PS2

Assignment 02 Report

Name: XU JUNPENG (许俊鹏)   **SID:** 12532726

1. Significant earthquakes since 2150 B.C.

Method

I loaded the Significant Earthquake Database (TSV) as `Sig_Eqs` .

Then I:

Cleaned types and parsed necessary fields.

- **1.1** Grouped by Country and summed Deaths (treating missing/non-numeric as 0), then sorted to print the top-10 countries by total deaths.
- **1.2** Filtered rows with `Mag > 6.0` , computed the yearly count worldwide, and plotted the time series.
- **1.3** Implemented a function `CountEq_LargestEq(country)` that returns:
    i.   the total number of earthquakes for the country, and
    ii. the date of the largest earthquake (tie-break by latest date if magnitudes are equal).

I applied it to all countries and reported results in descending order (by total number of earthquakes, then by magnitude/date as needed).

Results

- **Deaths (top-10):** My table lists the ten countries with the largest cumulative fatalities since 2150 B.C.
- **Mag > 6 counts per year:** The line plot shows variability across centuries/decades.
- **Largest events by country:** For seismically active regions, the largest events align with well-known historical earthquakes. Countries with many recorded events are often those with long, well-documented histories and high seismicity along plate boundaries.

2. Wind speed in Shenzhen (2010–2020)

Method

I used NOAA ISD hourly data (station 2281305, Bao'an Airport). The ISD CSV has a composite `wind` field WND with comma-separated parts (`dir,dir_qc,type,speed,speed_qc`). I parsed wind speed and speed quality code as follows:

- **Wind direction angle** (POS 61–63): definition and range 001–360 (999 missing/variable).
- **Direction quality code** (POS 64): meanings of 0/1/2/3/4/5/6/7/9.
- **Type code** (POS 65): e.g., N = Normal, H = 5-min avg, R = 60-min avg, V = Variable, etc.
- **Speed rate** (POS 66–69): meters s^{-1} with scaling factor 10, `9999` = missing. I therefore divide by 10 to convert to m/s and drop `9999`.
- **Speed quality code** (POS 70): meanings of 0/1/2/3/4/5/6/7/9. I kept {1, 5, 9} to ensure only values that passed all QC (1/5) or passed gross limits if present (9).

I then:

1. kept timestamps between 2010–2020,
2. set `datetime` as index and resampled monthly ('M') to get monthly mean wind speed,
3. plotted monthly means and a 12-month rolling mean for smoothing.

Results

- The monthly mean series exhibits seasonal variability and interannual fluctuations.
- The 12-month rolling mean does not show a strong, monotonic long-term trend from 2010–2020 at this station.
- Interannual ups and downs likely reflect synoptic variability and local exposure changes (e.g., roughness, measurement context). A robust climatological trend claim would require a longer, homogenized multi-station analysis.

3. Exploring a PM2.5 data set (北京、上海、广州、深圳)

Method

I used `全国主要城市空气质量.csv` .

- **3.1 Load and clean.**

`AQC = pd.read_csv('全国主要城市空气质量.csv', encoding='gbk')`

- Column 1 = city, Column 2 = date. I converted `date` to `datetime` , sorted by `[city, date]` , converted all non-(city/date) columns to numeric, and filled NaN → 0. I saved the cleaned file as `air_quality_citys_cleaned.csv.`

- **3.2 Plot time series (PM2.5, four cities).**

I filtered to the four first-tier cities and made:

1. Daily time-series figure with 4 subplots (English titles: *Beijing, Shanghai, Guangzhou, Shenzhen*) and month minor ticks on the x-axis.
2. Monthly-mean figure with one plot and four lines (one per city). I also saved their monthly means as `pm25_first_tier_monthly_means.csv` with columns: city, year, month, `YYYY_MM` , `PM2.5` .

- **3.3 Statistics (per city, based on daily means).**

I computed five simple stats (mean/median/std/min/max) and exceedance counts vs. the 24-hour limits used in my code:

- 35 µg m^{-3} (stringent daily criterion in my analysis),
- 75 µg m^{-3} (GB 3095-2012 Grade II daily limit commonly referenced in practice).

I also reported:

- the month (`YYYY-MM`) with the most > 75 µg m^{-3} days, and
- the annual mean PM2.5 per year.

I saved tables as `pm25_first_tier_stats.csv` and `pm25_first_tier_annual_means.csv`

Findings

- The **four daily series** and **monthly means** clearly show intra-annual variability and inter-city differences.
- The **exceedance tables** quantify days above **35** and **75 µg m⁻³** for each city, highlight the **month** with the most severe daily exceedances, and list **annual averages**—all directly computed from the code outputs saved to CSV.

Collaboration

- I used ChatGPT to help me translate this report from Chinese to English.