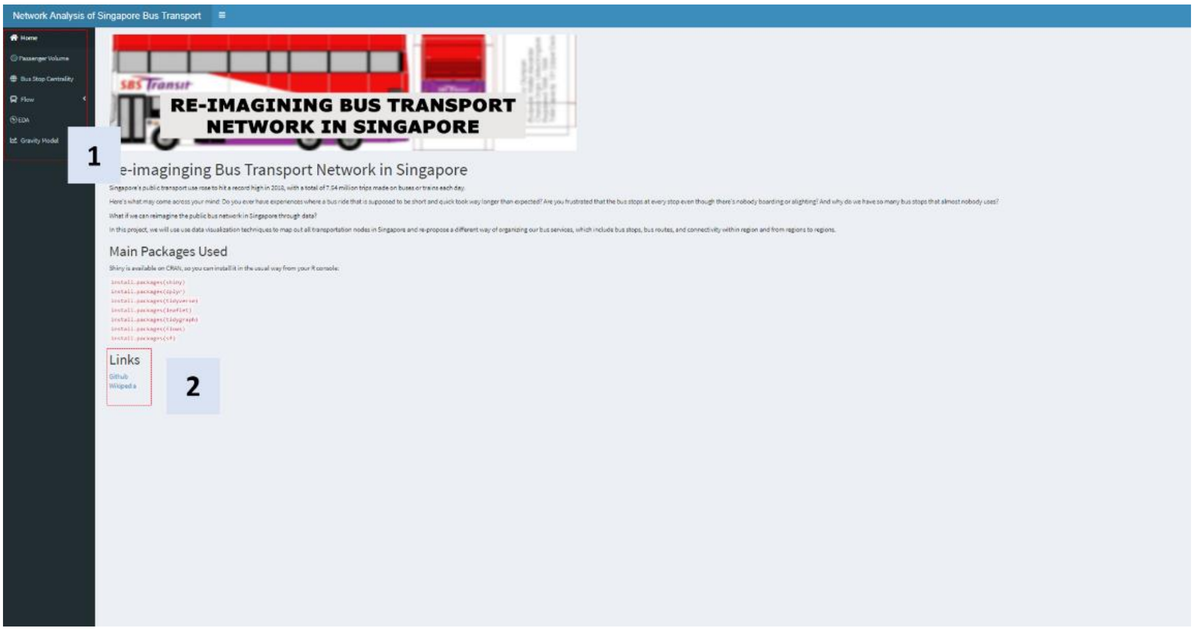
Re-imagining Bus Transport Network in Singapore – User Guide

Tab 1 – Home Page

The Home tab is the landing page of the app. It gives us a brief overview of the product, as well as the packages used and link to Github and Wikipedia.

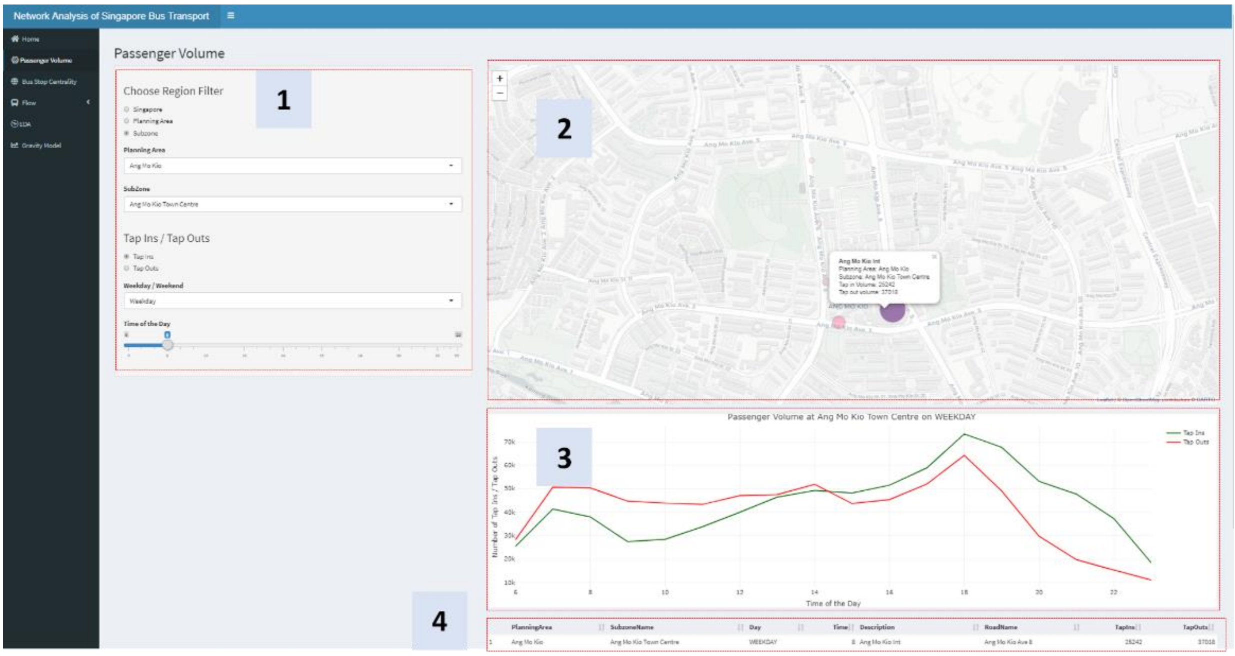


[1] On the left, you can find the navigator panel. There are 6 tabs in total, the Flow tab can be expanded to reveal 4 more tabs.

[2] The links to the project Github repo and wiki can be found here. You can find the codes to the R Shiny app on the Github link, and the Project Proposal, Research Paper, User Guide, and Poster in the Wiki link.

Tab 2 – Passenger Volume

This tab shows the passenger flow by Bus Stop, across Region (Singapore / Planning Area / Subzone), Weekday/Weekend, and Time of the Day. User view either Tap Ins or Tap Outs as the size of the node on the app.



[1] There are 4 filters in this visualization, Region, Tap Ins/ Tap Outs, Weekday / Weekend, and Time of the day from top to bottom respectively. The Region filter is a nested filter, choose which granularity of region filter (Singapore/Planning Area/Subzone), and after, choose the region you want. Data will be filtered to the input stated in the filters, and the maps and line graph below will change accordingly.

[2] The size and color intensity of the node represents the number of Tap Ins or Tap Outs at the chosen region, weekday/weekend, or time of the day. Interactivity is embedded into the map visualization, by brushing across the node, the name of the bus stop will appear. Full details, including name of the bus stop, planning area, subzone name, and tap in and tap out volume will appear when clicked on.

[3] The line chart below the map shows the passenger volume at the selected region by time of the day. The line chart is interactive, by hovering your mouse over the line, details on the number of tap ins / tap outs will show.

[4] The table below the line chart shows the data used to create the map and line chart. The table is sorted according to number of Taps Ins to show the highest passenger volume bus stop of from the selected data.

Tab 3 – Bus Stop Centrality

This tab shows the Bus Stop Centrality measure across Region (Singapore / Planning Area / Subzone). There are four centrality measures that we have included in this visualization, namely Betweenness Centrality, Degree Centrality, Closeness Centrality, and Eigenvalue Centrality.

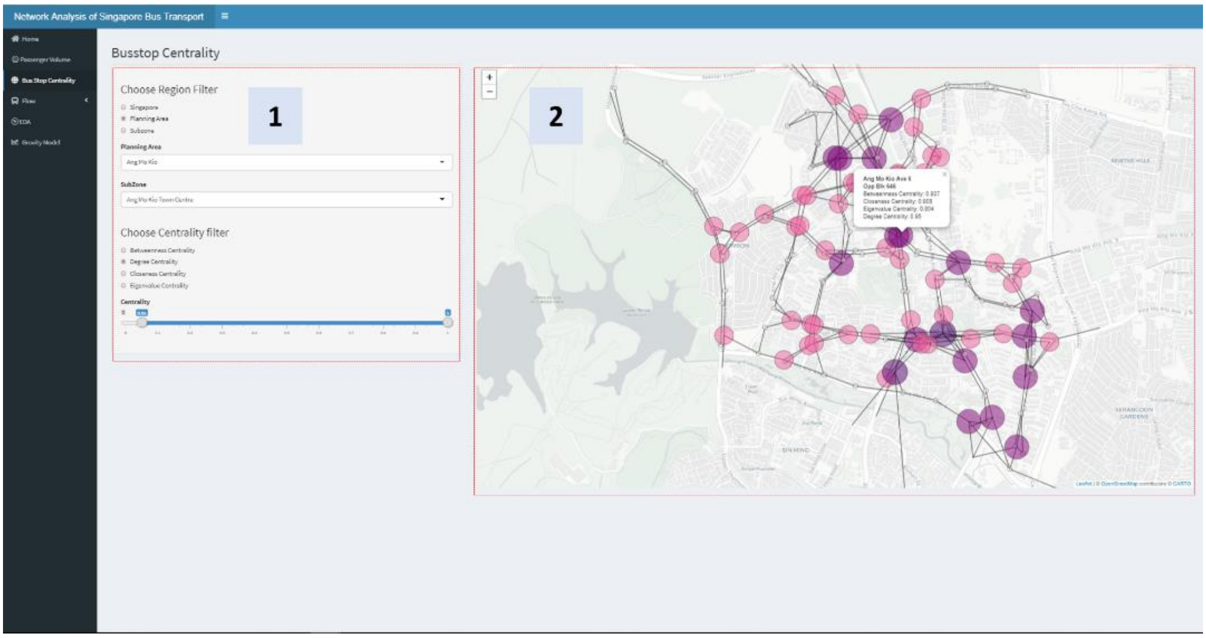
Definitions:

Betweenness Centrality – Defined as the average proportion of paths between as two nodes within the network that transverse the node in question, out of the total number of possible paths between these two nodes.

Degree Centrality – Defined as the proportion of nodes directly connected to the node in question out of the totality of nodes within the network.

Closeness Centrality – Defined as the inverse average distance between node in question and all other nodes within the network.

Eigenvalue Centrality – Defined as the measure of the influence of a node in a network. Relative scores are assignment to all nodes in the network based on its connections, the more important the connections, the higher the score.



[1] There are 2 filters in this visualization, Region and centrality. The Region filter is a nested filter, first choose which granularity of region filter (Singapore/Planning Area / Subzone), and after, choose the region you want. The centrality filter consists of a radio button and a scroll bar. Use the radio button to select the centrality measure you want to be displayed as the size of the nodes on the map, and the scroll bar to filter to a specific centrality range.

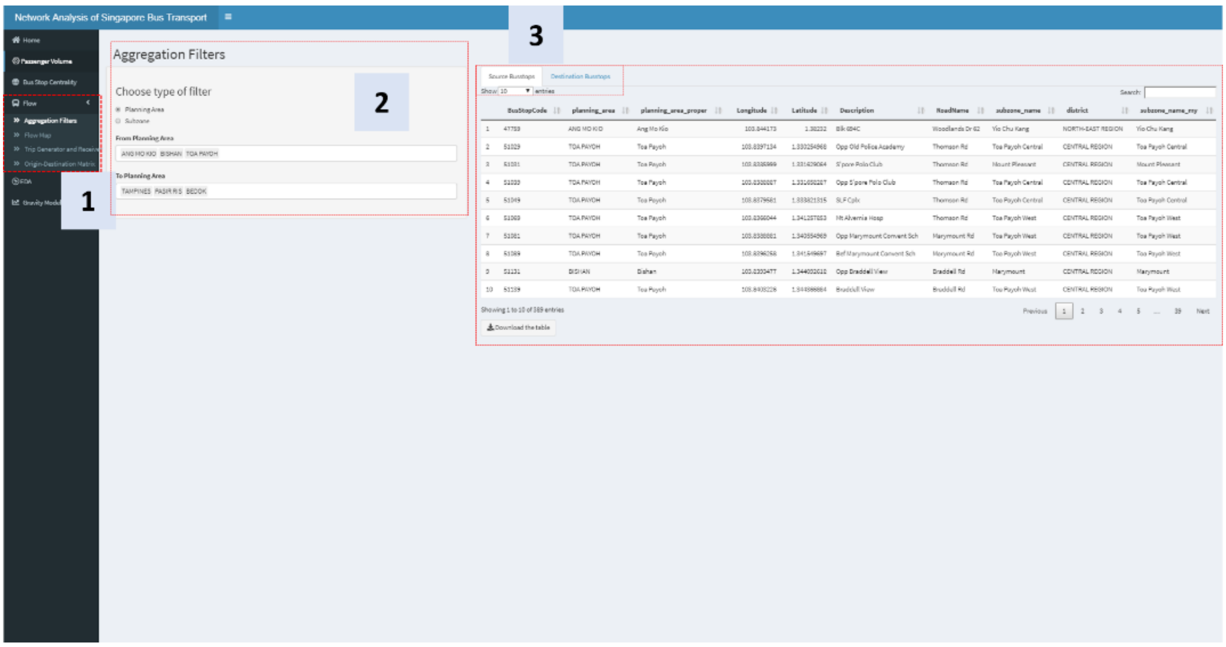
Data will be filtered to the input stated in the filters, and the maps will change accordingly.

[2] The size and color intensity of the node represents the selected centrality measure. Interactivity is embedded into the map visualization, by clicking on the node, full details including name of the bus stop, road name, and all four centrality measures will appear.

Tab 4 – Flow

Tab 4.1 – Aggregation Filter

This tab shows and analyze the Flow of passenger from region to region. The first sub tab 'Aggregation Filters' filters, aggregate and prepare the data for later analysis.



[1] After clicking on the Flow tab, 4 sub tabs will appear.

Important : 'Aggregation Filters' is the first step towards further flow analysis, it should first be filled before proceeding to the other tabs under Flow. The 2 tables produced are downloadable with the " Download table" button in '3' in figure shown.

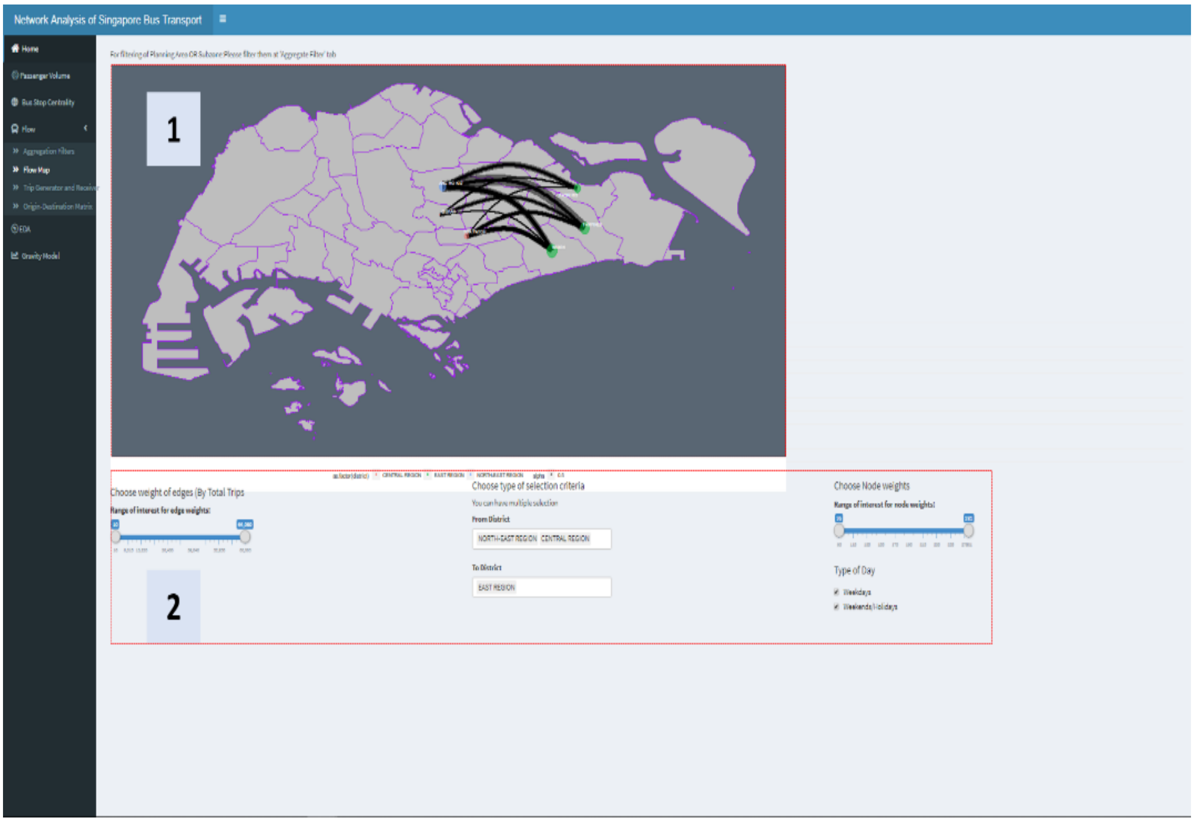
[2] Since we are looking at the flow of passenger from region to region, first we must choose which the Source and Destination of the flow. The filters consist of a radio button for 'Planning Area' / 'Subzone'. If 'Planning Area' is chosen, we have multiple select input options for 'From Planning Area' (Source) and 'To Planning Area' (Destination). If you choose 'Subzone', we have multiple select input options for 'From Subzone' and 'To Subzone'. You can input multiple regions in both 'From Planning Area' and 'To Planning Area' filter.

[3] Once you have done the selection from

[4] A data table here shows your selected Source Busstop and Destination Busstop. There is also a download button below the table to download the table if the user want to see the input data to further analysis in tabular form.

Tab 4.2 – Flow Map

Next, after filling in the inputs at Aggregation Filter, the map on Flow Map will be populated on the second tab of Flow.

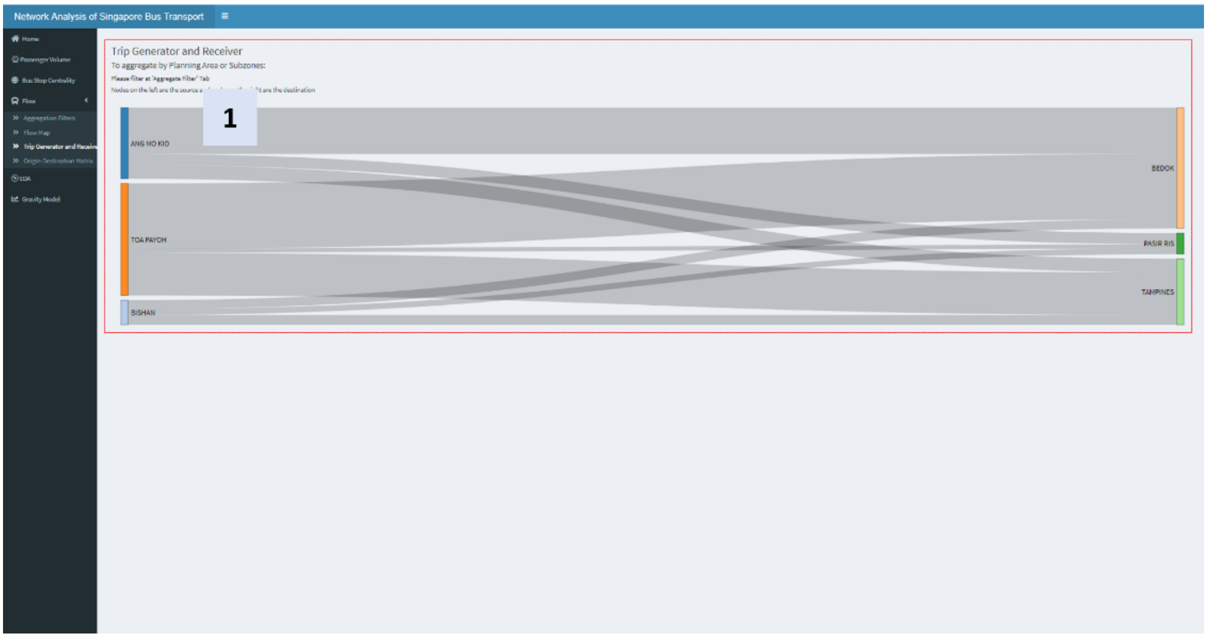


[1] The flow map shows the movement of people from region to region. The nodes are colored blue for Source and Green for Destination, the size of the nodes represents the number of people traveling to/from the regions. The thickness of the line corresponds to the volume of travel between the regions.

[2] Below the map there are filters for district, edge weights, node weights, type of day, and from/to district. Especially when the user choose many source to many destination which will create a noisy map, the filters would be helpful providing clarity. All the filters will be populated at full, users can use the filters or even filter away districts that they are not interested in. Districts are populated to the fullest. There can be multiple districts in source and destination nodes.

Tab 4.3 – Trip Generation and Receiver

This third tab of Flow shows a Sankey diagram of the selected Source and Destination from the Aggregation Filters tab. Where left is the source, right is destination.

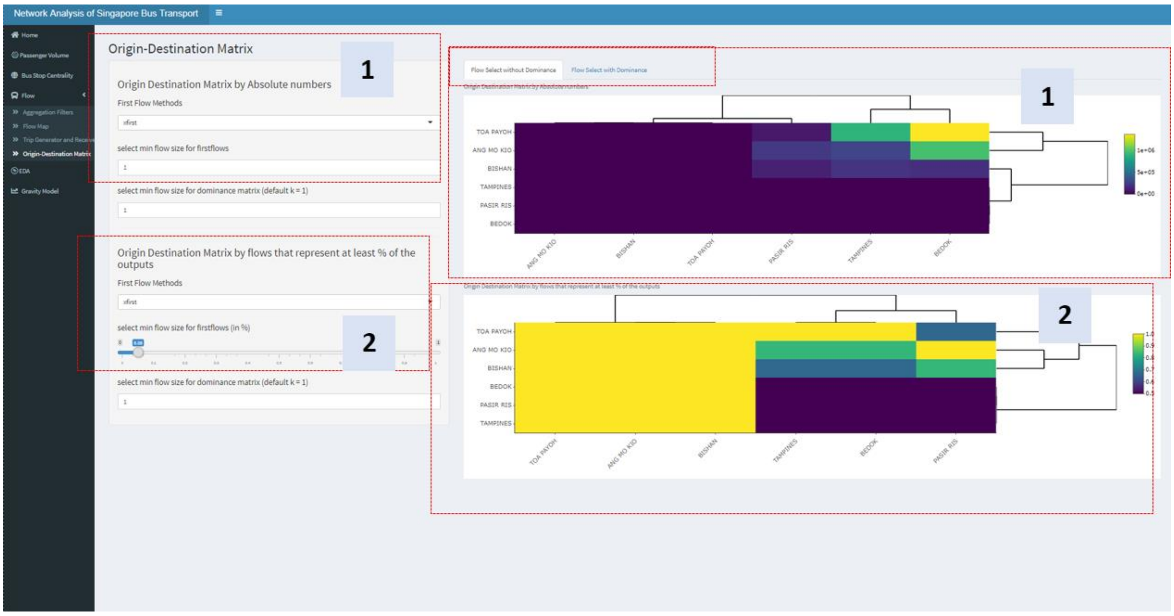


[1] The Sankey Diagram shows the flow of passengers from Source to Destination, with the width of the flow being the proportion of volume of passengers traveling from Source to Destination. This diagram is interactive, you can drag both the source or destination staring point up and down to rearrange the diagram.

Tab 4.4 – Sub Tab: Origin Destination Matrix (Flow Select without dominance)

In the final tab of 'flow', we have the origin destination matrix.

There are 2 sub tabs namely "Flow Selection without Dominance" and "Flow Selection With Dominance"

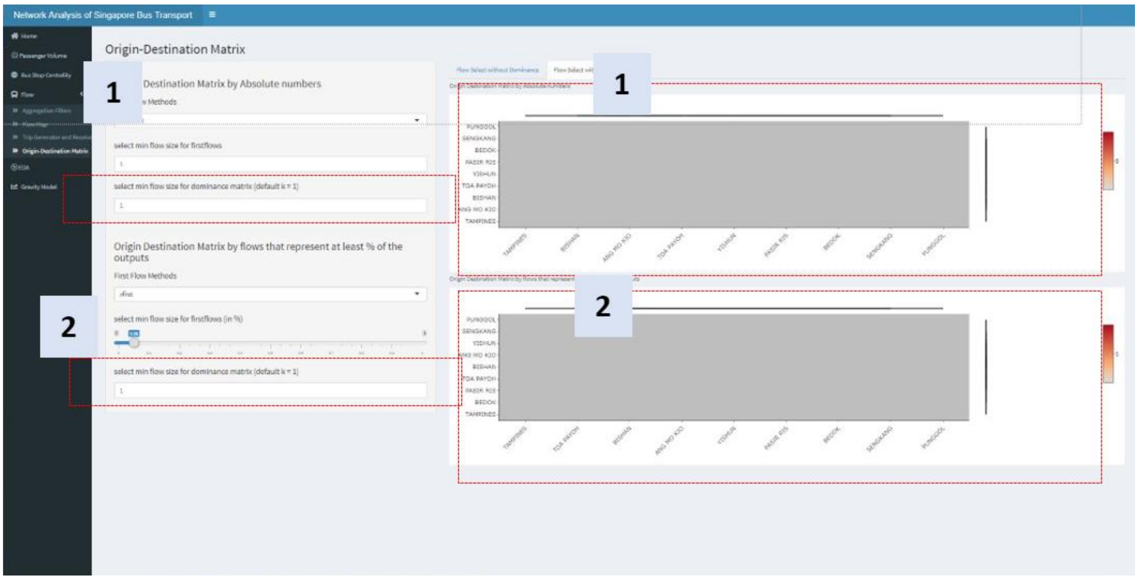


[1] On the left you can find filters relating to First Flow methods, minimum flow size for first flow, and select minimum flow size for dominance matrix.

[2] In the 'Flow Select Without Dominance' tab, the first 2 filters located in '1' on the left is to filter to get the matrix M on the right indicated as '1'. SImilarly for the filters in '2', it is meant for the matrix in '2' on the right. M = myflow * firstflows

Firstflows consists of 3 types namely : nfirst, xfirst and xsumfirst. Nfirst = the k first flows from all origins; xfirst = all flows greater than a threshold k; xsumfirst= as many flows as necessary for each origin so that their sum is atleast equal to k.

Tab 4.4 – Sub Tab : Origin Destination Matrix (Flow Select with dominance)



[3] After you filtered the filters mentioned in [1], you can proceed my filtering the filters in this image above, where '1' on the left is the filter for '1' on the right and '2' on the left is the filter for '2' on the right. M = myflow * firstflows * dominantflows. There are 3 components for the M matrix in "Flow Select with Dominance" that is why we have 3 filters for each Origin-Dominance matrix. It is natural to have empty plot because the dominance matrix in "dominantflows" is a sparse matrix.

Tab 5– EDA

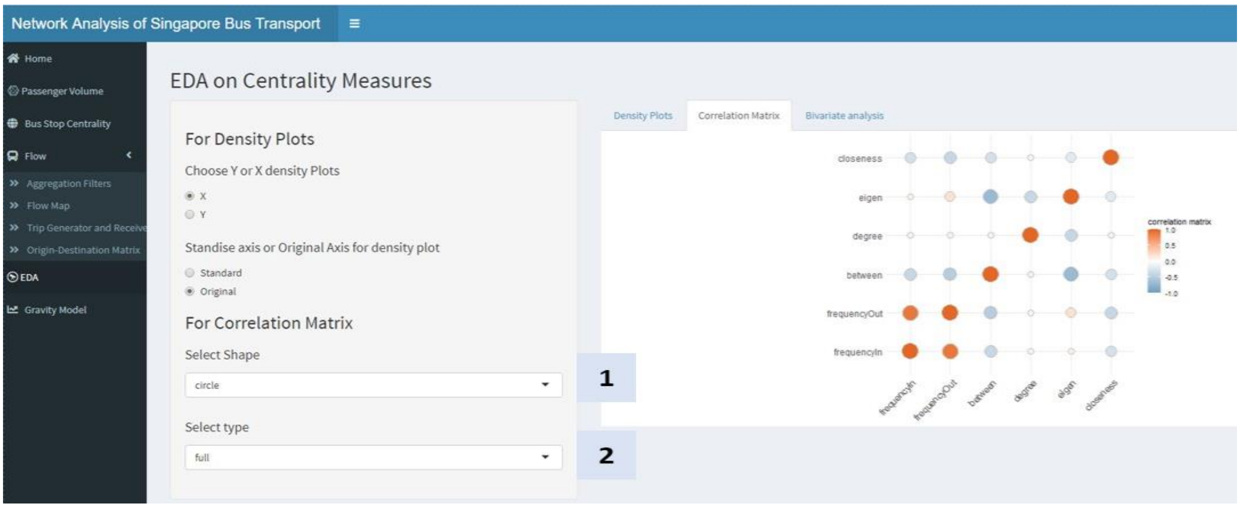
Tab 5.1 Density plot with Histogram

There are 3 subtabs.

Location 1: Density Plot with Histogram, Location 2: Correlation Matrix. Location 3: Bivariate Analysis scatterplot

[1] For location 1 on the picture and corresponding tab 1. On the sidebar, users can select to view density plot for the independent (X) variables or for the dependent (Y) variables. After that, users can select "Standard" to standardize the Y axis or select "Original" to get the original plot.

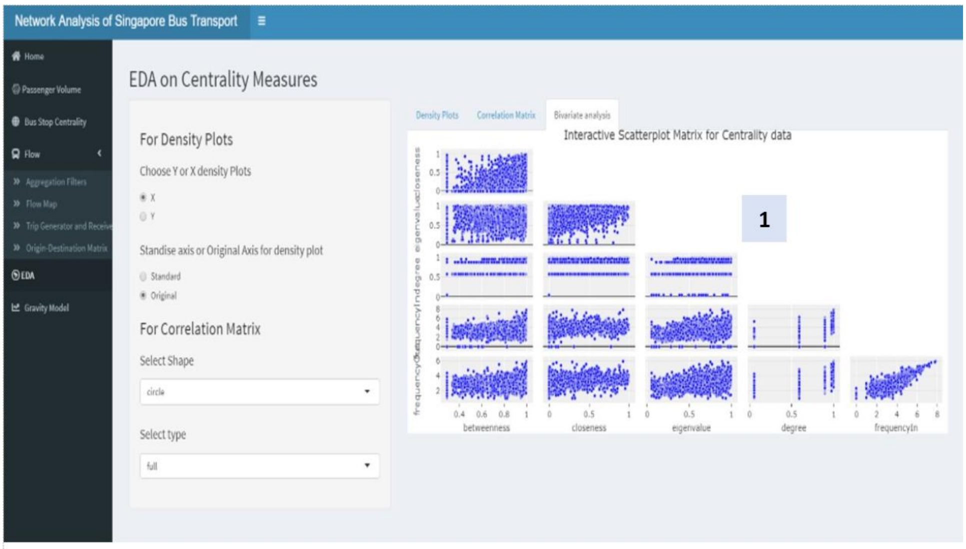
Tab 5.2 Correlation Matrix



[1] Tab 2 is the correlation matrix. Users can select under "Select Shape" to change the shape of the correlation matrix from circle to sqaure.

[2] Users can select under "Select Type " to visualise the correlation matrix as full, upper or lower.

Tab 5.3 Sub tab- Interactive Scatterplot

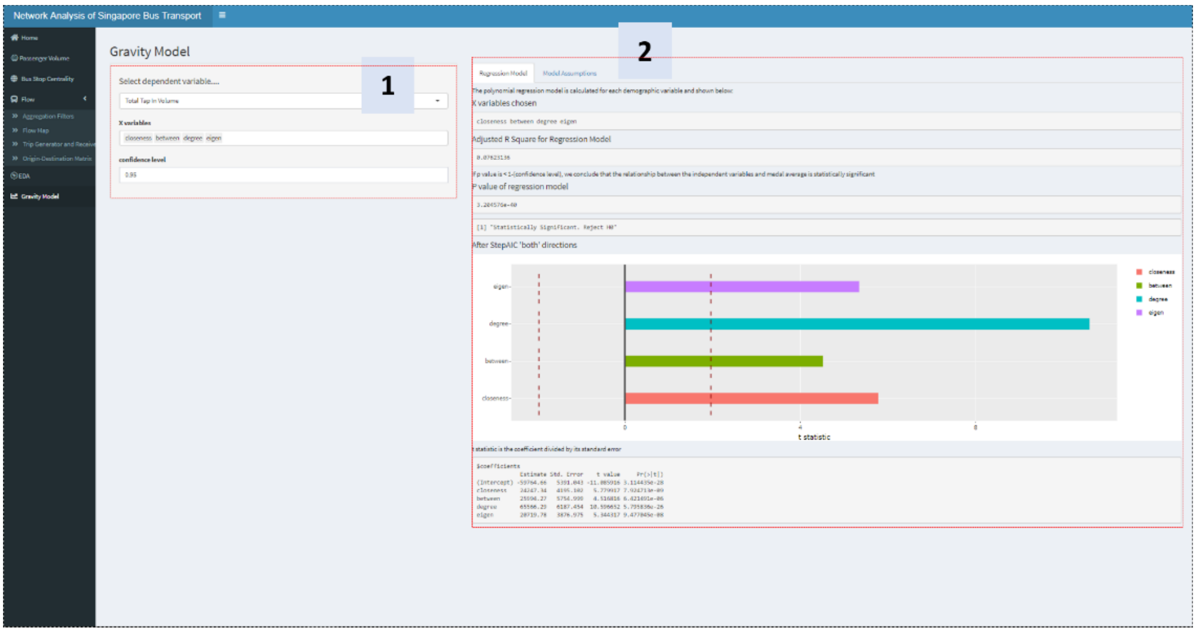


[1] User can hover over the points of the scatterplot to see the exact data. There are no filters on the sidepanel for this plot.

Tab 6 – Gravity Model

Tab 6.1 Sub tab – Regression Model

The Gravity Model plots a regression analysis of Centrality measures (X, Dependent Variable) against Tap in or Tap out (Y, Independent Variable). There are 3 common filters for the 2 subtabs namely: 1. Choose dependent Y variable under "Select Dependent Variable", choose independent variable under "X variables" and select the confidence level for regression model and Durbin watson's test under "confidence level".



[1] On the filter, user can choose the X variable, Y variables, and the confidence level. This confidence interval applies to both the hypothesis for regression model in current tab where the test is test for significance of regression: $H0 : \beta 1 = \beta 2 = \cdots = \beta k = 0$; $H1 : \beta j \neq 0$ for at least one $j \neq 0$, as well as the Durbin Watson's test under the second sub tab 'Model Assumptions'

[2] On the top of the result panel, user can toggle between Regression Model and Model Assumptions. Based on the confidence level chosen, the model will show if the model is statistically significant.

The coefficient of the model can also be found below. Since we normalized all centrality measures, the weight of the coefficient can tell us the importance of the centrality measure to the overall passenger flow. In the case above, we can see that degree centrality is the most important centrality measure (of the transformed model based in the report).

Tab 6.2 Sub tab – Model Assumption

Lastly, we did some testing using Durbin Watson test to detect the presence of autocorrelation.

[3] The filters in this current tab is same as that of “Regression Model” tab.