Introduction of Project, Objectives and Description

The project focuses on designing a data visualization algorithm to visualize the movement of people under the Covid-19 situation. The design of the data visualization is to provide insights into the patterns of a Covid-19 outbreak. Areas that have a lack of clarity or definition during a Covid outbreak will also be helped to identified. Therefore, this project will help people to have a general knowledge or pattern regarding the formation of Covid clusters.

Data visualization is a technique that technique that forms information into visual context. This enables people to identify patterns within large datasets. When having large datasets in dealing with the with the movement of people, data visualization technique is crucial for analysing the correlations for its datasets. Complex algorithms are easier to be identified with graphs than numbers.

A visualization interface is to be designed so that the variables can be interchanged to show the relationship between the datasets. A graphing algorithm is also implemented to show the relationship between the movement of different clusters of Covid. The visualization designed will help groups of people in achieving a much better understanding of the pattern of Covid-19 and how it spreads.

(Week 1)

i. Roles

For the conduct of the project, our team were to be separated into different roles. This is to define people to specialize in proper objectives instead of having multiple people working on one task at a time.

Roles were categorize into:

Leader – The person who manages the team and provides documentation for the project and creates the survey and evaluation form for the project.

Designer – The person who designs the data visualization for the project and its interface.

Main Code Developer – The person who is mainly working on the graphing algorithm of the project in python.

Assistant Code Developer – The person who is assisting in the graphing algorithm, collaborating with the designer to provide reliable information to the main code developer. As well as develop the website for the project.

ii. Project Analysis

A survey would be required to be designed to obtain opinions from other people to see which graphing techniques would be suitable for designing the data visualization.

Datasets would be required to be identified to determine which datasets were suitable to be used for the visualization, as samples with low potential would not meet the requirements for obtaining positive results in correlation. The relationships between the datasets could also be identified.

The project would require to have a working interface for adjusting the of variables. The interface is used to display the datasets in a graphical manner. This can be designed on the data visualization or Python.

A graphing algorithm was then to be designed so that the movement of clusters between regions could be identified.

iii. Planning and execution based on FILA FORM

The project will follow through a Design Thinking Process when conducting it. The Design Thinking Process is separated into four steps, Empathy, Ideate, Design, Prototype and Testing. Each of these steps provide a different objectives for the achievement of the project.

● Empathy (Week 1)

In the Empathy stage, we were required to search and identify visualization techniques for the design of the data visualization. We were required to be familiarized with different methods of graphing. This is required so that we would identify a suitable visualization technique for the project. A survey must also be formed and be given to multiple people for their opinion on the data visualization method.

From the resources provided, several of the data and methods of plotting were chosen for the survey. This is to provide a generalized opinion of how the people think and taking in as reference for the design of our data visualization for the project.

● Ideate (Week 2)

Datasets were taken in from the provided sources given in the assignment document and implemented to a program. The datasets chosen were the number of check-ins per day towards the infected cases per day. However, during the plotting, we found out that the graph shown does not show any positive correlation regarding the number of check-ins towards the infected cases per day. In a sense, the datasets were not complete to us, therefore several methods of implementation were used so that the data shown were much clearer to us.

Datasets chosen:

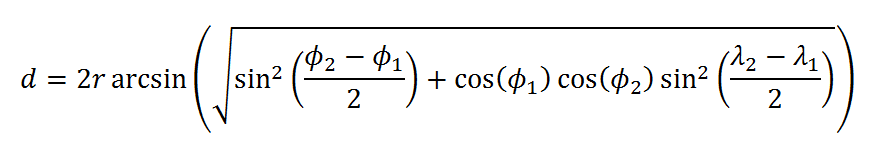
* Check-ins by state
* New cases by state
* New deaths by state
* ICU admission by state
* Vaccination by state
* Clusters

In our implementation, we decide to include the dates and phases of our movement control order (MCO) to make sense of the irregularities of the correlations regarding the number of check-ins towards infected cases per day.

● Design (Week 2)

Tableau is chosen to be our choice of data visualization software. Tableau is a popular interactive data visualization software with various chart features. Two dashboards are constructed to visualize the pandemic situation. The workflow is as described below:

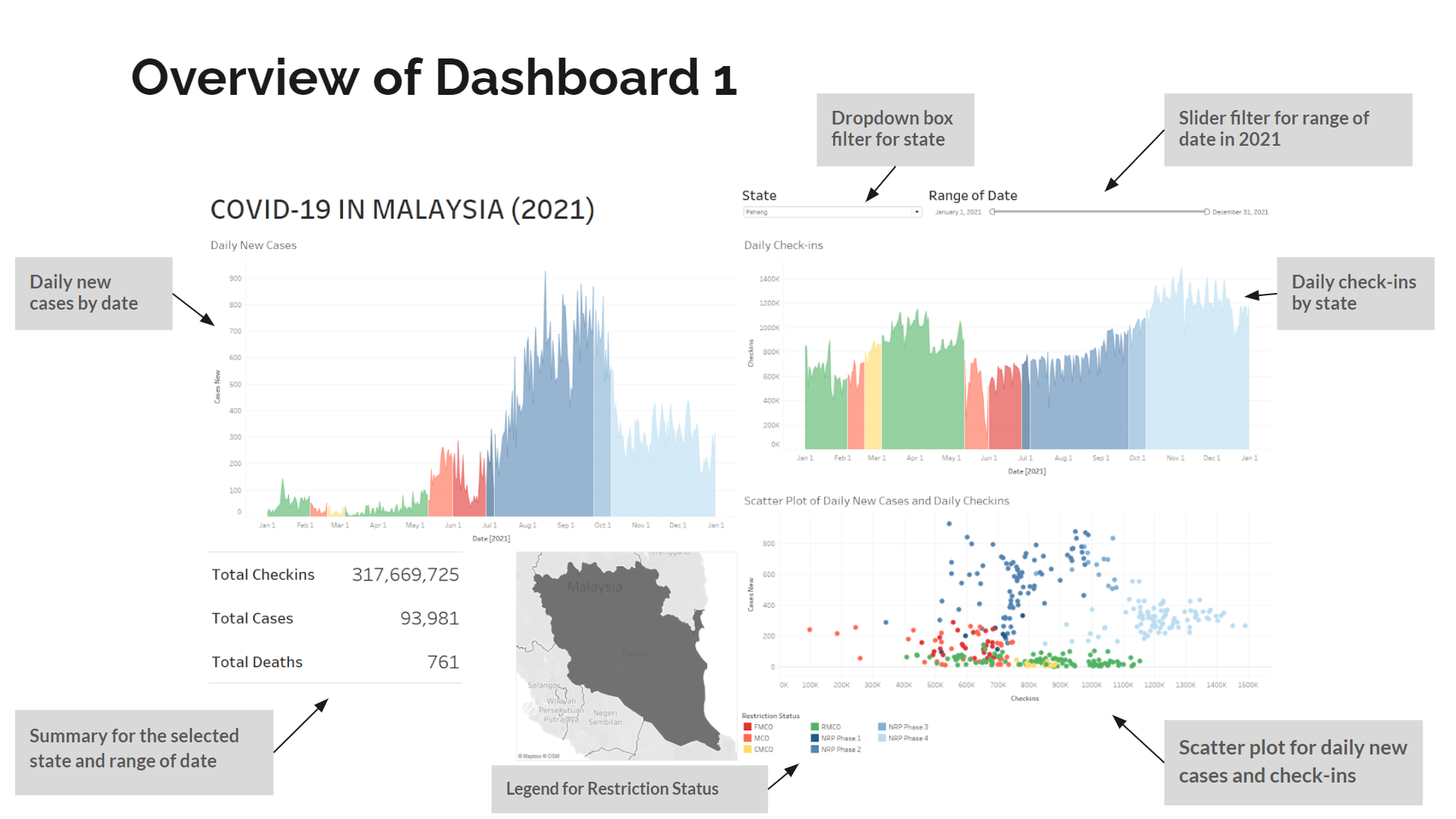
1. Data collection & cleaning
   * Collect restriction status of states from Wikipedia and news websites
   * Compile list of districts (daerah) and obtain coordinates by geocoding
   * Calculate distance between districts by using the Haversine formula which is used to calculate the distance between any two points on a sphere.



assuming r = 6378km for earth’s radius at the equator

* + Restructure clusters dataset for graph network

1. Import to Tableau
   * Import all relevant datasets
   * Merge dataset by date and state
   * Only include data points that are recorded in 2021
2. Create charts
   * Area chart, scatter plot, network graph, table
   * Labelling (color, size, text)
3. Construct dashboard
   * Multiple charts joined to create a dashboard.
   * Add state and date filters

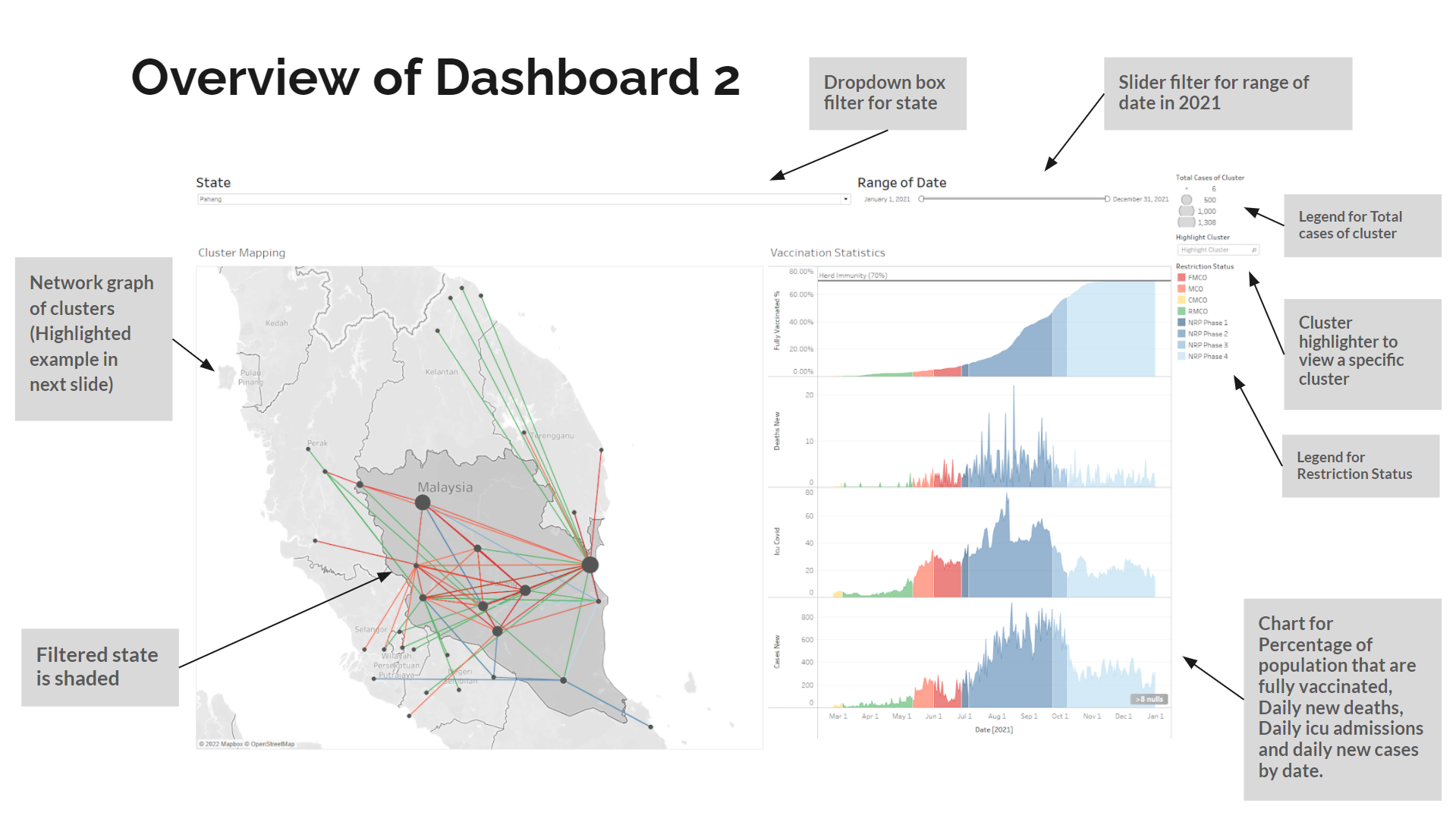
The following diagram is an overview of the first dashboard,

The elements of this dashboard are,

1. State and Range of date filter
2. Area chart for daily new cases by date for the selected state and range of date, color labelled according to the restriction status of the selected state.
3. Area chart for daily check-ins by date for the selected state and range of date, color labelled according to the restriction status of the selected state.
4. Table summarizing the total number of check-ins, cases and deaths for the selected state and range of date.
5. Small map with the selected state shaded
6. Scatterplot for daily new cases and daily new check-ins for the selected state and range of date, color labelled according to the restriction status of the selected state.
7. Colour legend for the restriction status (FMCO, CMCO, RMCO, MCO, NRP Phase 1 to 4)

● Design (Week 3)

The following diagram is an overview of the second dashboard,

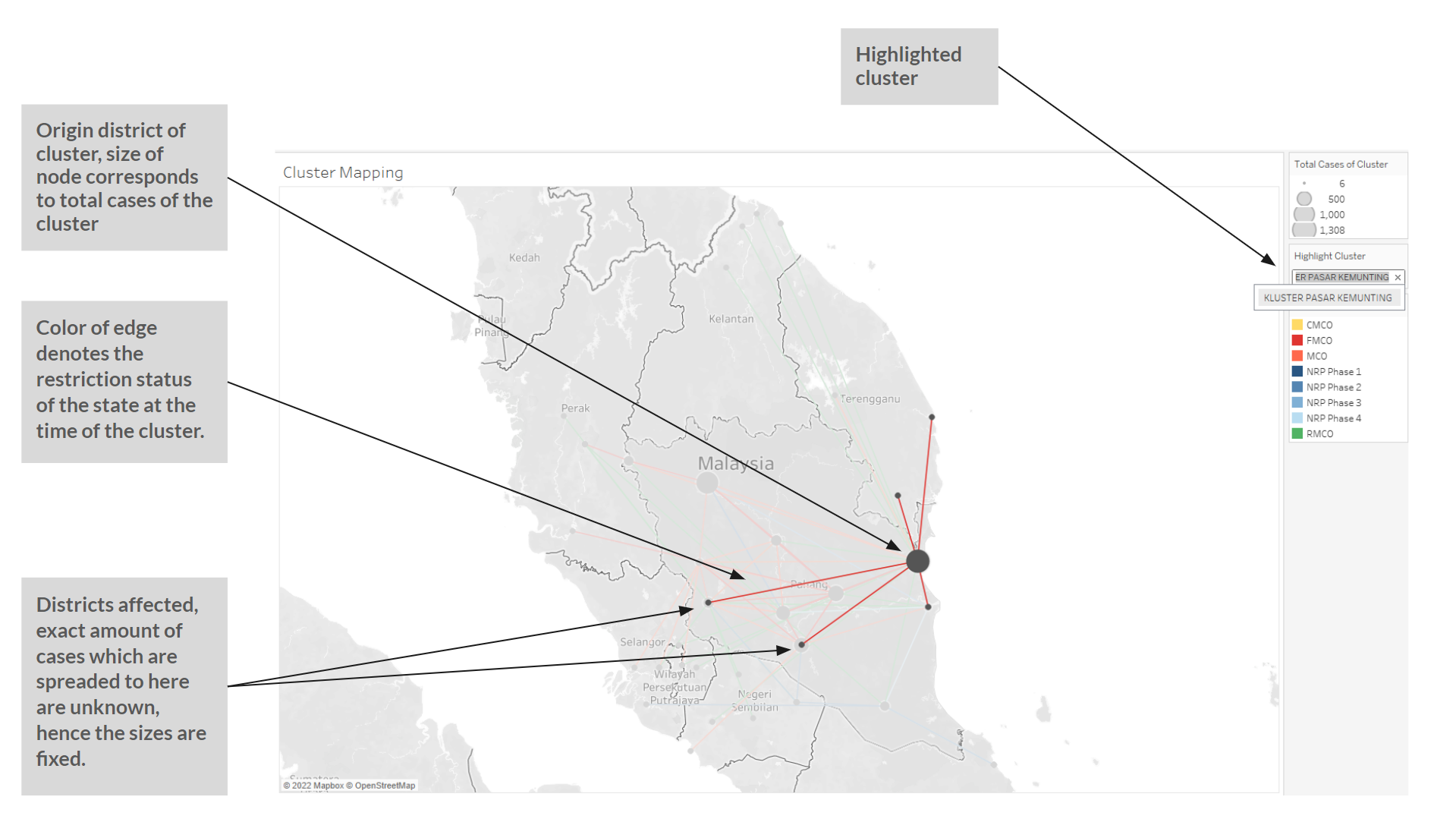


The elements of this dashboard are,

1. State and Range of date filter
2. Network graph of the clusters that originated from the selected state and range of date. Each node represents a district (daerah) and the size of the node denotes the total number of cases in that cluster. The color of the edges denotes the restriction status during the period of the cluster. The selected state is also slightly shaded. Edges that link to nodes out of the selected state indicate an interstate cluster.
3. Area charts of,
   * Percentage of population fully vaccinated by date
   * Daily new deaths by date
   * Daily ICU admissions by date
   * Daily new cases by date

for the selected state and range of date. This combination of charts can be used to show the efficacy of vaccines against hospitalization and deaths.

1. Size legend for total number of cases for a cluster.
2. Cluster highlighter to select and view a specific cluster, the selected view is as shown below,



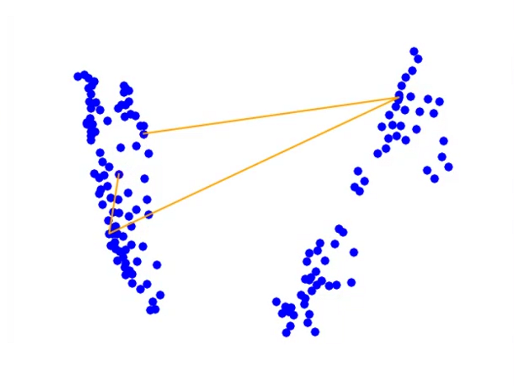
1. Colour legend for the restriction status (FMCO, CMCO, RMCO, MCO, NRP Phase 1 to 4)

● Prototype and Testing (Week 4)

The graphing algorithm is designed to identify the shortest path to move from one region to another. The graph algorithm is implemented into the interface by using Python. Python is required as Tableau does not have the functions for designing a graphing algorithm.

The data cleaned used for cluster mapping is used for the graphing algorithm with the data structure of adjacent lists and adjacent matrix.

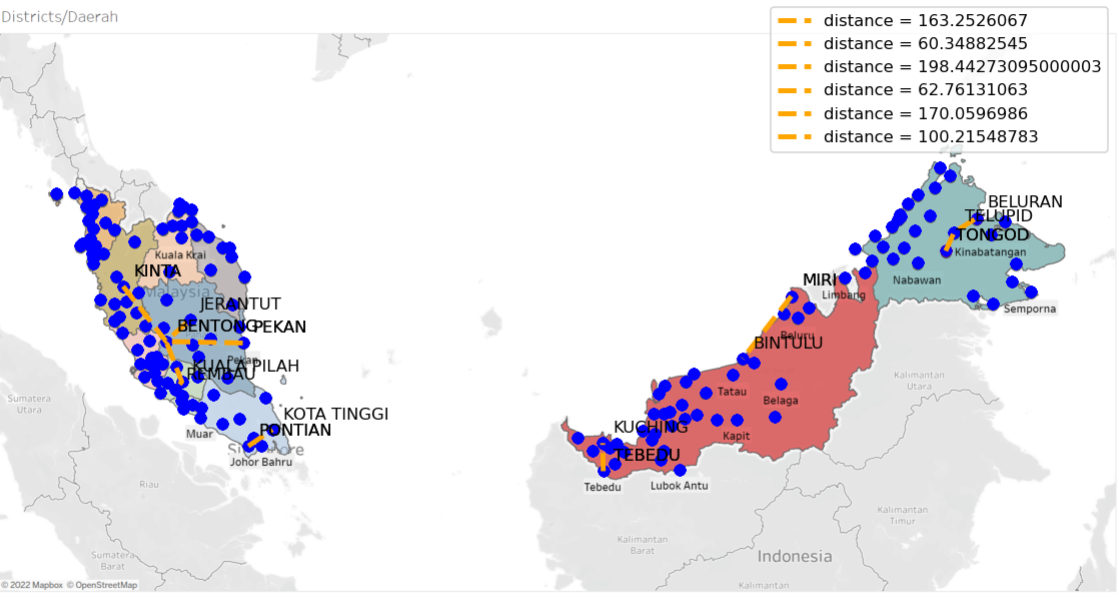
In using the graphing algorithm, the shortest travelling path between two different districts was discovered. Two methods were provided from the group with one being Dijkstra’s algorithm. One or both combined methods would be used for the graphing algorithm.



● Prototype and Testing (Week 5)

When running the group algorithm, a bug was found involving the inconsistent display of the node points. This occurs due to one of the datasets not being sorted correctly before converted into the proper format for the plotting. After implementing a sorting algorithm, the problem was resolved.

A map was also implemented into the algorithm in order to familiarize the coordinates and the Malaysia districts maps. This enables the users to be able to identify the districts of the clusters easier with the map designed.



iv. Overall reflection of the experience doing the project

While designing the project, we came across several problems when dealing with the team information as many of us have different skills in tackling this project. The team was able to provide many applications in designing the methodology and also website for the project. Discussions held were also able to information each other members with the condition of the project and their tasks clearly.

While being different, the tasks are divided specifically to match the skills that each team member possesses. While each team member has their own profession when designing the algorithm, there were several tasks such as , designing the survey, sorting the reliable data from the given datasets and implementing the python code to the website, could be deemed as tasks that requires more discussion and corporation with each other.

Overall, we learn to work within a team to design a grouping algorithm for the Covid 19 situation. Even though everyone has their own respective professions, yet this group project provides us an opportunity to learn to come together and work with people from different backgrounds.

v. Conclusion

The implementation of the grouping algorithm was a success as the data visualization interface was successfully implemented and able to identify the correlation between the number of infected cases vs. the number of check-ins per day. There were other environmental factors such as the SOP imposed or the presence of vaccination of the people being discovered as factors towards the irregularities of the correlation. Furthermore, a graphing algorithm was successfully implemented that identifies the distances between clusters from each district. This provides insight regarding the correlation between two clusters from different districts.