Group 32 Final Project Proposal 嚴偉哲 郭穎達

Part I: Dataset

The dataset we choose is about the Dengue cases confirmed in Taiwan throughout 27 years (1998~2024), which contains the following attributes:

- **Date Onset:** When the symptoms appeared
- **Date Confirmation:** When the case(s) was confirmed
- **Date Notification:** When the case(s) was reported

- **Sex:** Patient's gender
- **Age Group:** Patient's age category
- **Serotype:** Type of the dengue virus

- **Country/Township/Village Living:** Where the patient lives at
- **Country/Township/Village Infected:** Where the patient was contracted
- **Enumeration Unit:** The unit that collects the data
- **Dissemination Unit:** The unit that distributes the data

There are roughly 107k cases recorded in total, which can be seen here: https://www.kaggle.com/datasets/taweilo/taiwan-dengue-daily-confirmed-cases-1998-2024

Date_Onset	Date_Confirmation	Date_Notification	Sex	Age_Group	County_living	Township_living	Village_Living	Village_Living_C..
1998-01-02 2024-09-10	2004-05-16 2024-09-10	1998-01-07 2024-09-10	M 51% F 49% Other (1) 0%	70+ 13% 60-64 9% Other (83484) 78%	Kaohsiung City 46% Tainan City 45% Other (9514) 9%	Sanmin Dist. 11% North Dist. 9% Other (86449) 81%	None 9% Chenggong Vil. 1% Other (96969) 91%	None 6703100-001 Other (97072)
1998/01/02		1998/01/07	M	40-44	Pingtung County	Pingtung City	None	None
1998/01/03		1998/01/14	M	30-34	Pingtung County	Donggang Township	None	None
1998/01/13		1998/02/18	M	55-59	Yilan County	Yilan City	None	None
1998/01/15		1998/01/23	M	35-39	Kaohsiung City	Lingya Dist.	None	None
1998/01/20		1998/02/04	M	55-59	Yilan County	Wujie Township	None	None
1998/01/22		1998/02/19	M	20-24	Taoyuan City	Luzhu Dist.	None	None
1998/01/23		1998/02/02	M	40-44	New Taipei City	Xindian Dist.	None	None
1998/01/26		1998/02/19	F	65-69	Taipei City	Beitou Dist.	None	None
1998/02/11		1998/02/13	F	25-29	Tainan City	South Dist.	None	None
1998/02/16		1998/02/24	M	20-24	Kaohsiung City	Nanzi Dist.	None	None
1998/02/17		1998/02/23	F	30-34	Kaohsiung City	Fengshan Dist.	None	None
1998/02/23		1998/03/04	M	55-59	New Taipei City	Linkou Dist.	None	None
1998/03/05		1998/03/11	M	40-44	Kaohsiung City	Fengshan Dist.	None	None
1998/03/07		1998/03/12	M	30-34	New Taipei City	Tamsui Dist.	None	None
1998/03/10		1998/03/24	M	55-59	New Taipei City	Xinzhuang Dist.	None	None
1998/03/21	2006/11/29	1998/03/27	F	35-39	Taipei City	Neihu Dist.	None	None
1998/03/22		1998/03/27	M	30-34	Kaohsiung City	Qianzhen Dist.	None	None

Figure I.1: The dataset table

Part II: Visualization Method

Since this dataset contains the time and the place of the cases, we can use this information to analyze the spread and control of the disease over time. We plan to use the theme river and dot/bubble map to visualize the data. In the theme river plot, the x-axis shows the time,

and the y-axis represents the number of dengue cases. Furthermore, we can use the sex, age groups, and serotypes to categorize the cases and give the chart a more detailed look.

For the dot/bubble map, we will overlap the coordinates of the cases with the map of Taiwan. This is to show the distribution of the patients, and we can observe the severity of the endemic. We can also use the color of the dots to represent the sex, serotypes, or how long the time has passed since the case report. The size of the dot can also represent the number of cases in that region.

This can be further visualized by making the theme river brushable, so that we can control which time period we want to see in the map. By sliding the selected period, we can also see the spread of the disease.