水文变化的年际特征。这种双层次的可视化设计既能捕捉短期波动细节，又能展现长期演变规律，为理解水文动态提供了直观的视觉依据。

3.3 统计检验与分析

对水文数据进行了系统的统计特性探索，共完成了五项核心的统计检验。检验内容包括：数据分布的正态性评估、时间趋势的方向和显著性检测、变异程度的量化分析、序列自相关特征的考察以及分布形态的偏度和峰度测量。除了这些专项检验外，还计算了完整的描述性统计量，包括集中趋势、离散程度和分布形态的各类指标，全面揭示了水文数据的统计规律和特征。

结果

3.1 数据加载与清理

原始数据形状: (33959, 3)

数据列名: ['station_id', 'date', 'value']

数据基本信息：

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 33959 entries, 0 to 33958
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   station_id  33959 non-null   int64  
 1   date        33959 non-null   datetime64[ns]
 2   value       33959 non-null   float64 
dtypes: datetime64[ns](1), float64(1), int64(1)
memory usage: 796.0 KB
None
```

缺失值统计：

Series([], dtype: int64)

移除缺失值后数据行数: 33959 (移除 0 行)

检测到异常值数量: 3461

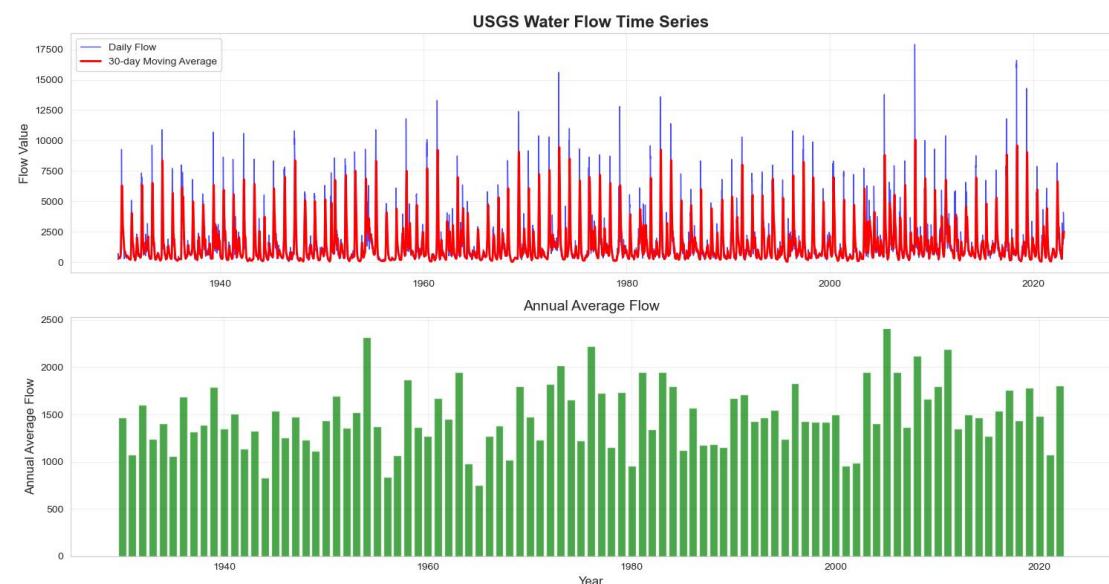
数据质量总结:

- 总记录数: 33959
 - 完整记录数: 33959
 - 数值范围: 33.50 – 17900.00
-

3.2 时间序列分析

使用时间列: date

使用数值列: value



时间序列图已保存为 'water_data_time_series.png'

3.3 统计检验与分析

进行 5 项统计检验:

1. 正态性检验: p 值 = 0.000000 - 数据不服从正态分布
2. 趋势检验: $\tau = 0.0662$, p 值 = 0.000000 - 显著上升趋势
3. 变异系数: 1.2384 - 高变异性
4. 一阶自相关: 0.9948 - 显著自相关
5. 分布形状: 偏度 = 2.6923, 峰度 = 8.8054

基本统计量:

- 均值: 1480.26
- 标准差: 1833.15
- 中位数: 785.00
- 偏度: 2.6923
- 峰度: 8.8054

- 最小值: 33.50
- 最大值: 17900.00
- 四分位距: 1285.00

=====

数据分析报告总结

=====

USGS 水文数据分析报告

数据来源: C:\Users\Administrator\Desktop\all\homework\code\1.xlsx

分析时间: 2025-11-12 14:25:45

1. 数据概况:

- 总记录数: 33959
- 分析时段: 1930-01-01 00:00:00 至 2022-12-31 00:00:00
- 数据完整性: 100.0%
- 监测站数量: 1

2. 主要发现:

- Shapiro-Wilk 正态性检验: 数据不服从正态分布
- Kendall 趋势检验: 显著上升趋势
- 变异系数: 高变异性
- 一阶自相关: 显著自相关
- 分布形状: 偏度: 2.692, 峰度: 8.805

3. 数据质量评估:

- 缺失值: 无
- 异常值: 已检测并标记
- 时间跨度: 33967 days 00:00:00

完整报告已保存为 'water_data_analysis_report.txt'

数据分析完成!