

Assignment 02

By Jie ou (ID:12432886)

1. Significant earthquakes since 2150 B.C

1.1 Top 20 Countries by Total Deaths

Step1: 数据清洗

在“Location Name”中根据字符的编写特征提取国家名的信息，及其对应的死亡数。如果仅有“Death Description”，则将其进行赋值：0—0，1（1~25）—25，2（51~100）—75，3（101~1000）—500，4（>1000）—5000）。由于国家名表示方式较为多样，有的包含了多个国家，将其死亡数平均给这些国家；有的具体到了国家的某个地区，将其统一修改为国家，而不需要具体到地区。

Step2: 分组合并统计各个国家地震造成的死亡数

Step3: 敏感性分析

由于“Death Description”为4为大于1000，其范围可以存在较大变化，观察将其赋值给2000、5000、10000的情况下，对死亡数最高的二十个国家的影响。表1可见，Death Description变化后依旧是这二十个国家，且仅对其中三个国家有小幅影响。

Table 1 Sensitivity of Top 20 Countries' Deaths with Varying Description 4.0 Values

4.0 赋值 2000			4.0 赋值 5000		4.0 赋值 10000	
Country	Deaths		Country	Deaths	Country	Deaths
1	China	2120987	China	2144987	China	2184987
2	Turkey	1181181	Turkey	1181181	Turkey	1181181
3	Iran	1028955	Iran	1034955	Iran	1044955
4	Italy	505643	Italy	505643	Italy	505643
5	Syria	385641	Syria	385641	Syria	385641
6	Haiti	323478	Haiti	323478	Haiti	323478
7	Azerbaijan	317719	Azerbaijan	317719	Azerbaijan	317719
8	Japan	285232	Japan	285232	Japan	285232
9	Armenia	191902	Armenia	191902	Armenia	191902
10	Israel	161754	Israel	161754	Israel	161754
11	Pakistan	145082	Pakistan	145082	Ecuador	147917
12	Ecuador	139917	Ecuador	142917	Pakistan	145082
13	Iraq	138450	Iraq	138450	Iraq	138450
14	Turkmenistan	117412	Turkmenistan	117412	Turkmenistan	117412
15	Peru	105491	Peru	105491	Peru	105491
16	Greece	89388	Greece	89388	Greece	89388
17	Portugal	84059	Portugal	84059	Portugal	84059
18	India	66276	India	66276	India	66276
19	Chile	65799	Chile	65799	Chile	65799
20	Taiwan	59702	Taiwan	59702	Taiwan	59702

1.2 Annual Earthquake Magnitude Analysis

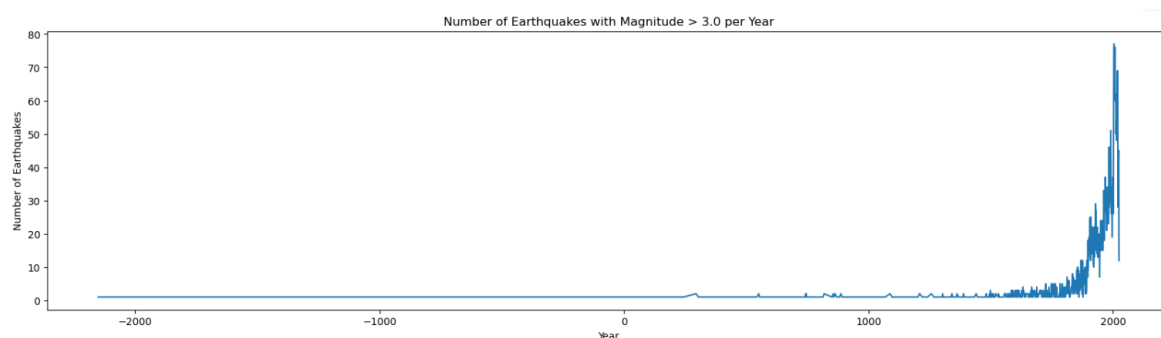


Figure 1 Number of Earthquakes with Magnitude > 3.0 per Year since 2150 B.C.

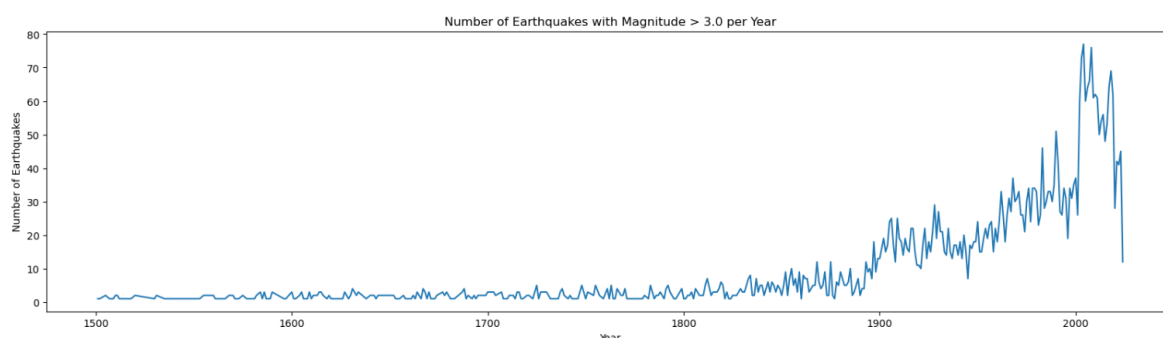


Figure 2 Number of Earthquakes with Magnitude > 3.0 per Year since 1500

从 1500 年前的统计数据来看，地震数量相对平稳且震级不高。然而，20 世纪之后地震记录数量显著增加。但这并非意味着地震活动实际增加，而是由于地震检测和记录技术的进步，使得我们能够更精确地捕捉到更多的地震事件。（19 世纪末，我们才拥有了能够测量地震波的校准仪器。）

1.3 CountEq_LargestEq function

Step1: 定义函数

对可能存在的多个最大震级地震进行检索，且根据各自数据日期的信息量给定相应的地震发生时间（如仅给定年月，则为年-月；如信息完全，则为年-月-日 时：分：秒）。

Step2: 对 Country 列表进行清洗

和 1.1 中类似，对国家名进行处理，并得到地震数据中出现所有不重复的国家名列表，用以后续遍历各个国家进行统计。

Step3: 调用函数输出结果

将得到的国家名列表及国家名处理过的地震数据传入定义好的函数中，按倒叙输出结果，见图 3。

	Country	Earthquake Number	Location and date
0	CHINA	625	INDIA-CHINA in 1950-8-15 14:9:30.0
1	JAPAN	415	JAPAN: HONSHU in 2011-3-11 5:46:24.1
2	INDONESIA	397	INDONESIA: SUMATRA: ACEH: OFF WEST COAST in...
3	IRAN	390	IRAN: DAMGHAN, QUMIS in 856-12-22
4	TURKEY	335	TURKEY: ERZINCAN in 1939-12-26 23:57:23.8, TU...
...
247	TEXAS	1	TEXAS: WEST in 2013-4-18 0:50:38.5
248	MADAGASCAR	1	MADAGASCAR: ANTSIRABE in 2017-1-11 22:6:58.7
249	ZAMBIA	1	ZAMBIA: KAPUTA in 2017-2-24 0:32:17.8
250	COMOROS	1	COMOROS: MAYOTTE in 2018-5-15 15:48:9.9
251	NORTH CAROLINA	1	NORTH CAROLINA: SPARTA in 2020-8-9 12:7:37.0

252 rows × 3 columns

Figure 3 Earthquake Data Summary by Country

2. Air temperature in Shenzhen during the past 25 years

Step1: 数据清洗

将“TMP”中用于描述数据质量的数字分隔开（用 `split` 函数，以“,”分隔）至新建的表明数据质量的“Quality”中。剔除数据中“+9999”的数据，并检查剩余数据的质量。经检查，剩余数据的质量仅有少数的 2（数据可疑，但不一定是错误的）和 5（通过了所有质量控制检查，且数据来源于 NCEI 的数据源），认为可用于后续的分析。

Step2: 对时间数据进行转换并绘图

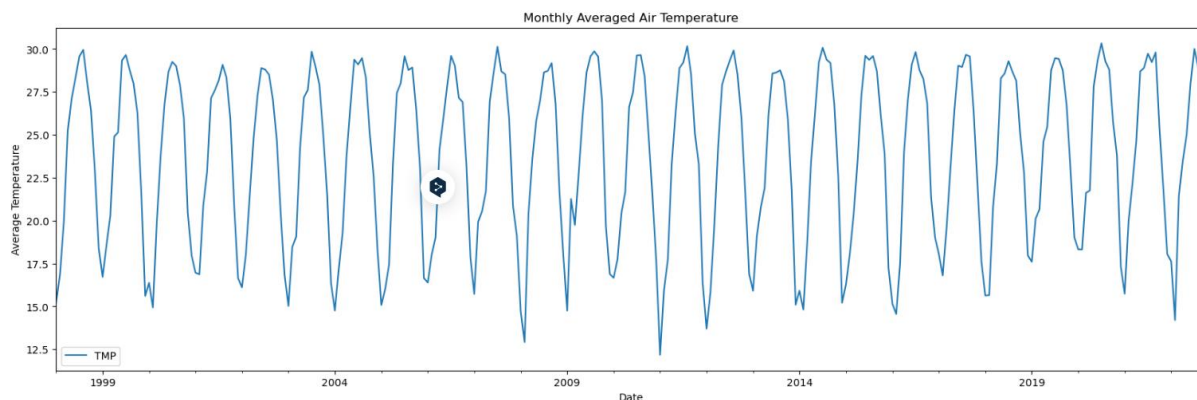


Figure 4 Monthly Averaged Air Temperature during the past 25 years

3. Global collection of hurricanes

3.2

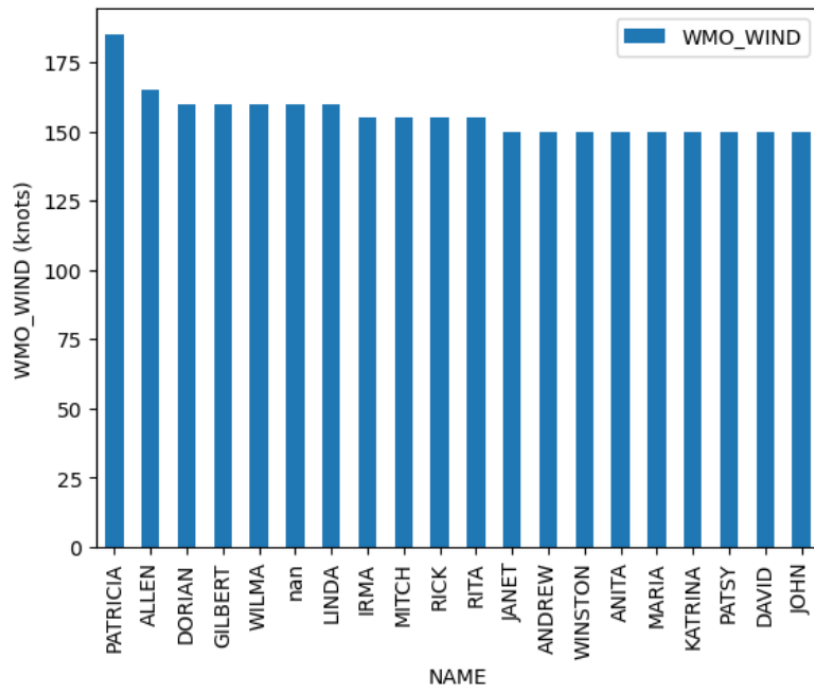


Figure 5 Top 20 Hurricanes by Maximum Wind Speed (WMO_WIND)

3.3

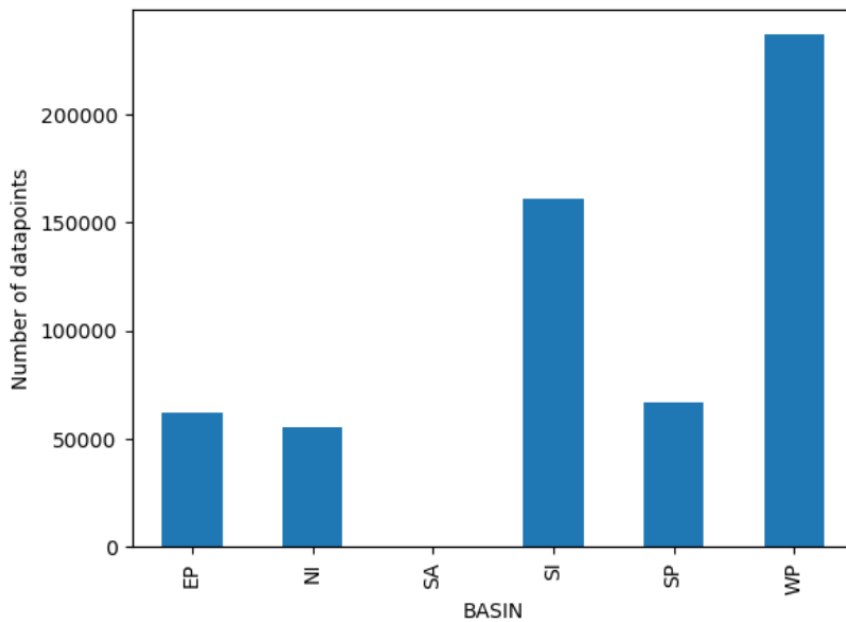


Figure 6 Count of all Datapoints by Basin

3.4

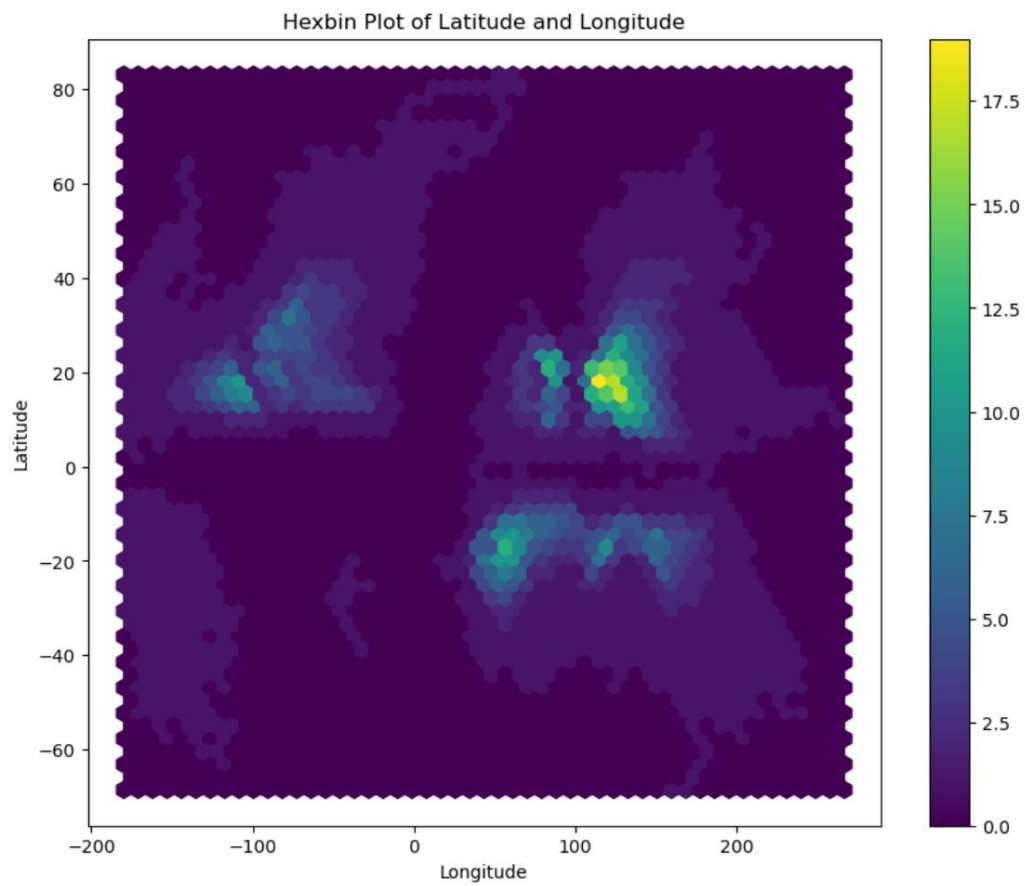


Figure 7 Hexbin Plot of Latitude and Longitude

3.5

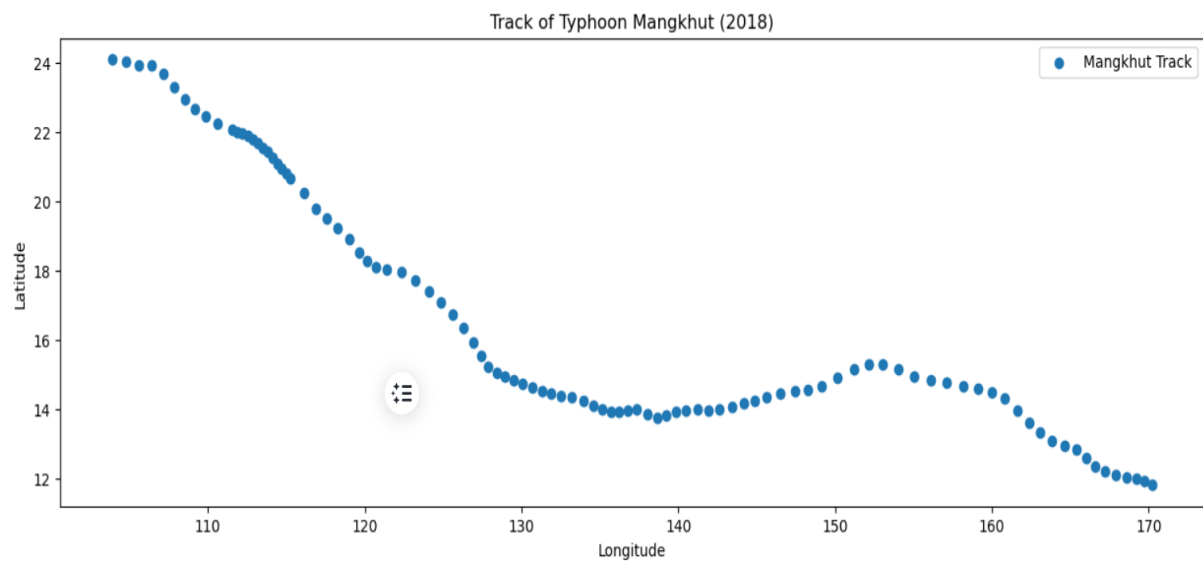


Figure 8 Track of Typhoon Mangkhut (2018)

3.7

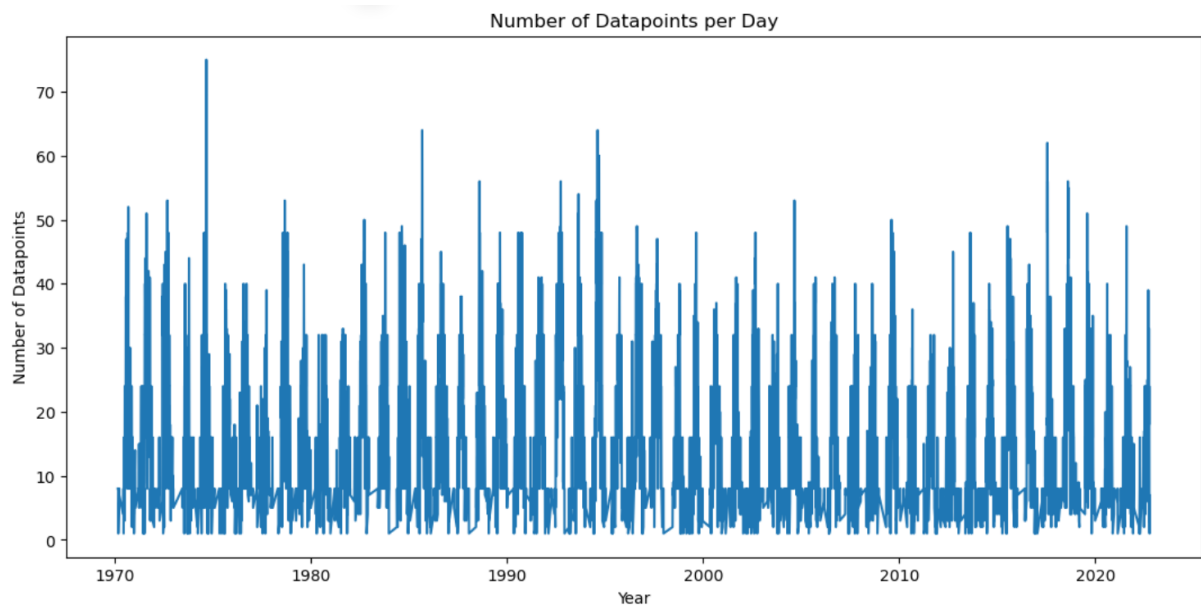


Figure 9 Number of Datapoints per Day

3.8

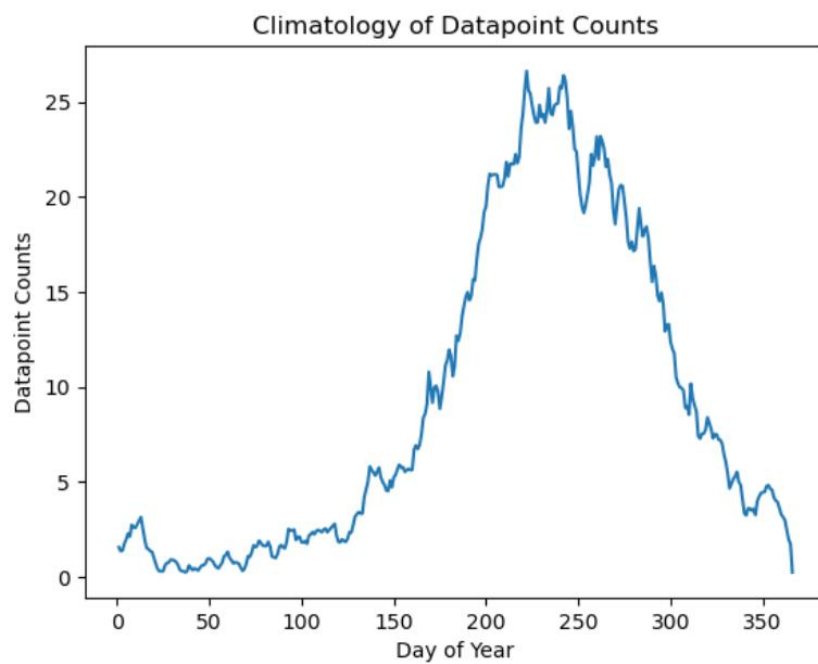


Figure 10 Climatology of Datapoint Counts

3.10

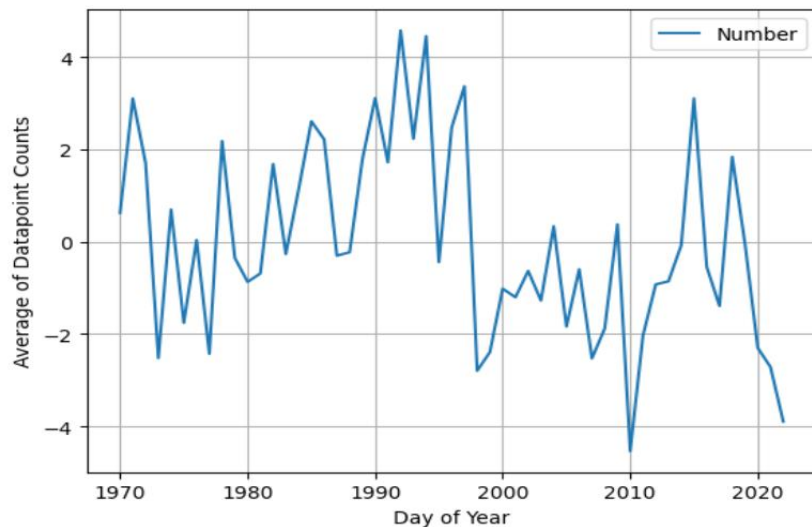


Figure 11 Anomalous Hurricane Activity

图中显示, 1971、1978、1985、1986、1990、1992、1993、1994、1995、1996、1997、2015 年份飓风活动相对较为活跃(Datapoint Counts >2); 1992、1994 年非常活跃(Datapoint Counts >4)。

4. Explore a data set

下载了站点“JAN MAYEN NOR NAVY, NO”(01001099999)在 2023 年一年中的气象数据, 主要针对风速数据进行了分析。

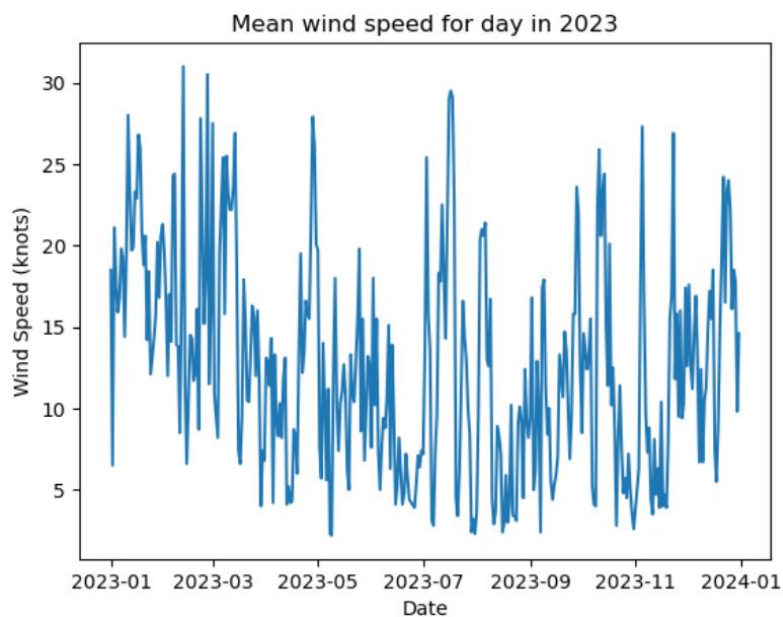


Figure 12 Mean Wind Speed for Day in 2023

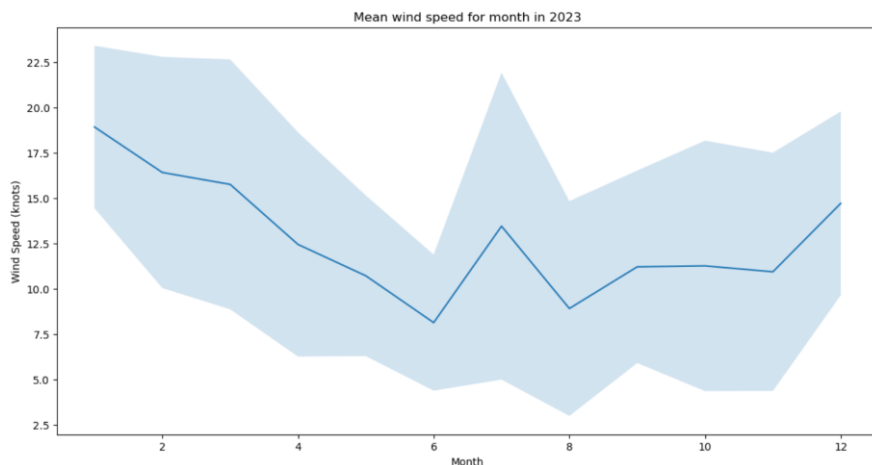


Figure 13 Mean Wind Speed for Day in 2023

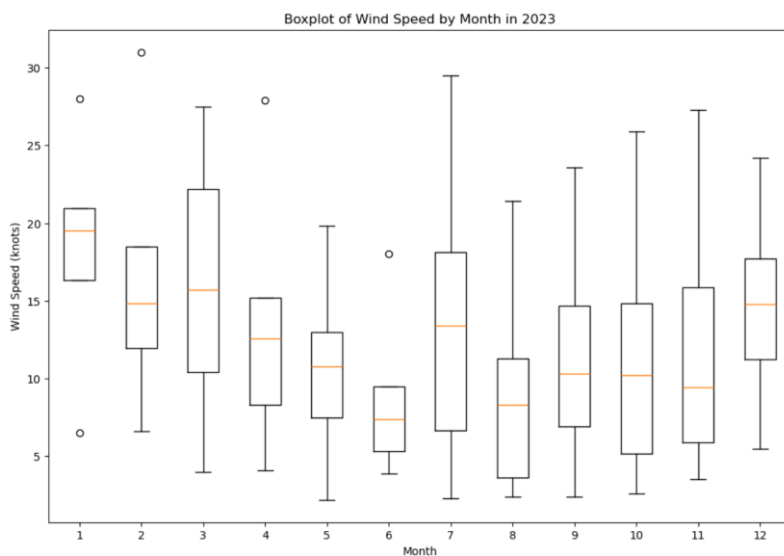


Figure 14 Boxplot of Wind Speed by Month in 2023

根据 2023 年 JAN MAYEN NOR NAVY, NO 的风速数据，该地区风速在一年中存在明显的季节性变化，冬季风速相对较高，而夏季的 6 月和 8 月风速相对较低。全年风速波动较大，尤其是 7 月份，表现出较高的不确定性和一些极端风速值。整体来看，风速的分布并不均匀，某些月份如 1、2、4、6 月出现了离群点，这可能与特定的天气事件有关。