# FloSG - Analysis of Commuter Flows (Bus) in Singapore using Spatial Interaction Models

Genice Goh
School of Information Systems, Singapore Management University
genice.goh.2018@scis.smu.edu.sg

Nor Aisyah Bte Ajit
School of Information Systems, Singapore Management University
aisyahajit.2018@scis.smu.edu.sg

Wong Wei Ling
School of Information Systems, Singapore Management University
wlwong.2018@scis.smu.edu.sg

## ABSTRACT

In light of the recent pandemic, the outbreak has caused daily lives throughout the world to be altered dramatically, with far-reaching implications for the global economy, trade, and mobility. Almost every facet of economic and social activity was – and continues to be – affected. In Singapore, people's faith in the health and safety of public transportation have also been harmed as a result of the pandemic, and they are being more careful and prioritizing a safe distance during commuting. This results in the need to view the spatial relation of commutes between subzones and predict commuter flows to help the Land Transport Authority of Singapore to strategise and ease the planning process of the bus infrastructure and timing in Singapore.

## 1. MOTIVATION

There is presently no analytics application that looks at public transportation commuter patterns across time and space. We hope that creating this application can bring benefits to multiple stakeholders: by allowing commuters to better plan their journeys, workplaces to effectively implement flexible working hours, and the relevant governmental organisations to implement measures to address peak hour crowds.

Since the pandemic, even though there are people who are less willing to take the public transport, there is still a crowd in ridership at certain times. An example would be when Mass Rapid Transport (MRT) services broke down in October 2020, this resulted in bus stops being packed with commuters looking for alternatives to get home. This is a cause for concern for commuters who are conscious of COVID-19 risks as it is tough to maintain a safe distance from everyone around them.

The government has implemented measures to address peak hour crowds, including increasing the frequency of buses and trains during these periods. While this does solve the issue, we wonder about the sustainability of increasing the frequency of public transport - particularly buses, when public transportation ridership is going to increase over the years.

## 2. REVIEW & CRITIC ON PAST WORKS

### 2.1 Commute Explorer in New Zealand

In his Shiny Application, Stefan Schliebs explores the commuting behavior of New Zealanders. Other than well-designed and intuitive maps, his app includes filtering options such as by work and education so that users are able to drill down into places of interest and understand the travel mode pattern. There are also other features such as buckets that allow the selection of specific regions of New Zealand and histograms of mode of travel by the commuters that allows users to see the distribution across the various places of residence and work. Although the scope of this application is not related to Singapore, this application demonstrates excellent use of interactive visualization with great colour management will lead to an attention grabbing information and trends on commuting behavior that will be useful and relevant in our analysis for the public transportation commuter patterns in Singapore.

### 2.2 Bus Transport Network in Singapore

Another Shiny application that can be found within the local context of Singapore is the Network Analysis of Singapore Bus Transport done by Group 8 of ISSS608 Visual Analytics 19/20 class. When analysing the flow of bus traffic from region to region, their application makes use of a Flow Map and an Origin Destination Matrix. These features will be relevant in our analysis as we seek to understand the movement of people from region to region.

The cons of these two applications are as follows:

- Unable to view the spatial relation of commutes between certain regions
- Unable to predict commuter flows

## 3. DESIGN FRAMEWORK

This application focuses on interactivity. To do that, it has features such as filtering, zooming in and out, hovering over and displaying significant details.

The main tabs of this application comprises of Exploratory Data Analysis (EDA) such as Spatial Point Map, Time-series Choropleth Maps and Desire Lines Map, as well as Spatial Interaction Models (SIMs) using Generalised Linear Models such as Unconstrained, Origin constrained, Destination constrained and Doubly Constrained Spatial Interaction Models.

Under the EDA tab, the Spatial Point Map shows the distribution of bus stops in Singapore. Filters such as colour and transparency of points are provided. Time-series choropleth maps reveal any outliers of the trips by origin or destination. Filters such as colour scheme, origin/destination, day type and hour/monthly interval and peak hours are provided. Users are able to select the filter by day type, which is either "Weekday" or "Weekend." Also, for the filter on hour/monthly interval, if user selects, "Month," 3 choropleth maps will be displayed (since this study takes data of months June, July and August). If "Hour" is selected, 24 choropleth maps will be shown for the hours 0 - 23. Next, not all hours are shown as there would not be any significant results. Hence, only peak hours are shown here. Lastly, the Time-series Desire Lines Map shows the movement of people from an origin to destination (Desire Lines - Piggraphy Alevel Skills, 2021). This is denoted by the number of trips. The filters are by day type, hour/month interval and peak hours. Explanations for the filters are similar as the one explained for the time-series choropleth maps.

Lastly, for the SIM tab, a scatterplot of the results of the various models are displayed. The filters provided are colour of points, types of model as well as origin and destination, where the options of origin and destination are population and income. A conditional filter is done here. If "Unconstrained" for types of model is selected here, the Origin and Destination filter will appear. If "Origin" is selected here, the Destination filter will appear. If "Destination" is selected, "Origin" filter will appear. Lastly, if "Doubly" is selected, it will display as it is and no further filters will be provided. At the bottom of the chart, Users can toggle between the different tabs at the bottom to view a summary and results of the selected model, the original matrix as well as the SIM matrix.

## 4. DEMONSTRATION - USE CASE

There are main tabs that the user can click on, they are namely: About, EDA and SIM. The About page contains a brief introduction to our project while the EDA tab contains our respective maps for the purpose of exploratory data analysis. The SIM data contains the results of the respective spatial interaction models we have performed.

The user can click on the EDA main tab to view 3 separate tabs, which are namely Spatial Points, Choropleth Maps and Desire lines Maps.

- Spatial Points
  - The user is able to view the geospatial distribution of bus stops in Singapore, with OpenStreetMap as the backdrop. The user can choose to zoom in and out to view the details of the respective bus stops. The user is also able to see a data table, depicting the details of the respective bus stops point data in table form.
- Choropleth Maps
  - If filters are Origin/Destination: Origin/Destination and Day Type: Weekday or Weekends/Holidays and Hour/Month Interval: Month
    - There will be 3 maps showing the origin/destination subzones which are outliers during Weekdays or Weekends/Holidays in June, July and August respectively.
  - If filters are Origin/Destination: Origin/Destination and Day Type: Weekday or Weekends/Holidays and Hour/Month Interval: Peak Hour
    - There will be 3 maps showing the origin/destination subzones which are outliers during Weekdays or Weekends/Holidays in the morning, afternoon and evening peak hours (7-9am, 10am-12pm, 5-7pm) respectively.
  - If filters are Origin/Destination: Origin/Destination and Day Type: Weekday or Weekends/Holidays and Hour/Month Interval: Both
    - There will be 9 maps showing the origin/destination subzones which are outliers during Weekdays or Weekends/Holidays in the morning, afternoon and evening peak hours (7-9am, 10am-12pm, 5-7pm), in the months of June, July and August.
- Desire lines Maps
  - If filters are Day Type: Weekday or Weekends/Holidays and Hour/Month Interval: Month
    - There will be 3 maps showing the movement of people from origin to destination subzones during Weekdays or Weekends/Holidays in June, July and August respectively.
  - If filters are Day Type: Weekday or Weekends/Holidays and Hour/Month Interval: Peak Hour
    - There will be 3 maps showing the movement of people from origin to destination subzones during Weekdays or Weekends/Holidays in the morning, afternoon and evening peak hours (7-9am, 10am-12pm, 5-7pm) respectively.
  - If filters are Day Type: Weekday or Weekends/Holidays and Hour/Month Interval: Both
    - There will be 9 maps showing the movement of people from origin to destination subzones during Weekdays or Weekends/Holidays in the morning, afternoon and evening peak hours (7-9am, 10am-12pm, 5-7pm), in the months of June, July and August.

The user can click on the SIM main tab to view 5 separate tabs, which are namely the SIM Scatterplot, SIM Summary, Original Matrix, SIM Matrix and SIM Results. The output of these pages will change according to the use cases stated below:

- Unconstrained, Origin: Population/Income, Destination: Income/Population
  - We will see the results of the unconstrained spatial interaction model which takes in the origin income/population and destination income/population, reflected in the 5 tabs mentioned above.
- Origin, Population/Income: Population/Income
  - We will see the results of the origin constrained spatial interaction model which takes in the origin subzone code and destination population/income, reflected in the 5 tabs mentioned above.
- Destination, Population/Income: Population/Income
  - We will see the results of the destination constrained spatial interaction model which takes in the origin population/income and destination subzone code, reflected in the 5 tabs mentioned above.
- Doubly
  - We will see the results of the doubly constrained spatial interaction model which takes in the origin and destination subzone codes, reflected in the 5 tabs mentioned above.

# 5. RESULTS & DISCUSSION

## 5.1 Exploratory Data Analysis

### *5.1.1 Spatial Points Map*

From plotting the spatial points of the bus stops data, there were 5 bus stops that were found outside of Singapore's boundary. As travelling to and fro Johor Bahru is possible with specific buses in Singapore, there are designated bus stops located at Johor Bahru. Hence, these bus stops are removed from the analysis.

### *5.1.2 Choropleth (Percentile) Maps*

Percentile maps are variations of common choropleth maps where the classification is designed to highlight extreme values at the lower and upper end of the scale, with the goal of identifying outliers.

By Origin, Weekday or Weekends/Holiday, Month

We can observe that subzones which are lower and upper outliers did not differ across the 3 months. From June to August, we can observe that there are 2 upper outliers in the eastern part, 1 upper outlier in the northern part and 1 upper outlier in the western part of Singapore.. Similar trends are seen in both weekdays and weekends/holidays.

By Destination, Weekday or Weekends/Holiday, Month

We can observe that subzones which are lower and upper outliers did not differ across the 3 months. Across June to August, we can observe that there are 2 upper outliers in the eastern part, 1 upper outlier in the northern part and 1 upper outlier in the western part of Singapore. Similar trends are seen in both weekdays and weekends.

By Origin, Weekday or Weekends/Holiday and Peak Hour

Our peak hours are defined to be 7-9am, 10am-12pm and 5-7pm. We can see slight changes in the subzones which are upper outliers across the different peak hours, but generally they are located in similar areas - the eastern, northern and western part of Singapore. Similar trends are seen in both weekdays and weekends.

By Destination, Weekday or Weekends/Holiday and Peak Hour

We can see slight changes in the subzones which are upper outliers across the different peak hours. In addition, we can observe that the subzones which are upper outliers have changed slightly as compared to the origin and peak hour choropleth maps - there is now 1 upper outlier in the central area. Similarly, there are upper outliers in the eastern, northern and western part of Singapore. Not much difference can be observed across weekdays and weekends.

By Origin, Weekday or Weekends/Holiday, Month and Peak Hour

We found that across the 3 months and 3 peak hours, the subzones which are upper outliers are generally the same for peak hours. Meanwhile, the subzones which are upper outliers differ slightly across the 3 months. Not much difference can be observed across weekdays and weekends

By Destination, Weekday or Weekends/Holiday, Month and Peak Hour

We found that across the 3 months and 3 peak hours, the subzones which are upper are generally the same for peak hours. Meanwhile, the subzones which are upper outliers differ slightly across the 3 months. Not much difference can be observed across weekdays and weekends

### *5.1.3 Desire Lines*

A desire line is a straight line on a map representing the movement of people from region to region. For the purpose of our analysis, we focused on only the top 2% of the desire lines for all the maps below.

By Month

Based on the results of the 3 maps of the months, there is a lower concentration of desire lines in June as compared to July and August. The difference can be observed from the lesser concentration of desire lines in the West and East region. However, not much difference can be seen from the months of July and August..

By Peak Hour

Out of the 3 categories of peak hours, the afternoon peak hour has the least amount of desire lines as compared to the morning peak hour where the desire lines are more concentrated. The desire lines in the afternoon peak hour are also sparse, and can only be observed in the North and East region and lesser at the Central region. For the afternoon peak

hour, the people around the CBD region do not travel by bus for their lunch and may have travelled by other modes of transportation like car or walking.

By both Month and Peak Hour

From our analysis, we found that across the 3 months and 3 peak hours, the pattern of the desire lines are generally the same for peak hours. The peak hour with the highest concentration of desire lines can be found in the evening peak hour while the lowest concentration is found in the afternoon peak hour. Across the months, minimal changes are observed whereby the concentration of desire lines in June is the least compared to the desire lines in August. This means that more people are commuting by bus and may be due to the ease of COVID restrictions.

## 5.2 Spatial Interaction Models

The aim of building Spatial Interaction Models in our analysis is to explain commuter flows from the spatial perspective. Commuter flows are regarded as an interaction between origins and destinations. There are four types of model created, Unconstrained, Origin Constrained, Destination Constrained and Double Constrained.

### *5.2.1 Unconstrained Model*

2 main variables were used in the Unconstrained Model, the population and median income. The R-squared value is 0.353 which is relatively low, implying that the unconstrained model failed to fit the empirical data well.

### *5.2.2 Origin-Constrained Model*

For Origin Constrained, the R-squared value is 0.622 when using population as the Destination variable. However, when income is used as the Destination variable, the R-squared value obtained was lower, about 0.547. Overall, we can observe that there is considerable improvement from 0.353 in the Unconstrained model.

### *5.2.3 Destination-Constrained Model*

For Destination Constrained, the R-squared value is 0.622 when using population as the Origin variable. However, when income is used as the Destination variable, the R-squared value obtained was lower, about 0.531. It is also observed that there is improvement from the Unconstrained model however, the R-squared value is generally lower than the Origin Constrained model.

### *5.2.4 Doubly-Constrained Model*

- Model: Origin Code, Destination Code;
  - R-squared value: 0.74
- Model: Origin Code, Destination Code, Origin Population, Destination Income;
  - R-squared value: 0.18
- Model: Origin Code, Destination Code, Destination Population, Origin Income;
  - R-squared value: 0.18
- Model: Origin Code, Destination Code, Origin Population, Destination Income, Destination Population, Origin Income;
  - R-squared value: 0.18

The Doubly constrained model with the highest R-squared value is the model with just the subzone codes as the variables with a R-squared value of 0.74. This value is also the highest compared to the other 3 models - Unconstrained, Origin constrained and Destination constrained. This means that variables like population and median income can be used to explain the flow between origin and destination however they are not the best variables to be used since the Doubly Constrained model has the highest R-squared value.

## 5.3 Usability of Application

This application is made to be as interactive as possible. Such features include being able to zoom in and out in maps as well as to select different filters based on what users are interested in.

Taking Jakob Nielsen's 10 heuristics for user interface design as a reference, this application has fulfilled (World Leaders in Research-Based User Experience, 2021):

- Consistency and standards
  - Filters for choropleth maps and desired lines map are similar, namely, being able to display the maps by months, by peak hours, or by both months and peak hours.
- Aesthetic and minimalist design
  - All tabs are designed to only display the necessary filters and maps with useful insights.
- Help and documentation
  - A user guide is linked to the About tab.
  - In the Choropleth Maps and Desired Lines Maps tab, the peak hours are made transparent to the users. This is shown as the map title.
  - A short explanation on using the SIM tab is at the top left corner
  - A short explanation at the bottom of the SIM Scatterplot tab to explain what Spatial Interaction Models are.

# 6. FUTURE WORK

## 6.1 Feature for users to upload their preferred origin and destination of bus trips data

Currently, the study focuses on the origin destination of bus trips data in the months of June, July and August in 2021. Hence, the results that users are seeing are only accurate to the aforementioned months. If there are opportunities to enhance this application, giving the users the option to upload their own data would be implemented. This way, users are able to use this application to determine the results of their preferred origin and destination of bus trips datasets.

## 6.2 Cover other modes of transport

In addition to only focusing on the origin and destination of bus trips data, this study can also be refined to cover other modes of transport like Mass Rapid Transport in Singapore.

# 7. REFERENCES


- Group08 application. Group08 application - ISSS608-Visual Analytics and Applications. (n.d.). Retrieved November 18, 2021, from https://wiki.smu.edu.sg/1920t2isss608/Group08_application.
- Harrison, J. (2020, October 21). Commentary: Singapore's new headache of fewer public transport commuters. CNA. Retrieved November 18, 2021, from https://www.channelnewsasia.com/commentary/public-transport-mrt-bus-capacity-breakdown-covid-car-share-968131.
- Lta.gov.sg. LTA. (n.d.). Retrieved November 18, 2021, from https://www.lta.gov.sg/content/ltagov/en/newsroom/2014/7/3/steps-taken-to-reduce-bus-crowding-during-peak-hours.html.
- Morning pre-peak fares. Public Transport Council. (n.d.). Retrieved November 15, 2021, from https://www.ptc.gov.sg/fare-regulation/bus-rail/morning-pre-peak-fares.
- RStudio Shiny Contest 2021 - Commute Explorer. Data Science Academy. (n.d.) Retrieved November 15, 2021, from https://nz-stefan.shinyapps.io/commute-explorer-2/
- Spreading out the peak hour MRT crowd. Ministry of Transport. (n.d.). Retrieved November 15, 2021, from https://www.mot.gov.sg/Transport-Matters/public-transport/Detail/spreading-out-the-peak-hour-mrt-crowd.
- World Leaders in Research-Based User Experience. (April 24, 1994). 10 usability heuristics for user interface design. Nielsen Norman Group. Retrieved November 15, 2021 from https://www.nngroup.com/articles/ten-usability-heuristics/.
- Sites.google.com. Desire Lines - Piggraphy Alevel Skills (n.d.). Retrieved November 15, 2021 from https://sites.google.com/site/piggraphy/home/cartographic-d/desire-lines ...