GPExploreR: Singapore Government Procurement Analysis

Application User Guide

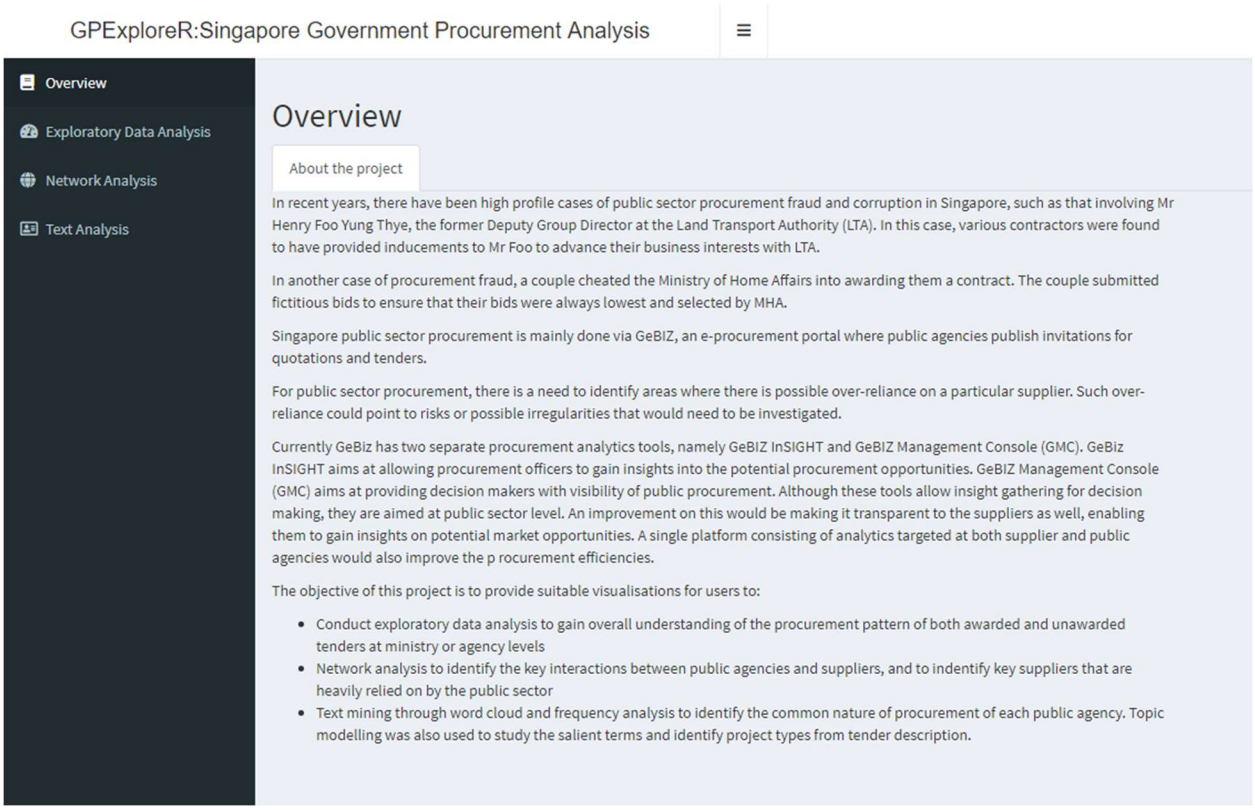
Contents

1. Overall Navigation

When users launch the App, the first screen displayed is the Overview tab which allows users to have an overall understanding of the project background.

The app contains 3 other main sections and each section has subtabs embedded. Users can select on the main section on the left sidebar panel, the subtabs are displayed under the main section title. The layout is as shown below:

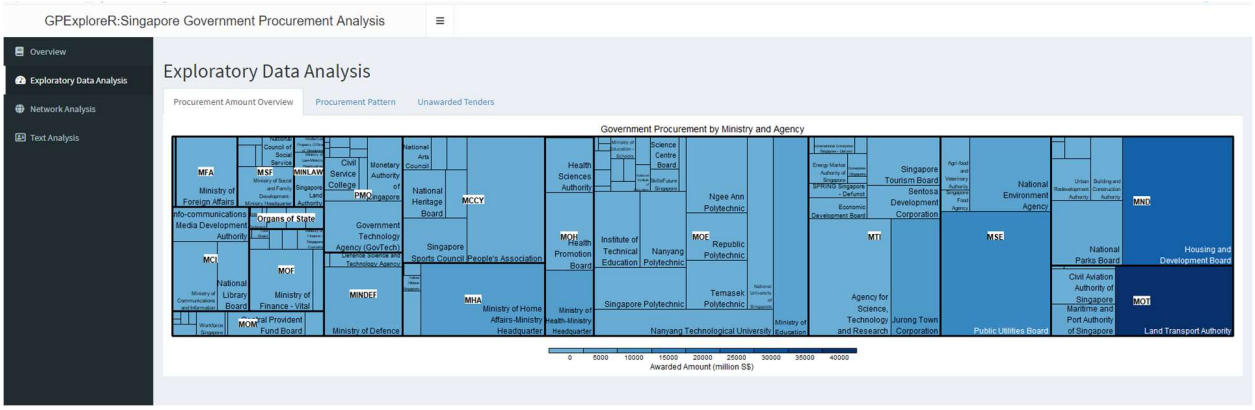
- Exploratory Data Analysis
    - Procurement Amount Overview
    - Procurement Patterns
    - Unawarded Tenders
- Network Analysis
    - Sankey Diagram
    - Network Diagram
- Text Analysis
    - Text Analysis of Tender Projects by Agency
    - Text Analysis of Tender Projects by Suppliers
    - Keywords by Project Type using LDAvis



2. TABS

2.1 Exploratory Data Analysis

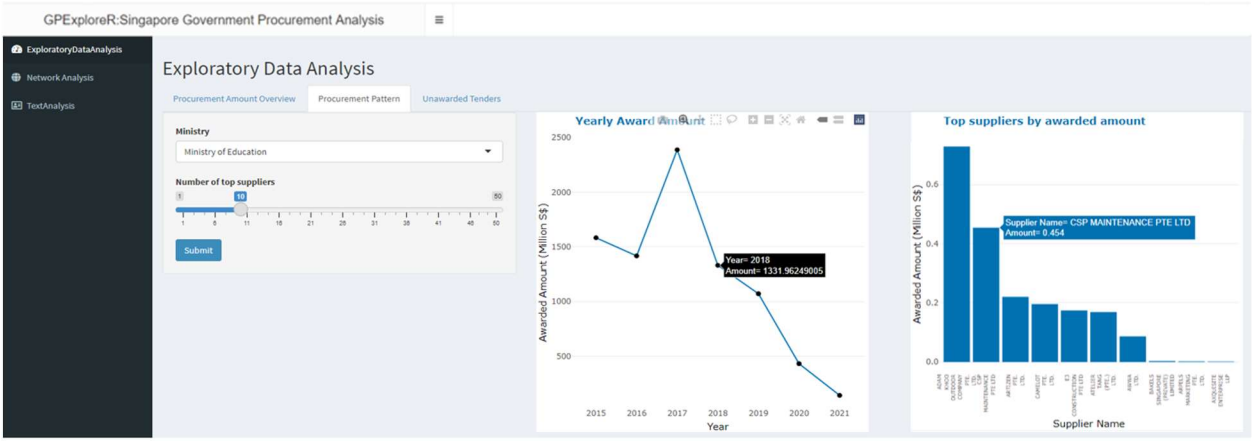
2.1.1 Procurement Amount Overview



On clicking on the "Procurement Amount Overview" tab item, the treemap showing procurement amount by ministry would appear on the main panel.

The size of the tiles represents the number of orders placed by each ministry or agency, the color of the tiles represents the total procurement amount during the time period. The legend below treemap shows the range of awarded amount in different colors.

2.2.2 Procurement Pattern

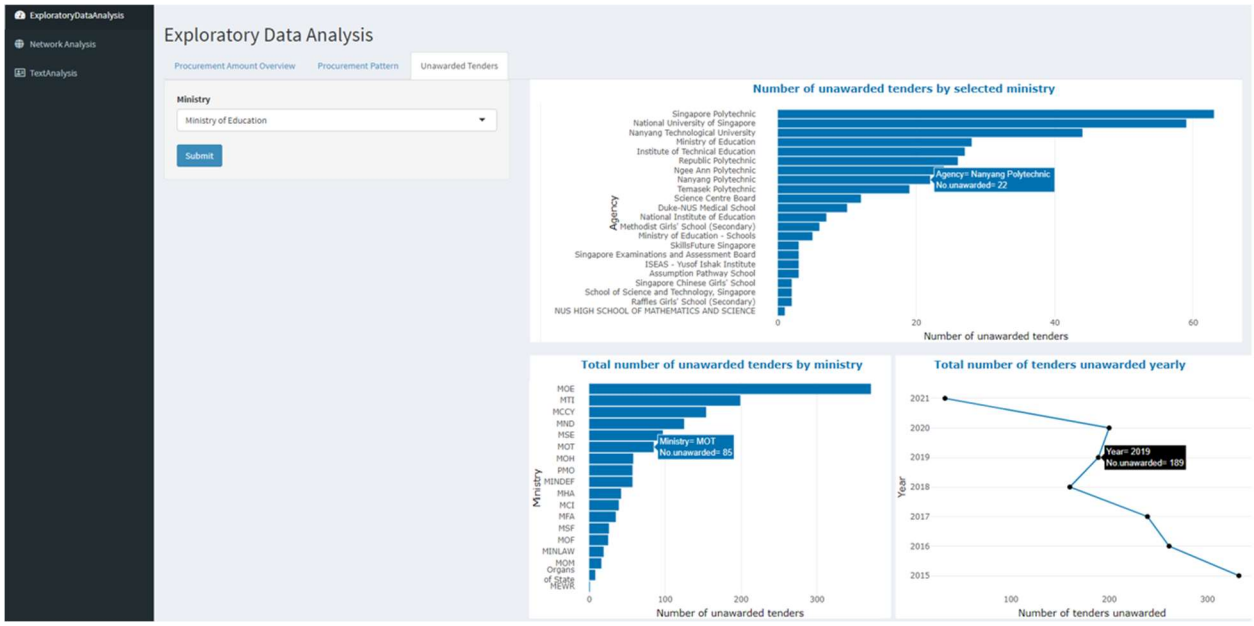


<u>Selection</u>

Users can choose a ministry under the "Ministry" dropdown list. To select a desired "Number of top suppliers", users can slide the number bar ranging from 1 to 50. Then click on "Submit" to view the visualizations.

When a ministry is chosen, the line graph will display the yearly award amount of the selected ministry during 2015 to 2021. The bar chart on the right will show the top suppliers by descending awarded amount under the selected ministry and specified number of top suppliers. Tool tips are available for easy references.

2.2.3 Unawarded Tenders



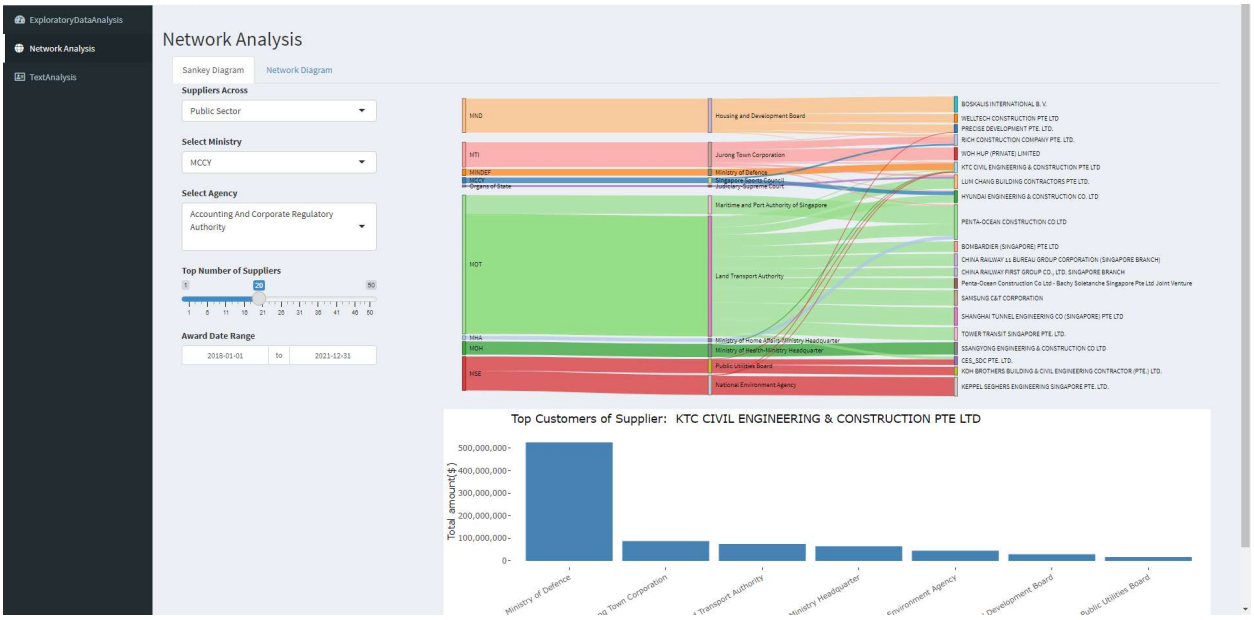
Under the “Unawarded Tenders” tab, users are able to select a ministry, and have an interactive view of the number of unawarded tenders by the agencies belong to the ministry.

On the bottom of the panel displays 2 static charts: the total number of unawarded tenders by ministry, and the total number of tenders unawarded yearly, both are aggregated numbers for the given dataset. Tooltips are available for each chart to see the detailed values.

2.2 Network Analysis

2.2.1 Sankey Diagram

On clicking on the “Sankey Diagram” tab item, the Sankey Diagram user selections and diagram would appear on the main panel.



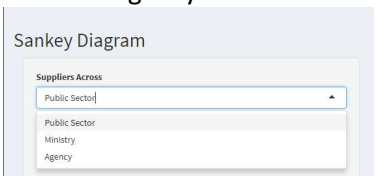
Level Selection

Under the “Suppliers Across” selection input, users can choose one of 3 levels of views:

“Public Sector” to view visualisation for agencies across the entire public sector

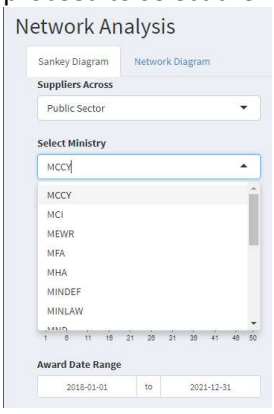
“Ministry” to view view visualisation for agencies across a particular ministry

“Agency” to view view visualisation for a particular agency



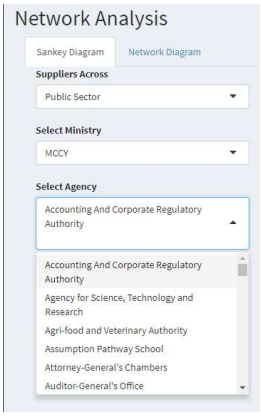
Ministry Selection

If the user has selected “Ministry” under the “Suppliers Across” selection input, the user can then proceed to select the ministry under the “Select Ministry” selection input.



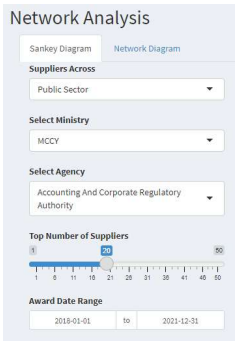
Agency Selection

If the user has selected “Agency” under the “Suppliers Across” selection input, the user can then proceed to select the agency under the “Select Agency” selection input.



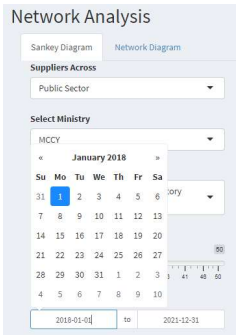
Number of Top Suppliers Selection

Using the slider bar, users can filter the visualisation to display the data of top number of suppliers of the selected ministry or agency.



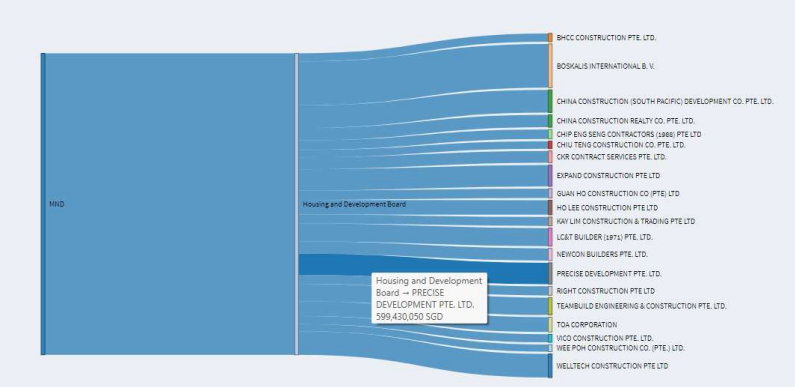
Award Date Range Selection

Using the Award Date Range fields, user can filter visualisation based on the date range for the tender award date.



Sankey Diagram

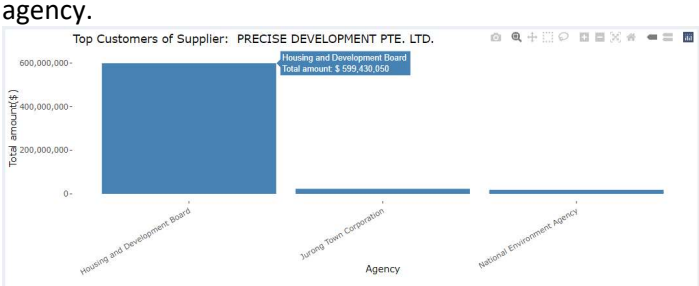
The mouse-over on a link in the Sankey diagram will display a tooltip showing the source and target, and the corresponding procurement value.



Clicking on the supplier node on the right end of the Sankey diagram will generate a bar char of top customers of the selected supplier (see section on Top Customers Bar Chart).

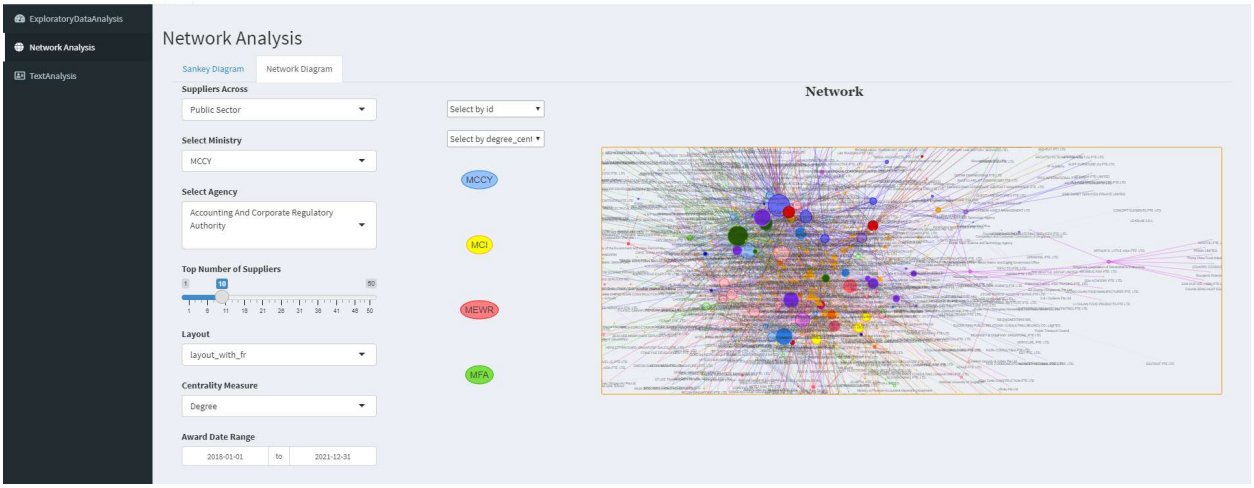
<u>Top Customers Bar Chart</u>

Mousing over a bar would display the agency name and the total amount awarded to the supplier by the agency.



2.2.2 Network Diagram

On clicking on the “Network Diagram” tab, the Network Diagram user selections and diagram would appear on the main panel.



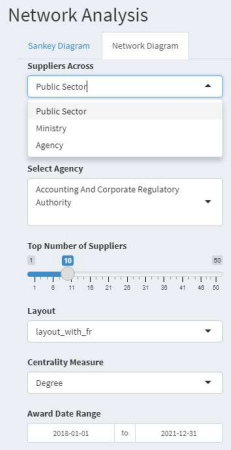
Level Selection

Under the “Suppliers Across” selection input, users can choose one of 3 levels of views:

“Public Sector” to view visualisation for agencies across the entire public sector

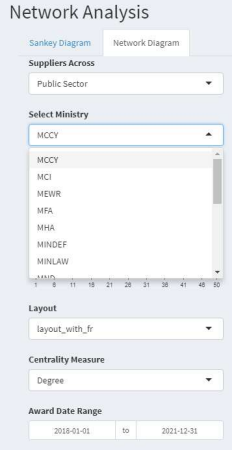
“Ministry” to view view visualisation for agencies across a particular ministry

“Agency” to view view visualisation for a particular agency



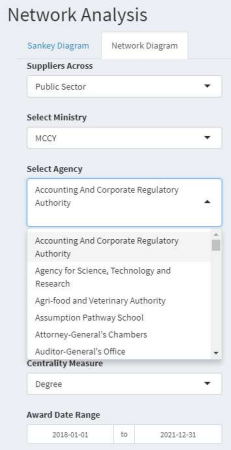
Ministry Selection

If the user has selected “Ministry” under the “Suppliers Across” selection input, the user can then proceed to select the ministry under the “Select Ministry” selection input.



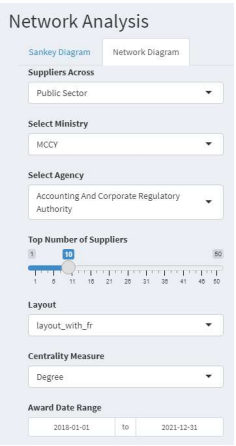
Agency Selection

If the user has selected “Agency” under the “Suppliers Across” selection input, the user can then proceed to select the agency under the “Agency” selection input.



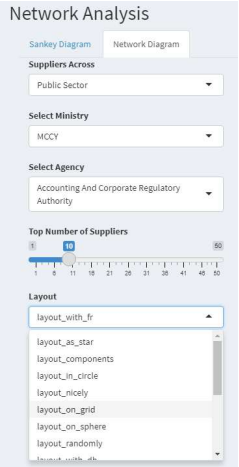
Number of Top Suppliers Selection

Using the slider bar, users can filter the visualisation to display the data of top number of suppliers the selected agency (if "Agency" is selected for "Supplier Across" input) or each agency across the public sector (if "Public Sector" is selected for "Supplier Across" input) or across the ministry (if "Ministry" is selected for "Supplier Across" input).



Layout Selection

In the "Layout" selection, users can select the type of layout for the network diagram.

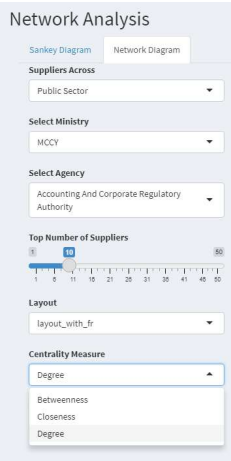


The following layouts are available:

- layout_as_star
- layout_components
- layout_in_circle
- layout_nicely
- layout_on_grid
- layout_on_sphere
- layout_randomly
- layout_with_dh
- layout_with_drl
- layout_with_fr
- layout_with_gem
- layout_with_graphopt
- layout_with_kk
- layout_with_lgl
- layout_with_mds

Centrality Measure Selection

Under the “Centrality Measure” selection input, users can select the type of centrality measure to display in the network diagram.



The following centrality measures are available:

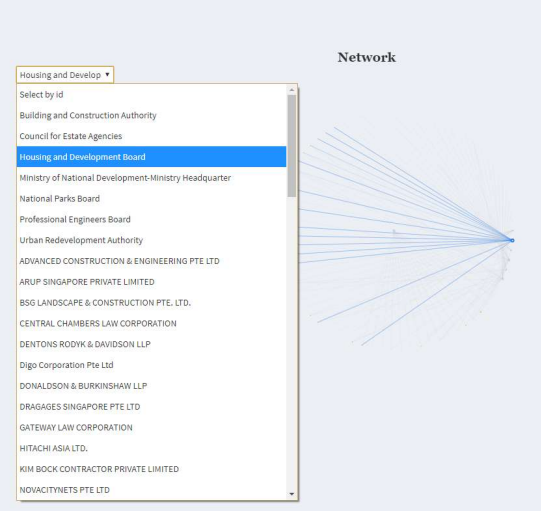
- Betweenness
- Closeness
- Degree

Award Date Range Selection

Using the Award Date Range fields, user can filter visualization based on the date range for the tender award date.

Select by id

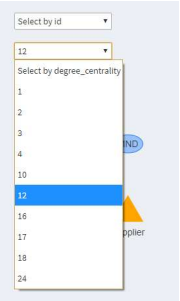
From the “Select by id” dropdown list, users can select the node and its corresponding connections to highlight on the diagram.



Select by centrality

From the “Select by centrality” dropdown list, user can select to highlight nodes with the same centrality measure.

The type of centrality measure will change based on the “Centrality Measure” selection input.

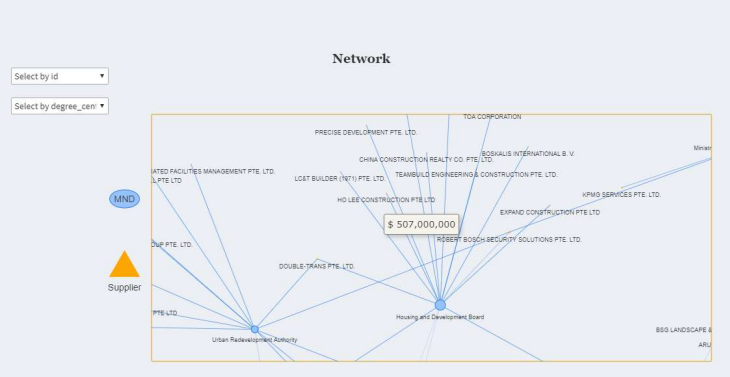


Network Diagram

Users can the mouse scroll to zoom in and out of the network diagram.

Agencies are denoted in blue circles. Suppliers are denoted in orange triangles.

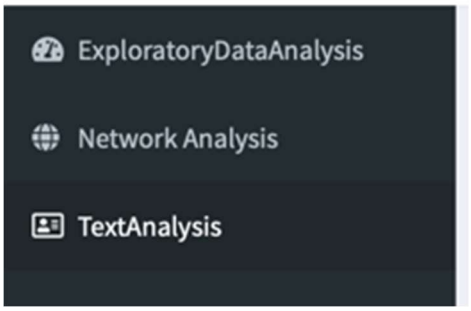
The size of each marker is based on degree centrality.
The thickness of each edge/line between the different markers is proportionate to the amount awarded.
Mousing over the edge/link will display the amount awarded.



2.3 Text Analysis

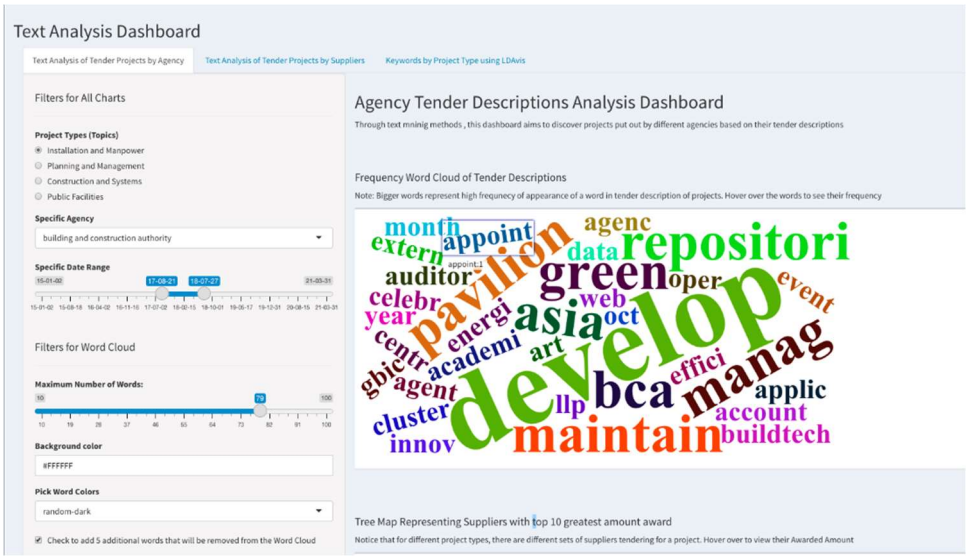
Navigating to the Text Analysis Dashboard

From the side navigation layout, click on “Text Analysis”. This will bring up the text analysis dashboards and their respective in page tabs.



2.3.1 Dashboard 1: Text Analysis of Tender Projects by Agency

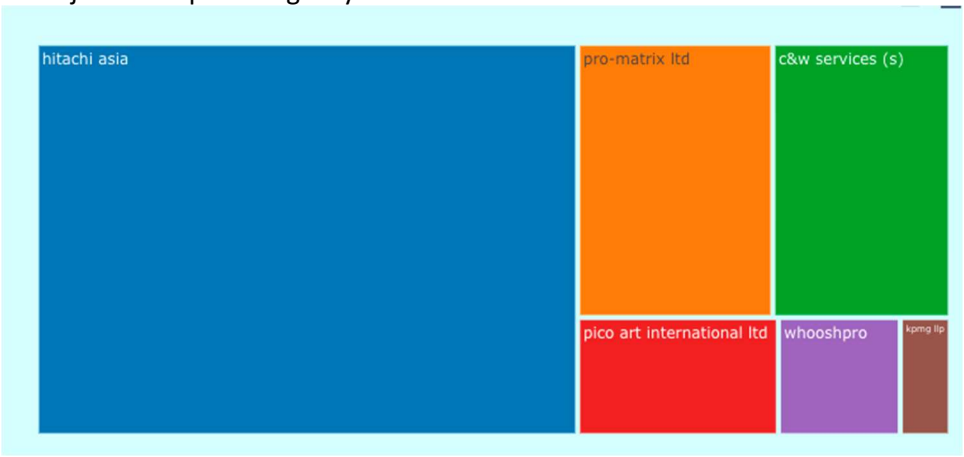
Notice that bringing up the dashboard page, there would be three different tabs. For the first analysis page, we will be navigating over to the in-page tab,” Text Analysis of Tender Projects by Agency”.



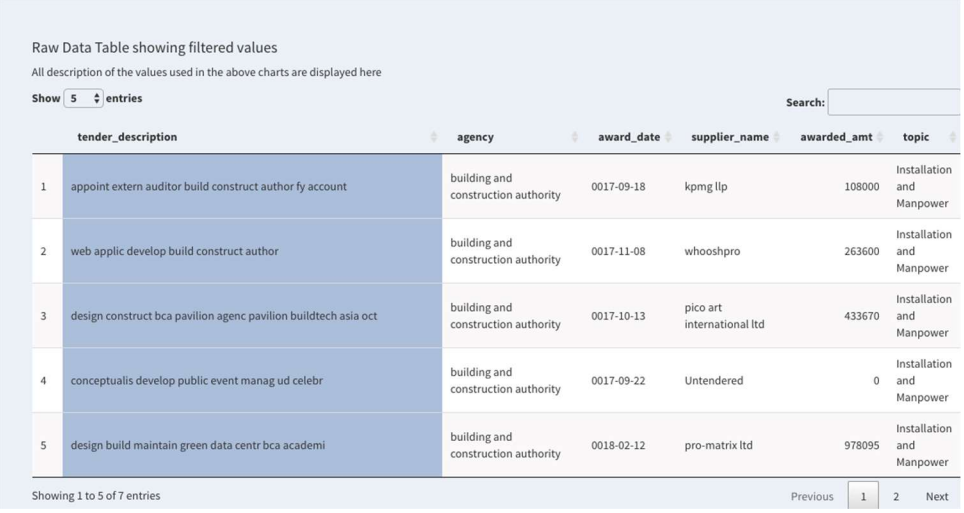
There are a total of three charts that will appear.

1. A Word cloud: Showing the tender description word frequency where bigger words represent higher frequency of a specific word. Hover over the word cloud to view the frequency of words.

2. Tree Map: Showing top 10 Suppliers the agency has awarded where the leave size is based on awarded amount. Moreover, over to view the amount awarded to the different suppliers tendering a job for a specific agency.



3. Table View: Showing the raw data used to create the above charts



Raw Data Table showing filtered values

All description of the values used in the above charts are displayed here

Show 5 entries

Search:

	tender_description	agency	award_date	supplier_name	awarded_amt	topic
1	appoint extern auditor build construct author fy account	building and construction authority	0017-09-18	kpmg llp	108000	Installation and Manpower
2	web applic develop build construct author	building and construction authority	0017-11-08	whooshpro	263600	Installation and Manpower
3	design construct bca pavilion agenc pavilion buildtech asia oct	building and construction authority	0017-10-13	pico art international ltd	433670	Installation and Manpower
4	conceptualis develop public event manag ud celebr	building and construction authority	0017-09-22	Untendered	0	Installation and Manpower
5	design build maintain green data centr bca academi	building and construction authority	0018-02-12	pro-matrix ltd	978095	Installation and Manpower

Showing 1 to 5 of 7 entries

Previous 1 2 Next

On the side notice that there will be a set of filter options for users to filter for specific values that will manipulate the visualizations on the right. Notice that there are two kinds of filters. “Filter for All Charts” applies to all three charts and “Filters for Word Cloud” apply only to the word cloud.

Firstly, we have the filter that will apply to all charts. These filters will allow the user to filter the charts for a:

1. Specific Project Type (Topic)
2. Specific Agency
3. Specific Data Range available in the Data

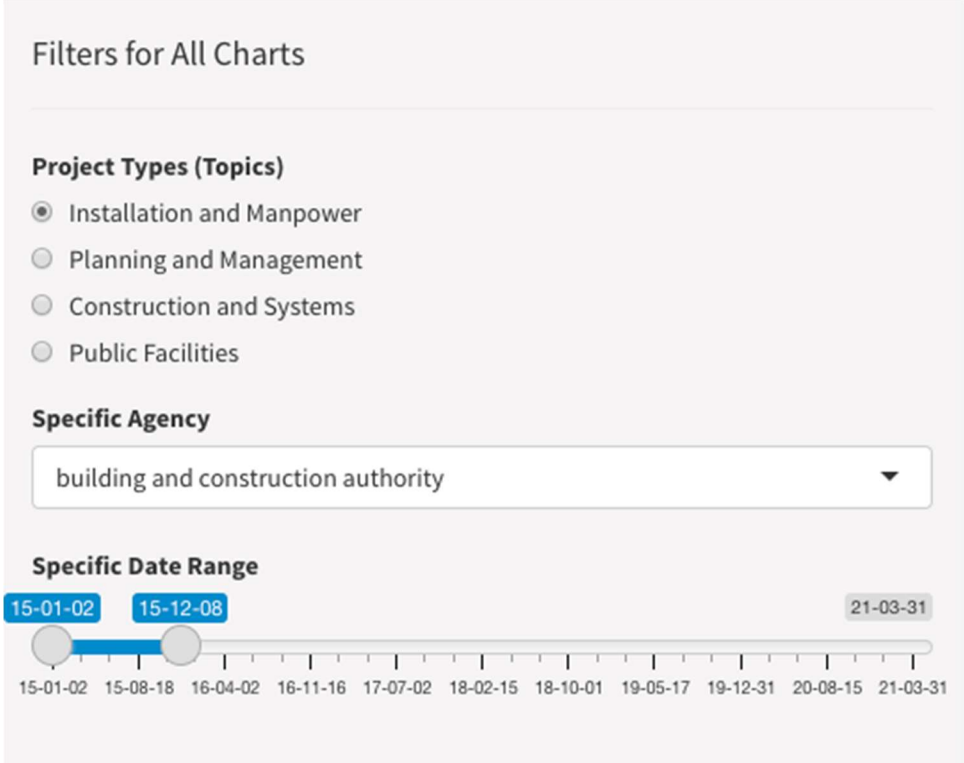
The purpose of this is to allow the user to analyze the kinds of tender jobs across different agencies, project types and date range. The analysis is to allow users to drill down to agency tenders needs across project types, agency, and date range.

Secondly, the main point on this section of the dashboard is to perform text analysis. As such, the textual information is rich in tender description. This gives rise to noticeable high frequency words that give context to the agency's needs. For example, notice that in the above dashboard, it seems like that for across all agencies the projects pertain to managing, building, design and maintenance of public organizations and roads. Additionally, having controls to change the number of words displayed on the chart would be a good addition.

Let us dive into its details.

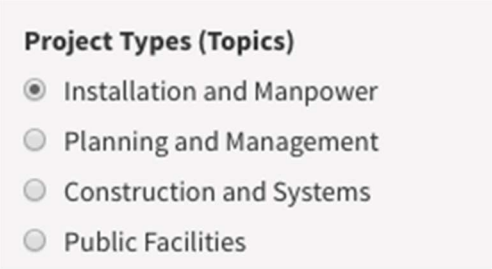
For filters pertaining to all charts:

1. Project Types (Topics)



The tender jobs cater for various kinds of projects in which a trained Latent Dirichlet Allocation Model was used to identify these main topics from tender descriptions. Notice that from the above, there are four topics found (with a good coherent score, explored in the "Keywords by Project Type using LDAvis which will be explained later"). The above will allow the user to filter for the four specific topics. All data pertaining to a specific project type would be shown.

2. Filtering for Specific Agencies

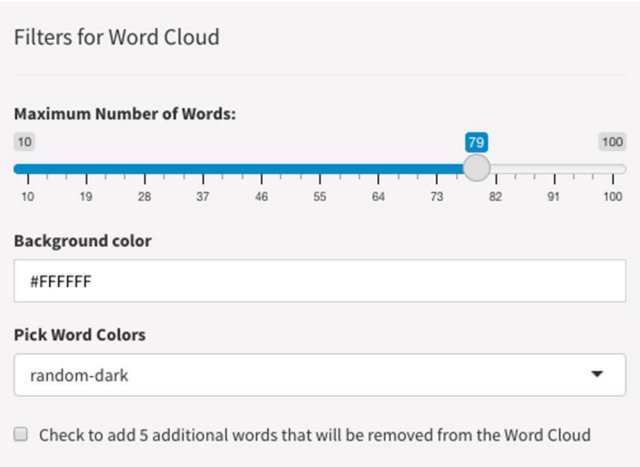


If the user specific wants to drill down to a specific agency, the user can do so, and all records pertaining to this agency would be shown.

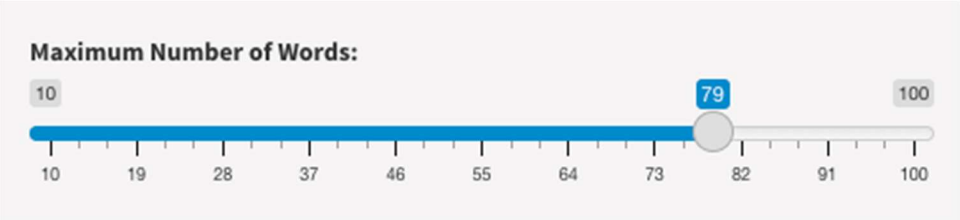
3. Specific Date Range Filter (YY-MM-DD)

The Specific Date Range filter will allow the users to specific a range of records pertaining to a date change. The use of a slider would allow the user to have specific and flexible controls over the data range.

For the next section we will explain the use of the filters for the word cloud.



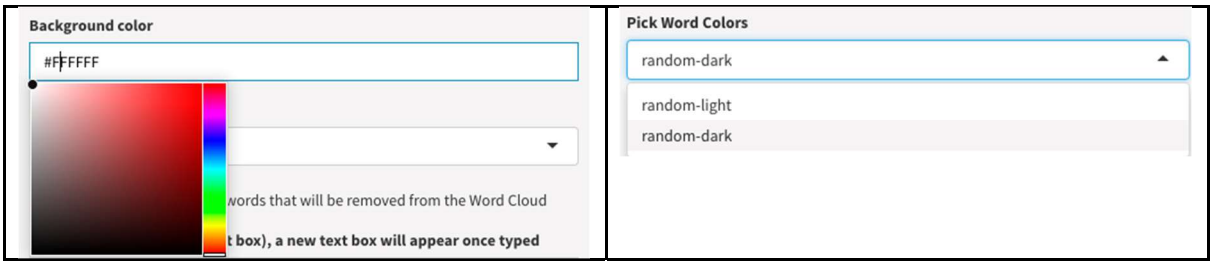
1. Filter for minimum number of words to display.



Based on the minimum number of words to display, the word cloud will pick based on a descended order list of words based on the word frequency. Users will be able to toggle to number of words to display. Often, many records can amount to a large amount of textual noise, having such an option is very much welcomed to filter out for data noise and only observe words of high frequency.

2. Background Color and Word Cloud

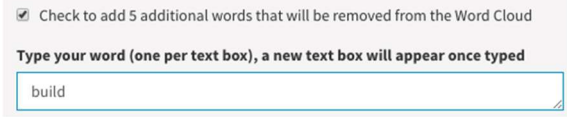
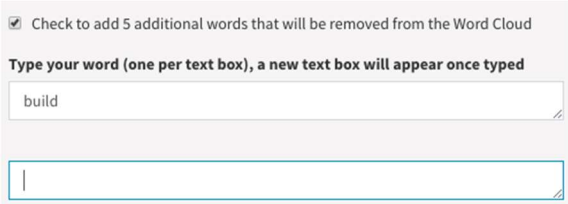
Background Colour Picker	Word Colour Picker



In addition to graph customization, users can filter for specific background color of the word cloud. This will allow for words to have a better contrast for different sets of audience. A color picker is used to select color based on the Hexa-color code which will allow users to select more specific colors as compared to say RGB only. In addition to background color, words can also be customized to assorted colors. It can be resource intensive to map every word to a specific color; especially for large databases. Due to the capacity of server resources, only two options for word colors are available for now. Users can select between random light or random dark as they contrast with different background colors.

3. Removal of Additional Words

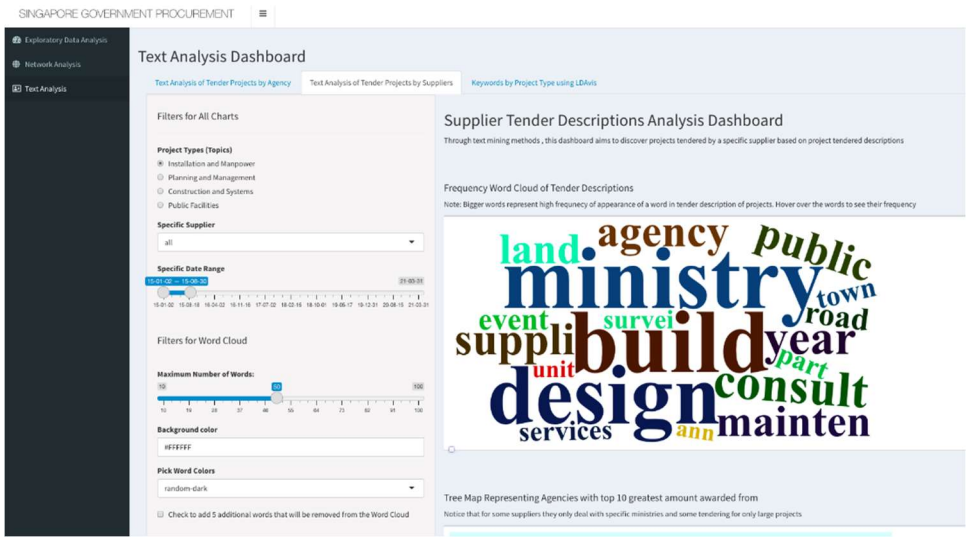
If users wish to remove more words from the word cloud, users can check on the check box as seen above. This will activate a conditional text box where users can input additional words in the text box to be removed from the word cloud.

Typing additional Words for Removal	After Typing additional Work for Removal
	

Once the words are typed in as such, the word will be removed from the word cloud as seen on the left. As more words are typed into the text box, a new text box will appear below for users to input more additional words to remove from the word cloud. Note that users can only remove up to 5 additional words in total and that this text box is CAPS sensitive. Below is the result of the word cloud after removing the word “build”.

Before Word Removal	After Word Removal
	

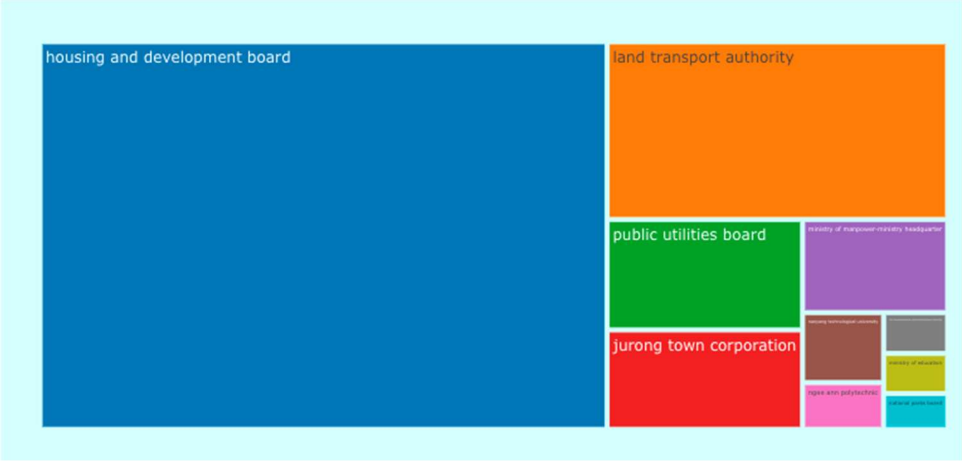
2.3.2 Dashboard 2:"Text Analysis of Tender Projects by Suppliers"



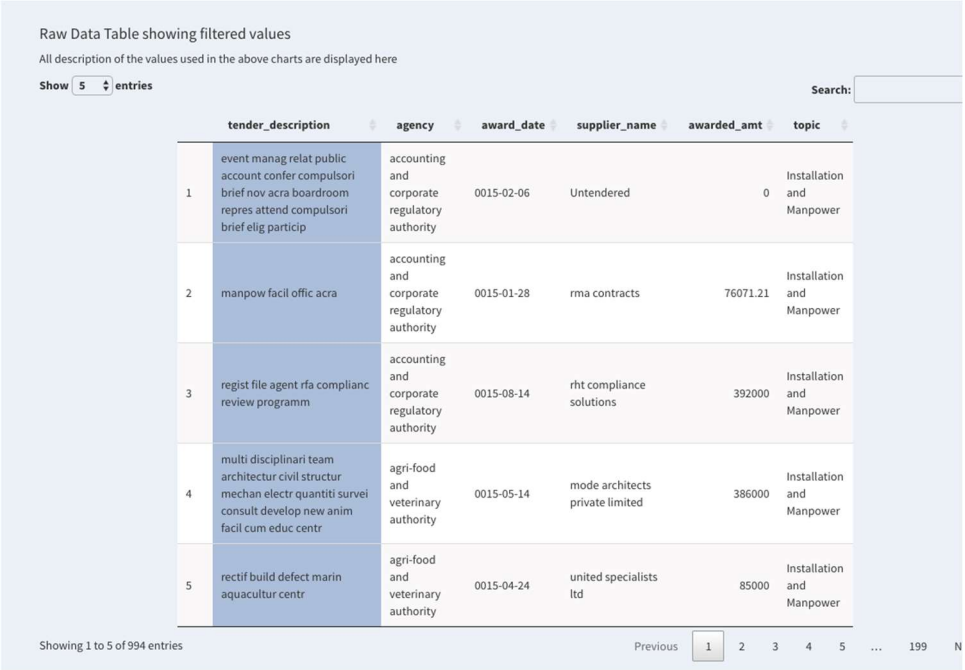
For the second analysis page, we will be navigating over to the in-page tab," Text Analysis of Tender Projects by Suppliers". The set of filters for this page is the same. The only difference is the kind of analysis and the tree map itself. Similarly, the charts provided are:

1. Word Cloud representing the Tender Description for the projects tendered by the different suppliers. This word cloud provides insights into the kind of project a specific supplier has tendered for based on the frequency of words that appear across all the projects tender description.

2. Tree Map: Showing top 10 Agencies the Supplier was awarded from where the leave size is based on awarded amount. Hover over to view the amount awarded to the by the different agencies.



3. Table View: Showing the raw data used to create the above charts



Raw Data Table showing filtered values

All description of the values used in the above charts are displayed here

Show 5 entries

Search:

	tender_description	agency	award_date	supplier_name	awarded_amt	topic
1	event manag relat public account confer compulsori brief nov acra boardroom repres attend compulsori brief elig particip	accounting and corporate regulatory authority	0015-02-06	Untendered	0	Installation and Manpower
2	manpow facil offic acra	accounting and corporate regulatory authority	0015-01-28	rma contracts	76071.21	Installation and Manpower
3	regist file agent rfa complianc review programm	accounting and corporate regulatory authority	0015-08-14	rht compliance solutions	392000	Installation and Manpower
4	multi disciplinari team architectur civil structur mechan electr quantiti survei consult develop new anim facil cum educ centr	agri-food and veterinary authority	0015-05-14	mode architects private limited	386000	Installation and Manpower
5	rectif build defect marin aquacultur centr	agri-food and veterinary authority	0015-04-24	united specialists ltd	85000	Installation and Manpower

Showing 1 to 5 of 994 entries

Previous 1 2 3 4 5 ... 199 N

On the side, notice that a similar set of filters are provided. The functionality and purpose of these filers are similar. “Filter for All Charts” applies to all three charts and “Filters for Word Cloud” apply only to the word cloud.

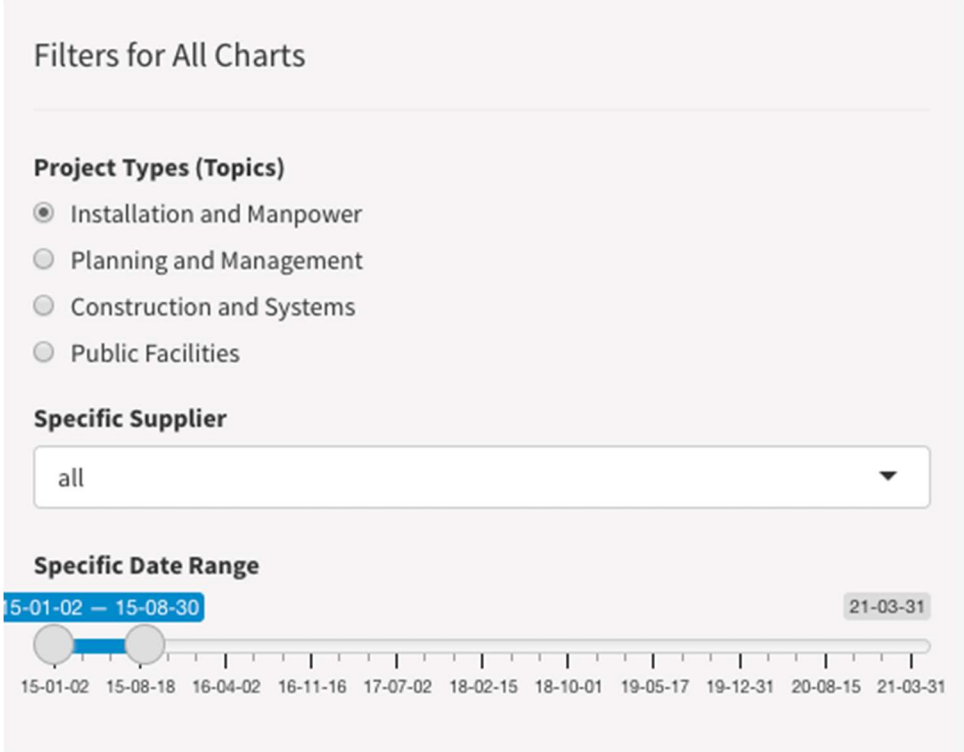
Firstly, we have the filter that will apply to all charts. These filters will allow the user to filter the charts for a:

1. Specific Project Type (Topic)
2. Specific Supplier
3. Specific Data Range available in the Data

The purpose of this is to allow the user to analyse the kinds of tender jobs across different suppliers, project types and date range. The analysis is to allow users to drill down to agency tenders needs across project types, agency, and date range. Similarly, this dashboard main point, is to allow users to perform some level of text analysis on tender description for projects done by the different suppliers. The main aim of this part of the dashboard is to highlight specialisations for specific suppliers. It also aims to provide a complimentary view of who the supplier has worked with, and which are the highly valuable projects tendered before by the different agencies as seen in the tree map. Lastly, the raw table view provides the raw data for users to perform a self-drill down on the data itself.

Let's dive into its details.

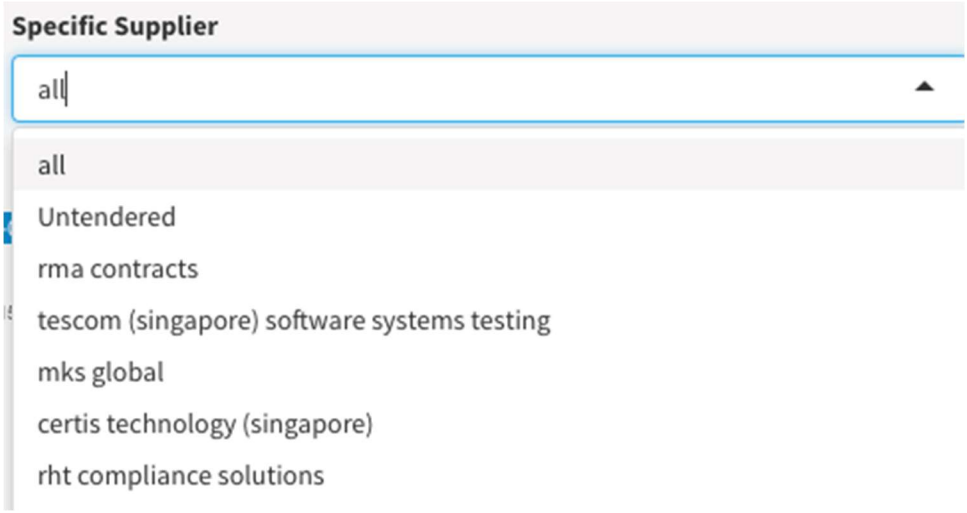
For filters pertaining to all charts:



1. Project Types (Topics)

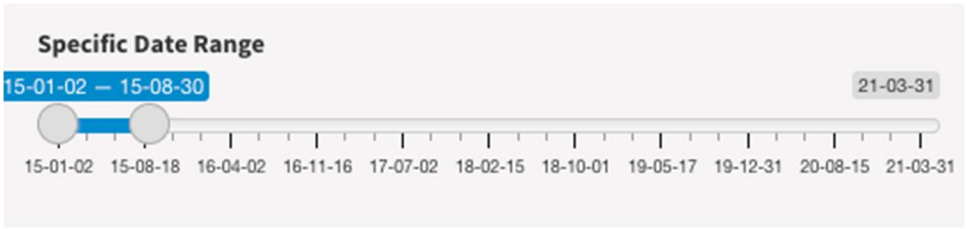
The tender jobs cater for various kinds of projects in which a trained Latent Dirichlet Allocation Model was used to identify these main topics from tender descriptions. Notice that from the above, these are the four related topics used. The above will allow the user to filter for the four specific topics. All data pertaining to a specific project type would be shown.

2. Filtering for Specific Supplier



If the user specific wants to drill down to a specific supplier, the user can do so, and all records pertaining to this supplier would be shown. This is the only significant change to the filters for the supplier dashboard for text analysis.

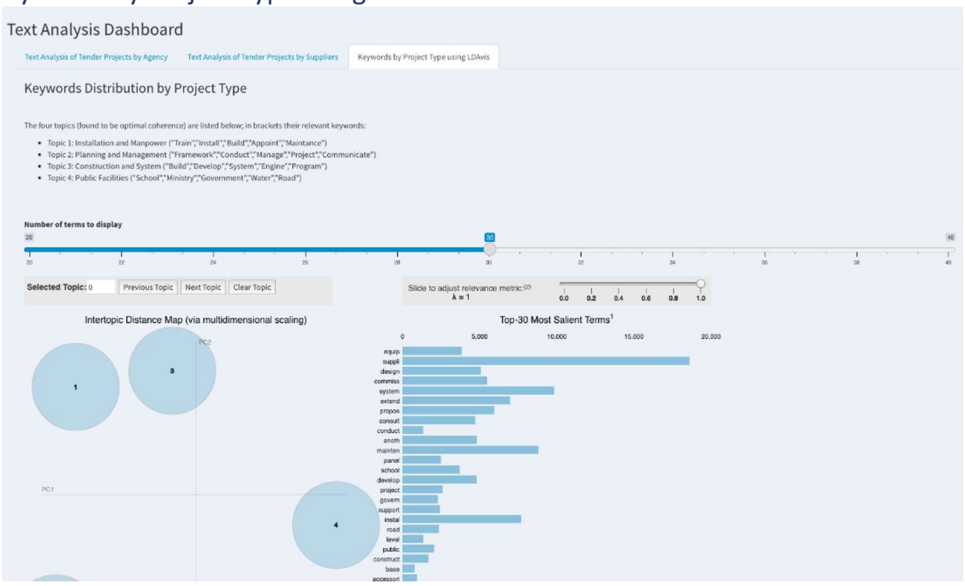
3. Specific Date Range Filter



The Specific Date Range filter will allow for the users to specify a range of records pertaining to a date change. The use of a slider would allow the user to have specific and flexible controls over the data range.

For the next section we will explain the use of the filters for the word cloud. Note that the usage of the word cloud is the same for that of “Text Analysis of Tender Projects by Agency”.

2.3.3 Keywords by Project Type using LDAvis



Lastly, “Keywords by Project type using LDAvis” is a dashboard to provide a visualisation on the topic model used to pick up the salient features of tender description that pertain to the four project types. In this dashboard, there are two parts.

1. Project Type Extraction

Based on the tender description, a topic model; specifically, the Latent Dirichlet Allocation was used to model the topics. From there, the salient features picked up from the tender description help model the topics. The top description shows the four topics found and the relevant words that amounted to its eventual category. For example,

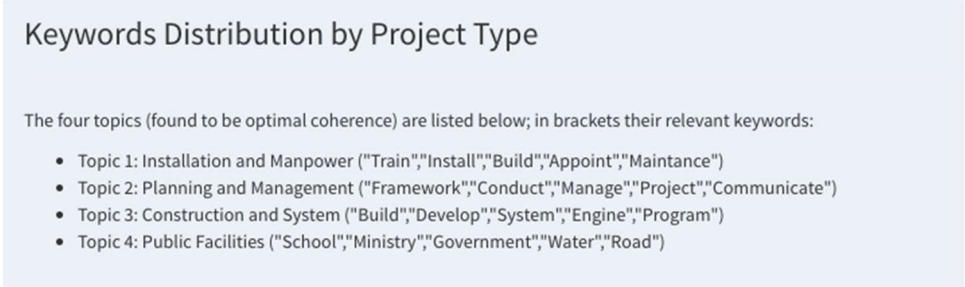
Topic 1: Pertains to the installation and manpower projects with salient terms like "Train", "Install", "Build" ,"Appoint", "Maintained" appearing.

Topic 2: Pertains to Planning and Management projects with salient terms like "Framework", "Conduct", "Manage", "Project", "Communicate" appearing.

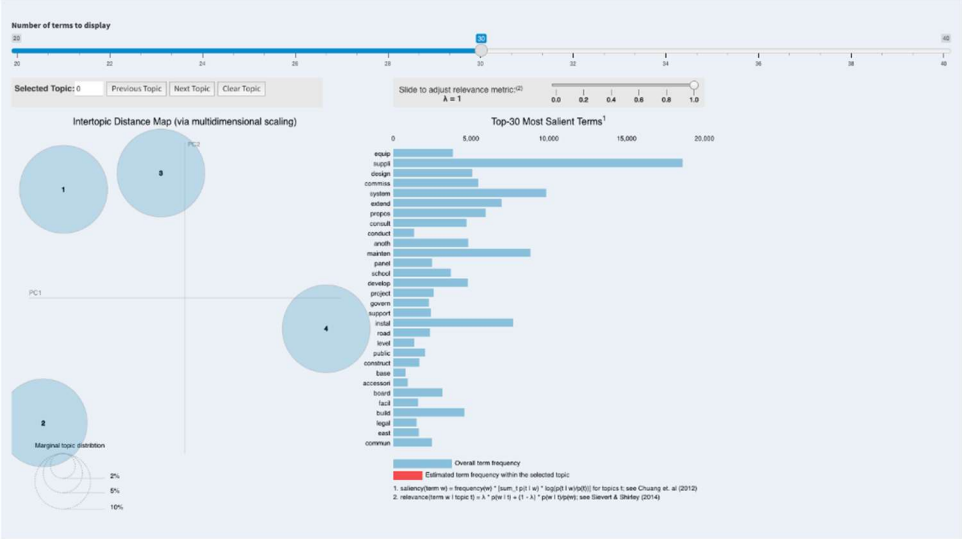
Topic 3: Pertains to the Construction and System projects with salient terms like "Build", "Develop", "System", "Engine", "Program" appearing.

Topic 4: Pertains to the public facilities projects with salient terms like "School", "Ministry", "Government", "Water", "Road" appearing.

This is as seen below:



1. Topic modelling visualisation: LDAvis

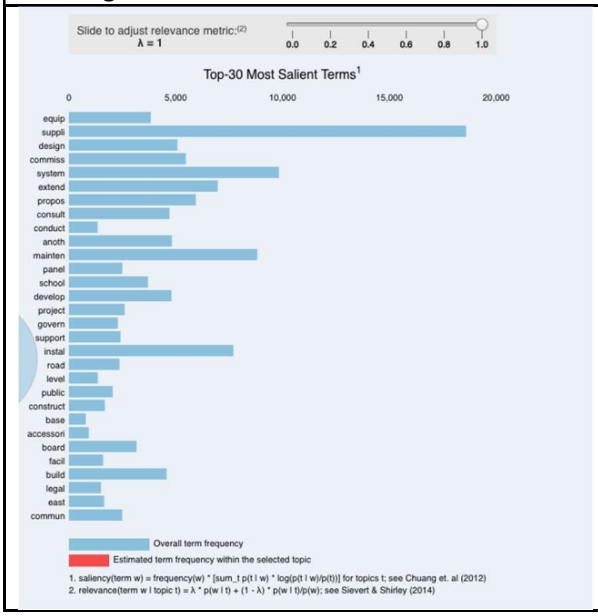
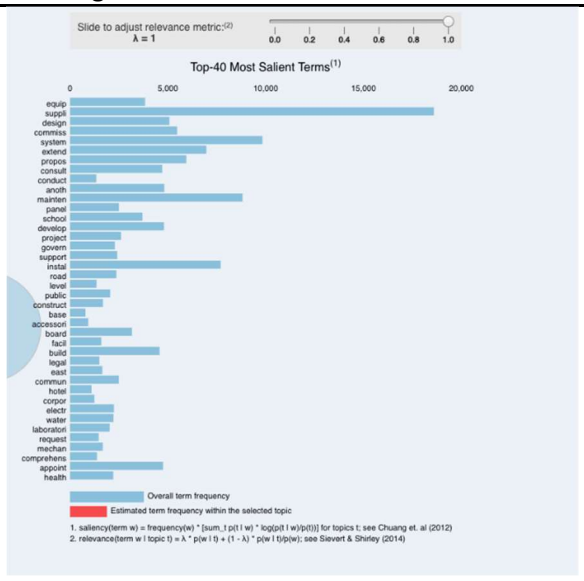


LDAvis is a visualisation built specifically for topic models. The idea is that users able to view the topic word clusters and drill down on the frequency of the words appearing in each cluster. The important part of topic modelling is the ability of the topic model to be able to identify salient terms that show cohesiveness.

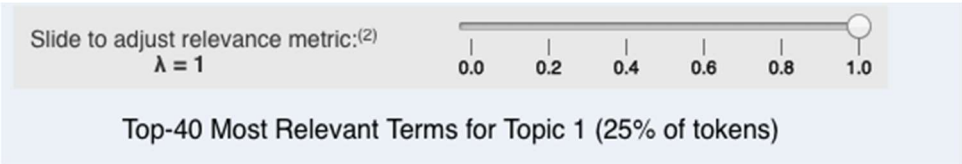
2. Filtering for more relevant salient terms ("Number of terms to display" filter)

On the top of the graph, there is a slider to allow users to slider for different number of terms to display in the LDAvis chart. Sliding the slider will either increase or decrease the number of words appearing on the right slide of the chart. Note that the default, amount is 30 and users can adjust to a minimum of 20 or a maximum of 40 relevant salient terms.

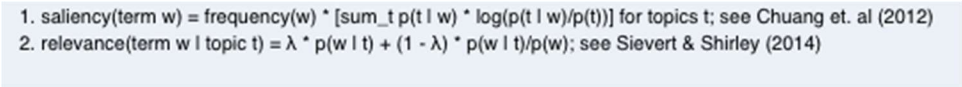
For example:

Filtering for 30 Words	Filtering for 40 Words
	

Notice that the last word is different. That is because, the total number of words to show has been increased. Additionally, just below, a lambda filter is used to adjust the relevance threshold of words for more technical users who wish to adjust the topic model parameters.

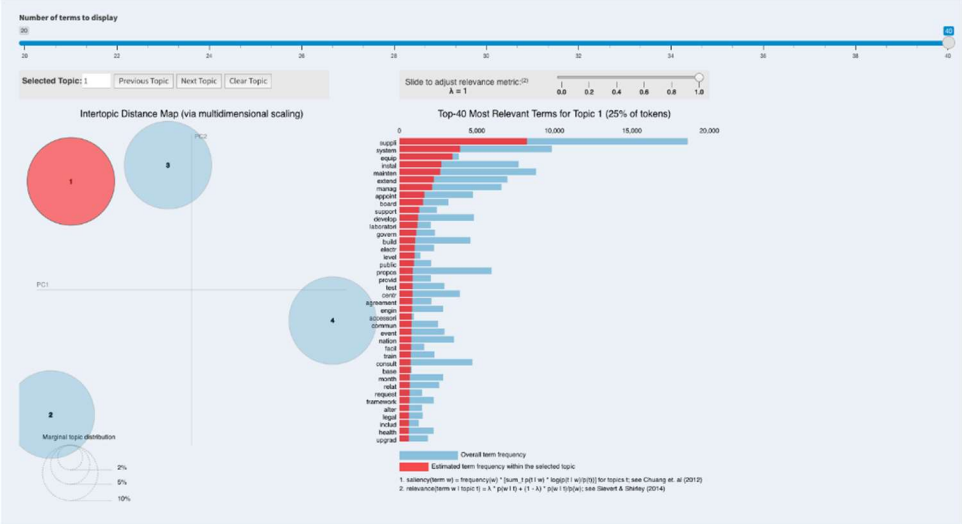


The description of how these were calculated are stated at the bottom as seen in the picture below :



3. Filtering words by Topic-word clusters.

On the left, users can select the topic word cluster which has an integrated filter to filter for the different topic words that will be highlighted in red by the graph on the left. This is depicted as seen below:



The highlighted words represent the frequency of each salient term that appeared in the topic-word cluster. This will allow users to draw understanding of how project types are named or categorized. As topic modelling is an unsupervised method, it would be important that such information is made transparent to for open discussion on what the project types should be. This will allow room for future improvement and suggestion by team members on how to name these different project type categories cohesively.