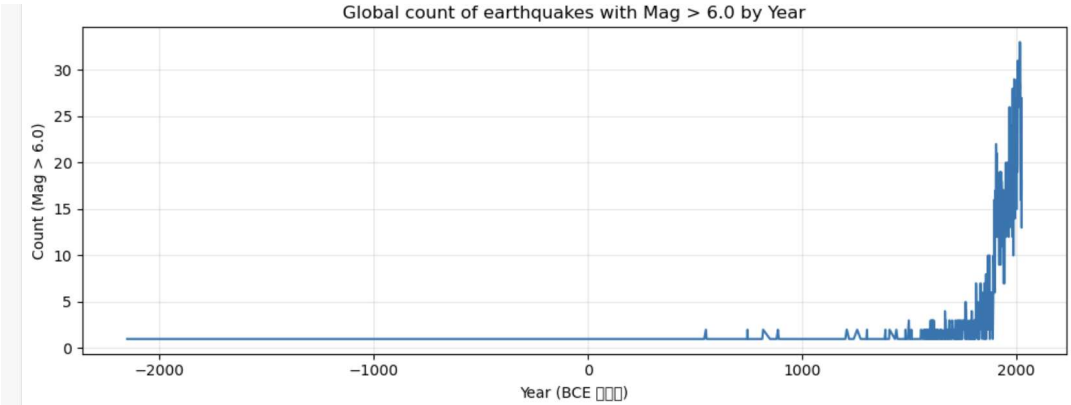
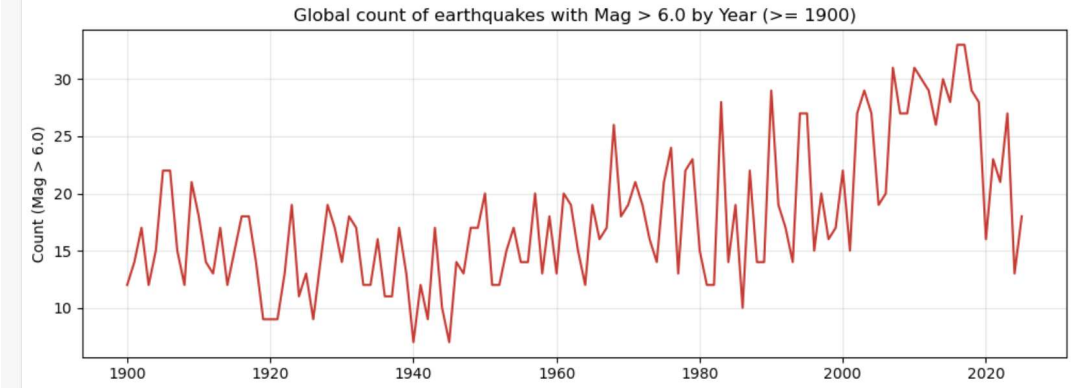
1.1

```
df = big_eqs.copy()

# 必要列检查
required = {"Year", "Total Deaths", "Location Name"}
missing = required - set(df.columns)
if missing:
    raise ValueError(f"缺少必要列: {missing}")

# 类型与过滤
df["Year"] = pd.to_numeric(df["Year"], errors="coerce")
df["Total Deaths"] = pd.to_numeric(df["Total Deaths"], errors="coerce")
df = df[df["Year"] >= -2150]

# 严格提取国家：仅取第一个冒号前的部分
country = df["Location Name"].astype(str).str.split(":", n=1, expand=True)[0].str.strip()
df = df.assign(Country=country)

# 去除空国家
df = df[df["Country"].notna() & (df["Country"] != "")]

# 汇总（默认 sum 会忽略缺失值，不做任何猜测填充）
top10 = (
    df.groupby("Country", as_index=False)["Total Deaths"].sum()
      .rename(columns={"Total Deaths": "Total_Deaths_Since_2150BC"})
      .sort_values("Total_Deaths_Since_2150BC", ascending=False)
      .head(10)
)

print("自公元前2150年以来各国地震致死总数（前十）：")
print(top10.to_string(index=False))
```

```
自公元前2150年以来各国地震致死总数（前十）：
   Country  Total_Deaths_Since_2150BC
     CHINA                  2106535.0
    TURKEY                  1008342.0
      IRAN                   745611.0
     ITALY                   423280.0
     SYRIA                   417724.0
     HAITI                   323782.0
AZERBAIJAN                   319251.0
     JAPAN                   318441.0
 INDONESIA                   282517.0
   ARMENIA                   189000.0
```

1.2





全时间趋势出现地震在近现代显著上升趋势，这可能与观测地震科技水平的提高有关，地震水平的提高使得我们观察到了更多的地震，而不是地震更加频繁。而 1900 年之后图年际波动明显（约 10–30 次/年），在现代观测期只见到“弱上升或近似平稳”的水平，不呈持续单调增长。个别年份偏高，多由大震序列（主震+余震群）驱动，例如 2004–2005、2011 前后。曲线末端的下沉常见于“最近一两年记录尚不完全”的边缘效应。

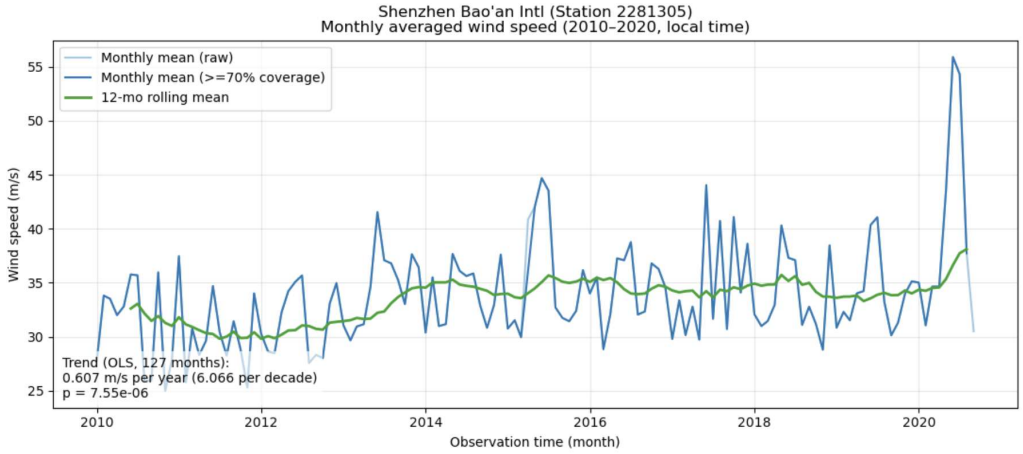
1.3

```
print(result.head(20))
```

	Country	total_count	largest_eq_date	largest_mag
0	CHINA	626	1668-07-25	8.5
1	INDONESIA	404	2004-12-26	9.1
2	IRAN	388	0856-12-22	7.9
3	JAPAN	364	2011-03-11	9.1
4	TURKEY	344	1939-12-26	7.8
5	ITALY	332	1915-01-13	7.5
6	GREECE	275	0365-07-21	8.0
7	PHILIPPINES	226	1897-09-21	8.7
8	MEXICO	204	1787-03-28	8.6
9	CHILE	200	1960-05-22	9.5
10	PERU	185	1716-02-06	8.8
11	RUSSIA	157	1952-11-04	9.0
12	BALKANS NW	118	1667-04-06	7.2
13	PAPUA NEW GUINEA	109	1919-05-06	8.2
14	CALIFORNIA	104	1857-01-09	7.9
15	INDIA	101	1897-06-12	8.0
16	TAIWAN	98	1920-06-05	8.2
17	ALASKA	80	1964-03-28	9.2
18	COLOMBIA	75	1826-06-18	8.2
19	ECUADOR	67	1906-01-31	8.8

2

```
print(f"p 值：{p_value:.4g} （常用阈值 0.05）")
```



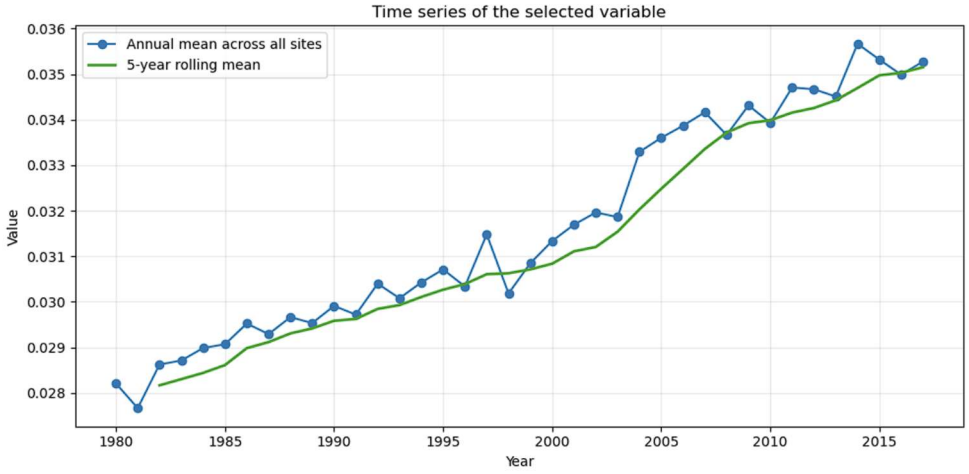

```
时间窗口： 2010-01-01 — 2020-12-31
通过完整性过滤的月份数：127 / 129
线性趋势斜率：0.6066 m/s/年 （每十年 6.0659 m/s）
p 值：7.549e-06 （常用阈值 0.05）
```

有显著上升的趋势，slope>0.6m/年

3.

```
report_basic_stats(ts)
```

```
原始形状： (24107, 41)
原始列名： ['lon', 'lat', '1980', '1981', '1982', '1983', '1984', '1985', '1986', '1987'] ...
清洗后形状： (24107, 40)（删除了 0 行全缺失记录）
年份列示例： ['1980', '1981', '1982', '1983', '1984', '1985', '1986', '1987', '1988', '1989']
```




```
=== Simple statistical checks ===
Count (non-NaN): 38
Mean: 0.0316368
Median: 0.0310926
Std: 0.00238399
Min: 0.0276654 at 1981
Max: 0.0356664 at 2014
5th–95th percentiles: 0.0285627 — 0.0352822
Skewness: 0.220 | Kurtosis: -1.355
Lag-1 autocorr: 0.973
Linear trend slope: 0.000209878 per year
Outliers (>+3σ): 0 | (<-3σ): 0
```