# Analysis of Climate Impact on Dengue Cases

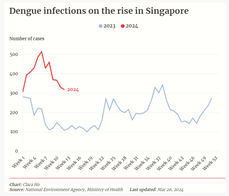
ISSS608 Group 3: CI Hui, Colin JIANG Kelin, SUN Yiping
Advisor: Prof. KAM Tin Seong

SMU SINGAPORE MANAGEMENT UNIVERSITY

## Introduction & Motivation

In 2012, WHO ranked dengue as “the most important mosquito-borne viral disease in the world”.

The continuously growing number of reported cases has made it one of the most concerning public health issues in Singapore.



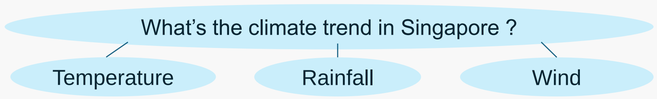

(link)

This project aims to develop a user-friendly tool for people without coding background to explore the impact of weather on dengue cases and forecast the future dengue cases.

## Objectives

The project aims to leverage the visualization techniques and advanced analytics to analyze and forecast

- ❑ The climate trend in Singapore from 3 aspects




- ❑ The weekly dengue cases reported in Singapore
- ❑ Forecast future dengue cases using univariate and multivariate approaches

## Tools & Methodology

Multiple tools and methodologies have been adopted.

**R studio** was used for data preparation, visualization and model development

**Shiny** was used to create the web application for users to analyze data and calibrate models

**Methodology**

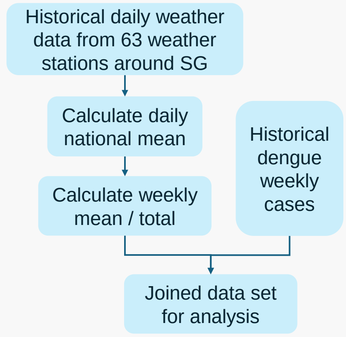
- ❑ Various **visualization** methods were adopted using R packages such as ggiraph, ggstatsplot, tmap etc.
- ❑ **TSLM model** was used to understand the impact of climate on dengue cases
- ❑ **ARIMA, ETS, VAR models** were used to forecast future dengue cases

## Data Preparation

The data used in this project were obtained from two sources:

- ❑ **Historical climate data**: Meteorological Service Singapore
- ❑ **Historical dengue weekly reports**: SGCharts: Outbreak

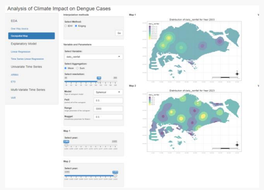
The two data sets were cleaned and processed before joining to form the final analytical data set.




## Data Exploration

### One-Way ANOVA Test

Pair-wise comparisons of the means to check if the two groups are significantly different.


### Spatial Interpolation

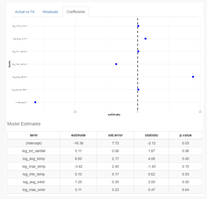
Interpolates the climate data for the area without weather stations using the data from the existing 63 weather stations.

## Feature Importance

**TSLM model** was used to study the impact of the climate on dengue cases. **Data transformation methods** are provided for the users to tune the model and get more accurate coefficients of the factors.


Users can choose to perform log transformation, min-max or Z scaling on the selected factors.

Coefficients can be visualized to understand if the factors have positive or negative impact on dengue cases, as well as the strength of the association.




| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | 46.36 | 7.70 | -6.10 | 0.00 |
| log_lot_rainfall | 0.11 | 0.06 | 1.87 | 0.06 |
| log_avg_temp | 8.65 | 2.17 | 4.09 | 0.00 |
| log_max_temp | -3.42 | 2.40 | -1.43 | 0.15 |
| log_min_temp | 0.10 | 0.17 | 0.62 | 0.53 |
| log_avg_wind | 1.25 | 0.38 | 3.09 | 0.00 |
| log_max_wind | 0.71 | 0.23 | 3.47 | 0.64 |

## Time Series Forecast

Three time series forecast models are provided in the Shiny app.

### Univariate Time Series Model

**ARIMA**


**ETS**


Forecast the future dengue cases purely based on the trend and seasonality in the historical data.

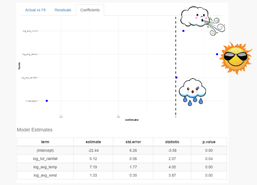
### Multivariate Time Series Model

**VAR**


Assuming climate does have an impact on dengue cases, adding additional information in the forecast model would result in more accurate forecasts.

## Insights

The results from TSLM model show that the dengue cases is indeed influenced by the climate factors. The screenshot on the right tells that **all the three factors** (i.e., temperature, rainfall and wind) are **significant with positive coefficients**.




| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -22.44 | 6.29 | -3.56 | 0.00 |
| log_lot_rainfall | 0.12 | 0.06 | 2.07 | 0.04 |
| log_avg_temp | 7.78 | 1.77 | 4.40 | 0.00 |
| log_avg_wind | 1.33 | 0.38 | 3.87 | 0.00 |

1. **Temperature**: more dengue cases during warmer periods
2. **Wind**: more dengue cases during the periods with stronger wind
3. **Rainfall**: more dengue cases during the periods with more rainfall

## Future Work

The project team managed to create an app for the users to analyze the climate data using different visualization methods and to customize their own models to forecast the dengue cases, despite the short timeline and the struggles among other course and work assignments.

A few improvements could be considered for the future work:

- ❑ **Correlation Analysis:** time series correlation analysis to understand the relationship among the climate data, and with dengue cases.
- ❑ **Spatial Interpolation:** incorporate the results from spatial interpolation might improve the model’s forecast accuracy.
- ❑ **Model Comparison:** it’s a bit difficult to compare the forecast results from different models due to the current design in the Shiny app. This can be improved if we have more time as we are more familiar with Shiny App now.