GeVIZ: Who Supplies the Needs of Singapore? Introducing A Way to Streamline Operations and Improve Efficiencies

Joel CHOO Peng Yeow, Manu George MATHEW, Yusuke MINAMI

Abstract— Launched in 2000, the Government Electronic Business (GeBIZ) Portal is an integrated portal for use by all Singapore government agencies to conduct business electronically with their suppliers. There are more than S$10 billion worth of business opportunities published annually to 30,000 suppliers registered with GeBIZ. Over the years, GeBIZ has accumulated a knowledge base of procurement data and massive amounts of data is a valuable source of market knowledge. A Business Intelligence (BI) platform was introduced in 2007 to help mine this information to help improve the efficiency and effectiveness of government procurement.

However, like any tool it does have its advantages and disadvantages. As such, we will be exploring ways to help improve current tools and increase the ability to garner insights. Predominantly, we will be exploring network graphs (vizNetwork & tidygraph) using a R shiny application to help visualise the relationships between the Ministries, Agencies and Suppliers. This will be an additional feature which many off-the shelf commercial BI tools are still lacking which could serve beneficial.

Index Terms—GeBIZ, Procurement, Business Intelligence, Network Graphs, visNetwork, tidygraph, R Shiny.

MOTIVATION OF THE APPLICATION

Due to the vast number of quotations & tenders each year, it becomes extremely challenging to track transactional patterns and entities who are involved in each of these contracts. Some of the issues that might have arisen includes:

- Tedious for potential supplier to research past tenders, quotations and period contracts of similar purchases across the entire public sector to determine quotation prices
- Lack of Ministry oversight on how the budgets were spent in the individual sectors and service categories
- Inability to identify reliable suppliers that many agencies and their respective ministries are purchasing from
- Recommend appropriate procurement categories and suggest possible suppliers to invite during the tender notification process

With the provision of GeBIZ procurement data, current analysis is limited to Agencies and Supplies and we will not be able to view the interactions across ministries. Furthermore, information on the type of contracts were also embedded in long text descriptions which makes it difficult to analyse how have budgets been spent.

In view of current constraints, we are motivated to create a dynamic and interactive dashboard to help provide ministries, agencies and suppliers a holistic view on the procurement contracts made thus far.

REVIEW AND CRITIC ON PAST WORKS

The development of Business Intelligence in GeBIZ started in early 2006. GeBIZ BI initiatives can be broadly divided into two areas. The first area entails the development of GeBIZ InSIGHT. It leverages on Machine Learning (ML) techniques to help individual procurement users research historical buys and gain market insights.

The second area covers the development of GeBIZ Management Console (GMC). GMC enables macro-level portfolio management and performance management in the public sector by its features such as filtering, pivot tables, and charts. As mentioned, the existing tools could help gain market insights but there are still ways to help improve how we analyse information with the use of visual analytical techniques. [2]

DESIGN FRAMEWORK

The overview of our approach can be seen in the following figure.

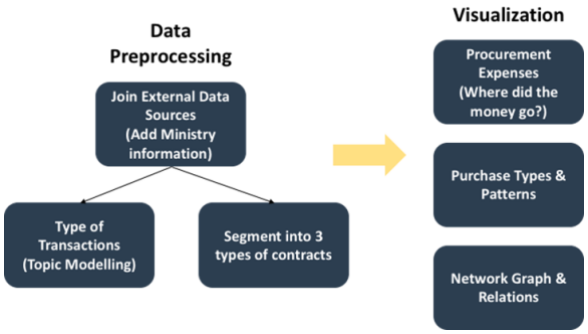


Fig. 1: Overview of the approach

Prior data pre-processing will be carried out to add in the ministry that these agencies belong to. In addition, we will segment the contracts based on their dollar value and topic modelling will also be implemented to determine their respective procurement categories.

With a three–pronged approach, we will help explore how the procurement expenses look like across ministries, agencies and suppliers. We will also investigate if there are seasonal patterns and purchase types specific to an entity. Lastly, we will leverage on the use of network graphs to help us understand their relationships and extract their procurement information at ease with an interactive interface.

Network Visualization

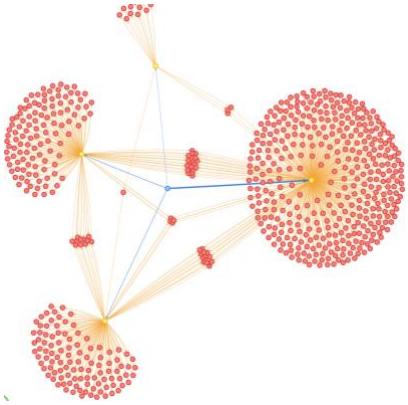


Fig. 2: Visualization by visNetwork

visNetwork was chosen as our preferred library to visualize network relations as it fosters interactivity which a user will be able to appreciate when dealing with complex nodes and edges. However, due to the lack of functions to implement centrality metrices like betweenness, we have made use of ***tidygraph*** to help us with the implementation and extraction of these additional columns.

In graph theory, betweenness centrality [9] is a measure of centrality in a graph based on shortest paths. A node with high betweenness would mean that it would have more control over the network, because more information will pass through it. With this metric in the procurement context, we would like to identify key suppliers which deals across multiple Agencies.

Definition: Betweenness centrality of a node v is given by

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where σ_{st} is the total number of shortest paths from node *s* to node *t*, and σ_{st} (v) is the number of those paths that pass through *v*.

The transition from ***tidygraph*** to ***visNetwork***, however requires us to manipulate our data before it will be compatible for the individual libraries. In tidygraph, we need to ensure that the columns are renamed to "source" and "target" while visNetwork uses "from" and "to". Using dplyr, we can easily perform dataframe manipulation.

As tidygraph also creates a tbl_graph which contains two tibble object where we can use a combination of *activate* and *as_tibble* functions to help extract the node and edge table.

```
A tbl_graph: 6093 nodes and 12451 edges

A directed acyclic simple graph with 1 component

Node Data: 6,093 x 4 (active)
   id label              group  betweenness_cent…
<int> <chr>              <chr>              <dbl>
    1 Ministry of Finan… Minis…                 0
    2 Organs of State    Minis…                 0
    3 Ministry of Natio… Minis…                 0
    4 Ministry of Trans… Minis…                 0
    5 Ministry of Trade… Minis…                 0
    6 Ministry of Cultu… Minis…                 0
... with 6,087 more rows

Edge Data: 12,451 x 5
 from    to value source target
<int> <int> <int>  <int>  <int>
   18   131     1     18    131
   18   132     1     18    132
   18   133     1     18    133
... with 1.245e+04 more rows
```

Fig. 3: tbl_graph

Once we have the relevant information, we can create a "Group" column and visNetwork will be able to display these groups as distinct colours. Setting the number of contracts as "value" also help visNetwork plot edges with difference thickness. The thicker the edge, the more contracts that are involved between 2 nodes.

Treemap

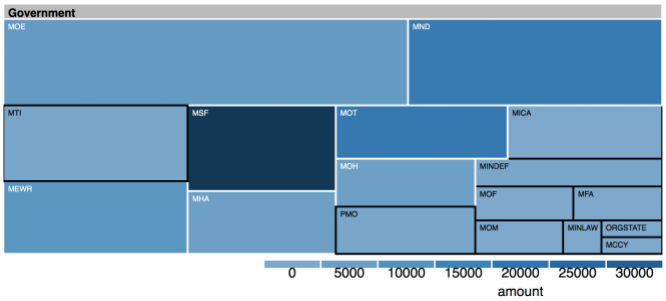


Fig. 4: Treemap

Treemap is a powerful way to visualize hierarchical data using nested rectangles. The advantage is that all data can be visualized in a single page utilizing rectangle area for the primary measure, and color for the second measure. Treemaps can be generated by a d3treeR library (https://github.com/d3treeR/d3treeR) in R.

While looking at Procurements, visualizing expenses at the Ministry and Agency level help us identify key spenders. This is useful when we want to know which ministry and which agency in that ministry tend to spend more. We use a Treemap Diagram to visualize the expenses at the ministry Level. The size of the tiles in the Treemap represent the number of orders placed by the ministry and the color indicates the amount spent. In the Sankey, we show the spending at ministry level for its top N agencies and the top N suppliers within each individual agency.

Sankey Diagram

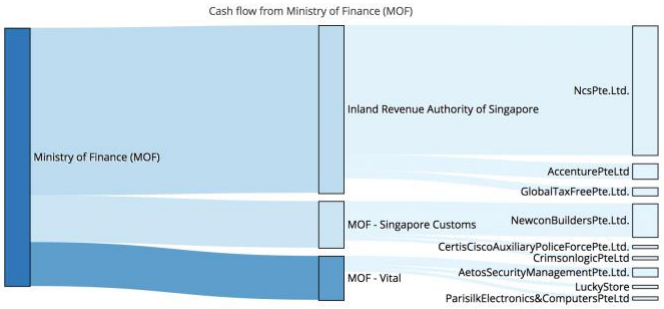


Fig. 5: Sankey Diagram

Sankey Diagram is another effective way to visualize a network graph as flow. The advantage of Sankey Diagram over Sunburst Diagram and regular network visualization is that nodes (and thus node labels) and edges are aligned in line, and the primary measure (count of procurement or monetary value SG$ in our dataset) represented by edge width can be compared with each other easily. In addition, the second measure can be represented by the color. Sankey Diagram can be generated by Plotly library (https://plot.ly/r/sankey-diagram/) in R.

Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is an example of Topic Modeling technique that enabled us to extract procurements type information from text descriptions. LDA assumes each word in each text sample

was chosen from topics which consist of words which have different probabilities to be chosen. Given the number of topics, LDA determines 2 distributions; distribution of words for each topic, and distribution of topics for each text sample. LDA can be run by lda library (https://cran.r-project.org/web/packages/lda/lda.pdf) in R. LDA was used to extract information from the text description column in our dataset. The output of LDA can be visualized on the plane of principal component axes by LDAvis library in R. (https://nlp.stanford.edu/events/illvi2014/papers/sievert-illvi2014.pdf)

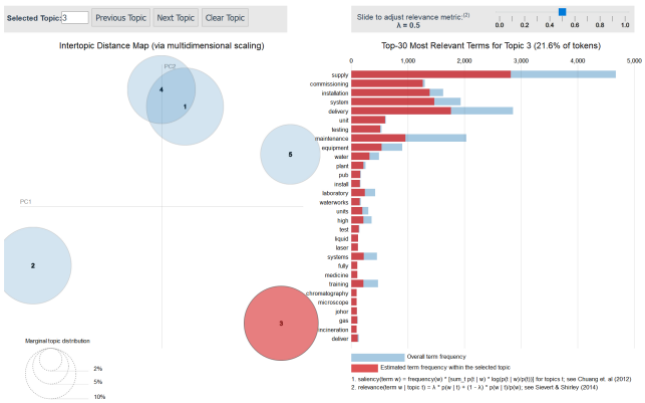


Fig. 6: Visualized result of LDA

LDA by setting the number of topics to 5 extracted 3 distinct topics (Topic 2, 3, and 5) on the plane of principal components as shown in the above screenshots. Based on the interactive visualized results, the 5 topics were labeled as follows.

Topic 1: Communication (relevant words: invitation, appointment, proposal, request])
Topic 2: Task (relevant words: works, engineering, consultancy, upgrading, building, road])
Topic 3: Product support (relevant words: system, installation, maintenance, testing)
Topic 4: General (relevant words: period, term, month, option)
Topic 5: Public/Organization (relevant words: school, boards, state, departments, organs)

Time Series

1. Calendar Plot

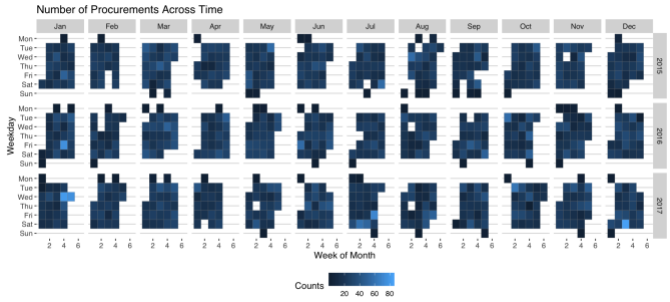


Fig. 7: Calendar plot

The essence to plot a calendar chart requires us to be able to manipulate data time to the necessary formats before we can obtain our desired graph. More importantly, string variables needs to be converted as factors so that weekdays and months could be interpreted as ordinal variables.

Using ***ggplot2*** with its ***geom_tile*** function, we can visualise the number of procurements across time in a calendar chart. This will help us identify if there are seasonality or cyclical patterns within an agency or supplier. For this example, one of the ministries tend to procure more during the end of the month. This will be useful for suppliers to help prepare for these peak bidding cycles.

2. Stacked Line Chart

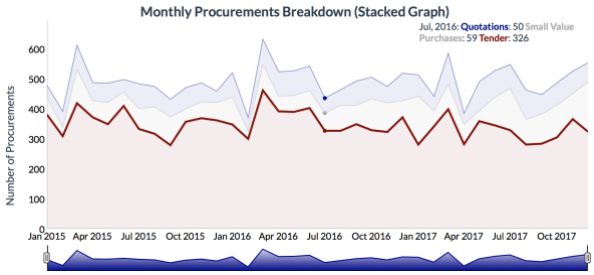


Fig. 8: Stack line chart

The stacked plot was created using the ***dygraph*** library which is a fast, reliable open source JavaScript charting library. Aggregating the number of contracts by month using ***dplyr***, we are able to identify trends & proportion of contracts contributed by the individual tenders segments. Upon inspection, a cyclical pattern becomes obvious with a spike in the number of contracts before April, which coincides with the end of the financial year.

DEMONSTRATION – A MOF CASE STUDY

Considering the limitations of current Business Intelligence (BI) tools for GeBiz in place [2], we will place more emphasis on the use of network graph to help identify the hierarchical relationships amongst entitles in this section.

Network Overview

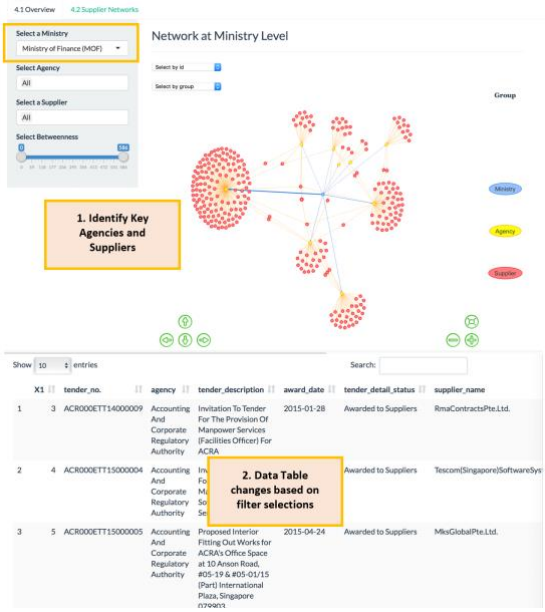


Fig. 9: Network overview

Using Ministry of Finance (MOF) as an example, we will be able to identify its respective Agencies and also their suppliers where they have individually engaged with thus far. This provides an oversight on who the **budgets** were spent on in the individual sectors.

As the chart is interactive, we will be able to drag around the nodes and zoom in to uncover who have made the most number of contracts. Here, MOF – Vital have sent out the greatest number of tenders amongst other agencies from MOF.

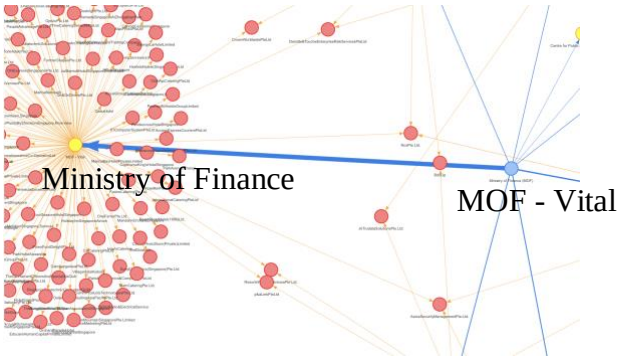


Fig. 10: Network involved with MOF

If we are interested to know the details of the tenders MOF – Vital have made, we can filter based on Agency and view the details in an interactive data table below. Sorting by awarded date, we will be able to find out the most recent contracts.

Fig. 11: Sorting and search

The inclusion of the search function also allows us to find out past quotations of similar procurement types. This will help suppliers research past tenders, quotations and period contracts of similar purchases across the entire public sector to determine quotation prices. This will not be made possible without the use of LDA where it has helped us generate a series of procurement types which a tender could belong to.

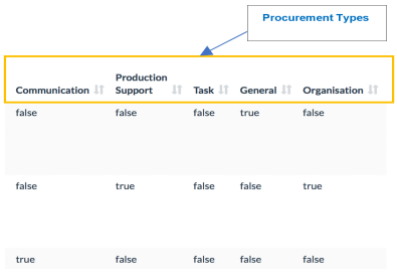


Fig. 12: Procurement type columns

This is an additional feature which have been added compared to the current BI systems where they use Support Vector machines to predict multi-class labels. Here, LDA provides a probability distribution across topics for 1 observation. This means that a procurement contract could be multi-labelled instead. Now, we can suggest possible suppliers to invite during the tender notification process.

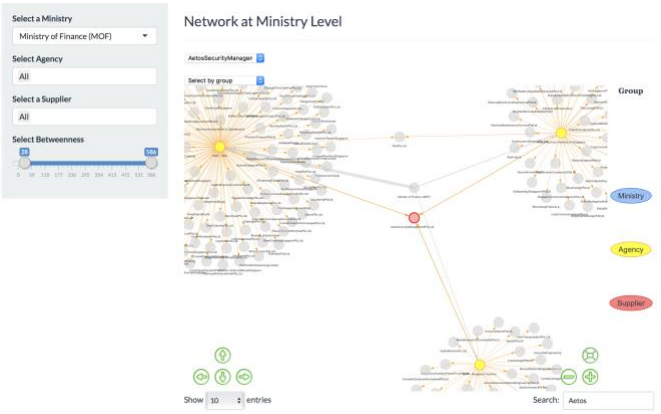


Fig. 13: Filtering by betweenness centrality

With the introduction of the betweenness centrality metric, we will now be able to identify the reliable suppliers that many agencies and their respective ministries are purchasing from. Using the interactive data table, one will be able uncover that Aetos Security Management Pte. Ltd helps with the provision of armed security personnel across government agencies.

Supplier Analysis

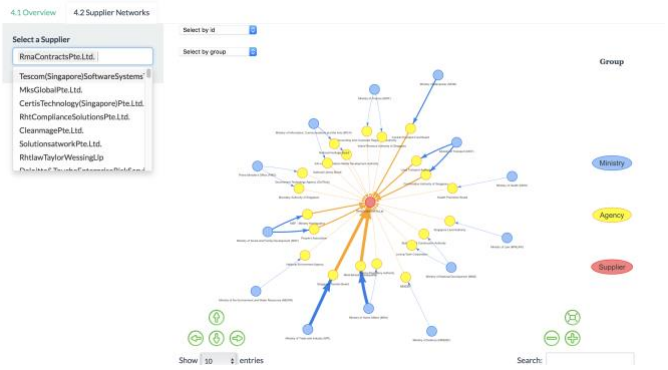


Fig. 14: Network involved with a supplier

If we are interested to find out the Agencies that a particular supplier has previously dealt with, we will also be able to do so on the Supplier Network tab. This is essential for Agencies when they would like to evaluate on potential suppliers for the tenders they have put up. This also fosters collaboration across government Agencies where they could garner feedback on these suppliers which aids them in their decisions.

Similarly, a data table have been included below which will react to the filters chosen and provide the details of past contracts if necessary.

DISCUSSION

We presented our work at the Visual analytics Conference and Poster Presentation held in Singapore Management university on 12-Aug 2018. We showcased the various features of our application and received positive comments on the capability of the Application to

bring up a Network of people associated with various ministries and suppliers.

Members of the audience were particularly impressed by the Network Visualization and one of them commented on the layout of the network visualization which we had, stating that this was one feature he hasn't seen in other relevant works using the visNetwork package for network visualization. Also, many were amazed at the amount of visualizations possible and the smooth interfacing in R Shiny. The audience were reaffirmed that the whole application has been done in pure R.

FUTURE WORK

1. Fixing the view of vizNetwork

When the betweenness scale is adjusted, the network graph's layout changes and it might be difficult to identify the suppliers who are not well connected. Using igraph's layout, we will be able to fix the layout and adjust the filters in visNetwork. This should make it more intuitive for users when they interact with the graph

2. Enhancement of Supplier Information

In the dataset, information about supplier is only the name. It will be beneficial to add supplier information, such as industry and financial health and performance information such as revenue, gross profit, net profit, and change of stock price.

3. Utilize Information Extracted from Text Description

Utilizing the new columns created based on LDA which indicates whether each procurement order is related to the topic, we can filter out the dataset and analyse the pattern in network for each topic.

4. Enhance Information Extraction from Text Description

The dataset includes text description, and more useful information would be able to be extracted. Tuning of topic modelling can be done by removing more words that disturb the output and trying other numbers of topics and the random seed. Other techniques such as Named Entity Recognition could be helpful as well.

ACKNOWLEDGMENTS

The authors greatly thank Dr Tin Seong KAM for his guidance and suggestions.

REFERENCES

[1] GeBIZ. (n.d.). Retrieved August 8, 2018, from https://data.gov.sg/dataset?q=GeBIZ

[2] Defence Science & Technology Agency, "Business Intelligence in Government Procurement – DSTA" (n.d.). Retrieved August 13, 2018, from https://www.dsta.gov.sg/docs/default-source/dsta-about/business-intelligence-in-government-procurement.pdf?sfvrsn=2

[3] Programmes - Civil Service College Singapore. (2018, February 02). Retrieved August 13, 2018, from https://www.cscollege.gov.sg/Programmes/Pages/Display Programme.aspx?ePID=2rblsavgweogwh2qu9lgtunfma

[4] D3treeR. (2018, February 06) Retrieved August 8, 2018, from https://github.com/d3treeR/d3treeR

[5] Sankey Diagram. (n.d.). Retrieved August 8, 2018, from https://plot.ly/r/sankey-diagram/

[6] visNetwork (n.d.). Retrieved August 8, 2018, from https://datastorm-open.github.io/visNetwork/

[7] CRAN - Package lda. (n.d.). Retrieved August 8, 2018, from https://cran.r-project.org/web/packages/lda/

[8] LDAvis (2018, April 25). Retrieved August 8, 2018, from https://github.com/cpsievert/LDAvis

[9] Betweenness Centrality (n.d). Retrieved August 6, 2018 from https://en.wikipedia.org/wiki/Betweenness_centrality

APPENDIX

I. INSTALLATION GUIDE

1.1 Data Preparation:

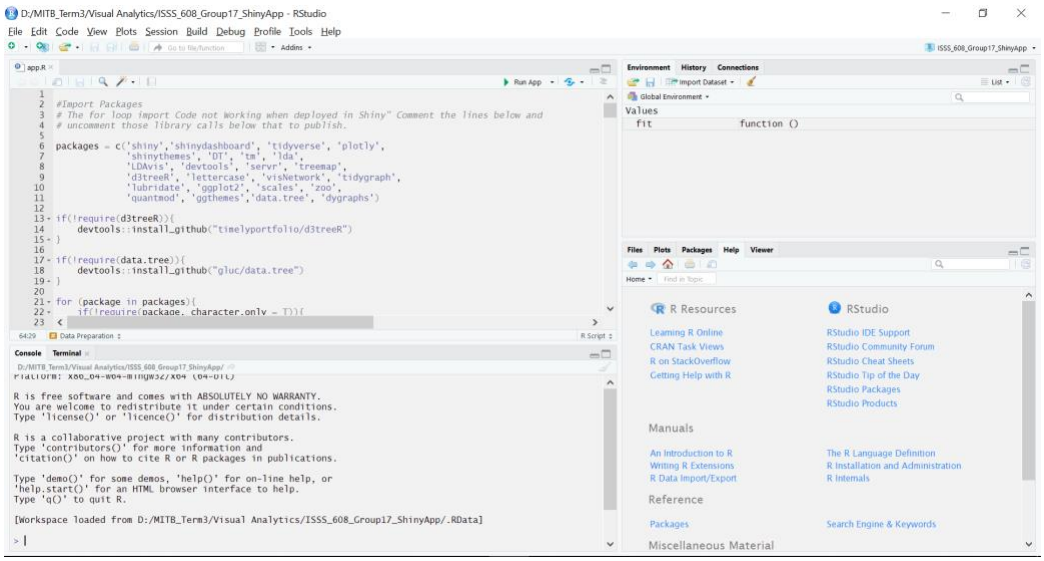
- Datasets for upload need to be in .csv format
- The Dataset from GeBIZ needs to be in .csv format and saved in the data folder of the Project under the name “government-procurement-via-gebiz.csv”
- Ensure that the date column in the data is in the format: "yyyy-mm-dd”

1.2 Online Use:

The recommended web browser would be Google Chrome for accessing our application webpage.
https://isss608-2017-18-t3-group-17.shinyapps.io/ISSS_608_Group17_ShinyApp/
Besides that, no other special tools are required.

1.3 Local Use:

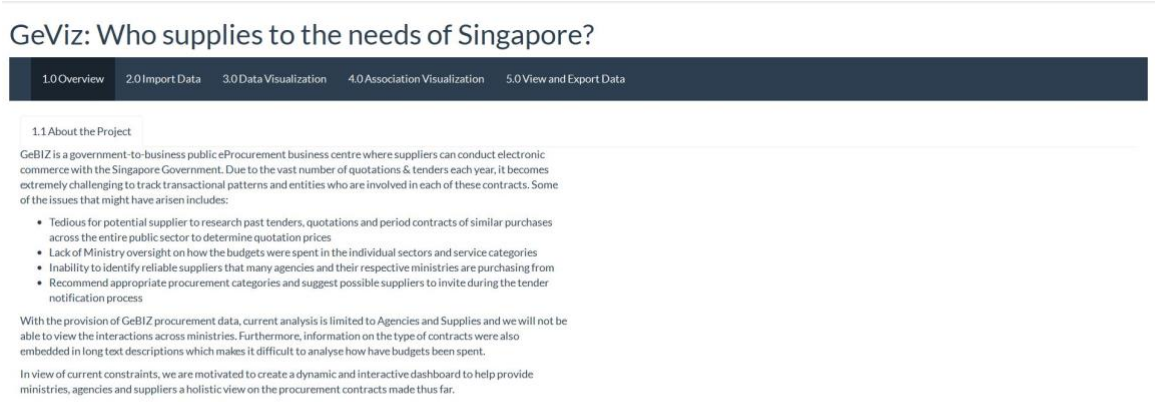
At the point of the project, R version 3.4.1 and R Studio version 1.1.453 was used to create the application.
1) Install R Studio version 1.1.453 from: https://www.rstudio.com/
2) Download the project folder, unzip it and double click on the R project “ISSS_608_Group17_ShinyApp”.



3) Click on the Run App button to run the Application.

II. USER GUIDE

When you launch the App, the first screen that welcomes you will be:



The App has 5 main sections. Each section has its own subsections where required. The Layout is as shown.

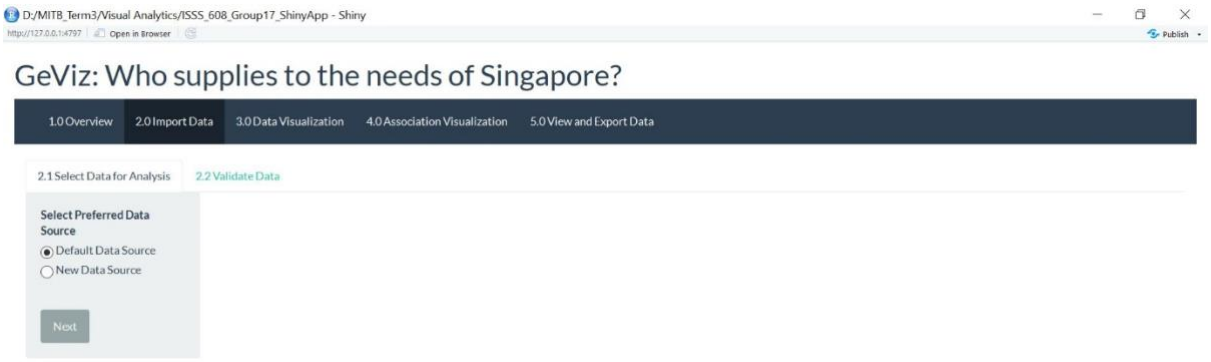
1. Overview
   1.1. About the Project
2. Import Data
   2.1. Select Data for Analysis
   2.2. Validate Data
3. Data Visualization
   3.1. Procurement Expenses
   3.2. Ministry and Suppliers
   3.3. Order patterns
   3.4. Topic Modeling
4. Association Visualization
   4.1. Overview
   4.2. Supplier Networks
5. View and Export Data

1. Overview Page

This page gives a brief introduction to what we are doing in this project.
While we can navigate to different tabs directly from the overview page (As we have incorporated default data), it is recommended that we move to the Import Data page.

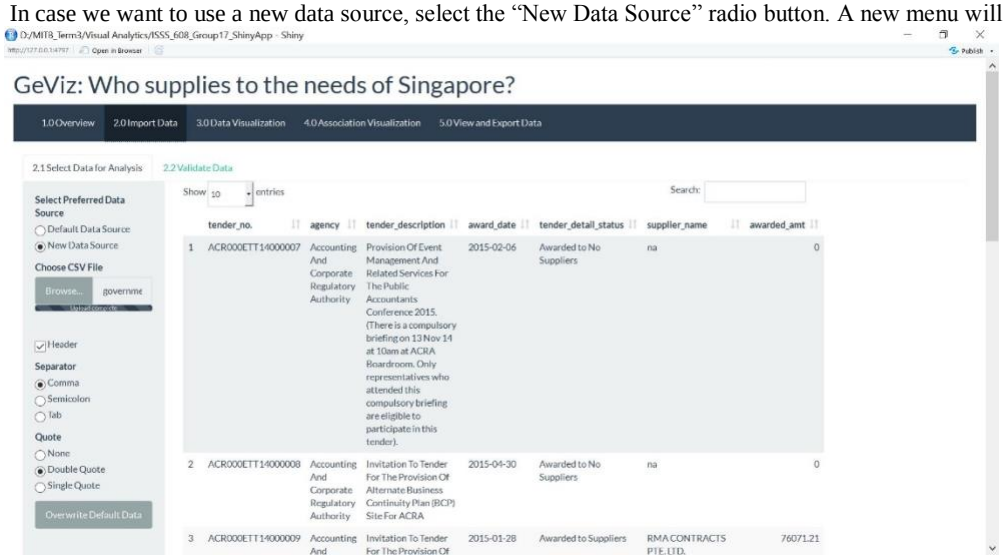
2. Import Data

2.1 Select Data for Analysis



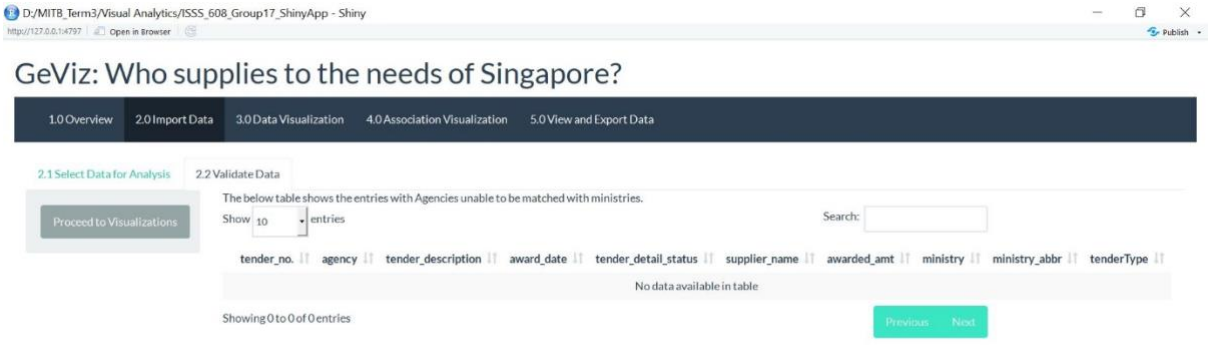
In this page, we decide which dataset to use in our analysis. We have 2 options here.

- Use the Default Data (Data that was used for building the app)
    - Suppose we decide to use the Default data source, you can proceed to click the "next" button to navigate to the Validate data section
- Use an updated version of the dataset downloaded from https://data.gov.sg/dataset?q=GeBIZ
    - In case we want to use a new data source, select the "New Data Source" radio button. A new menu will become visible.



    - In the menu, click on "Browse" button and use the resulting screen to navigate to and select the updated data saved locally in your PC.
    - Once it is selected, the app loads it and shows a preview of how it appears. You can use the options to adjust the import until it looks as shown above.
        - Header: Specify whether the dataset contains a header column. (Ideally it should have one.)
        - Separator: Specify what is the separator in CSV file.
        - Quote: Specify the format of quotes in the data.
    - Overwrite Default Data: This option allows you to overwrite the default data with the data you have currently loaded.
    - Once the data looks okay, click on the "next" button to navigate to the next page.

2.2 Validate Data



This page shows you if here were any agencies in your data, that was not able to be mapped onto ministries. You will have to manually find the ministries for these agencies and update the "MinistryReference.csv" file.
If the data looks proper, we can navigate to the Procurement Expenses page in the Data Visualizations tab using the button "Proceed to Visualizations".

3. Data visualization

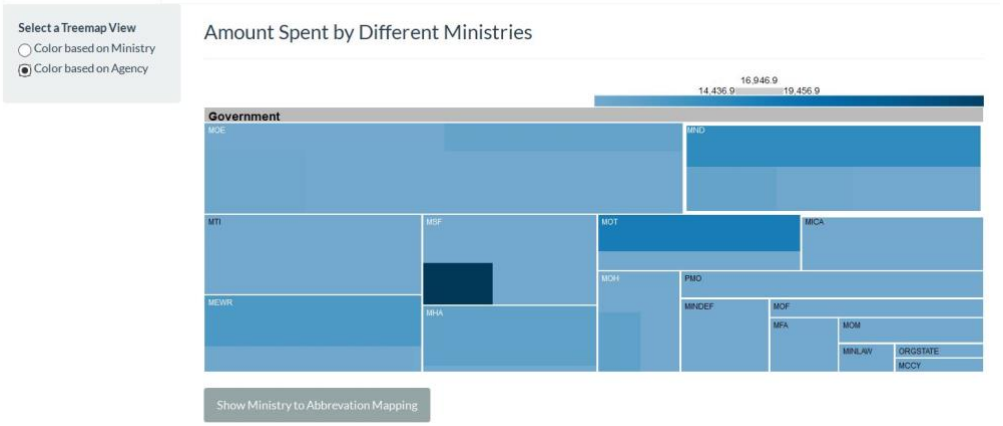
3.2 Procurement Expenses

The section has the procurement related expenses visualized in Tree map as well as Sankey format. The visualizations are under 2 sections

- Amount Spent by Different Ministries
    - Default View is the Color Based on Ministry

    - If you Hover over the tiles, the underlying Agency level tiles are visible. If you click on a tile you can zoom in. by clicking on the title, E.g.: "Government.MSF", you can navigate back to the normal view.

- If you want to see all the agency level tiles at the same time, click on "Color based on Agency".



- If you click on "Show Ministry to Abbreviation Mapping", you can get a pop up screen which shows you the Ministry names and their corresponding abbreviations.

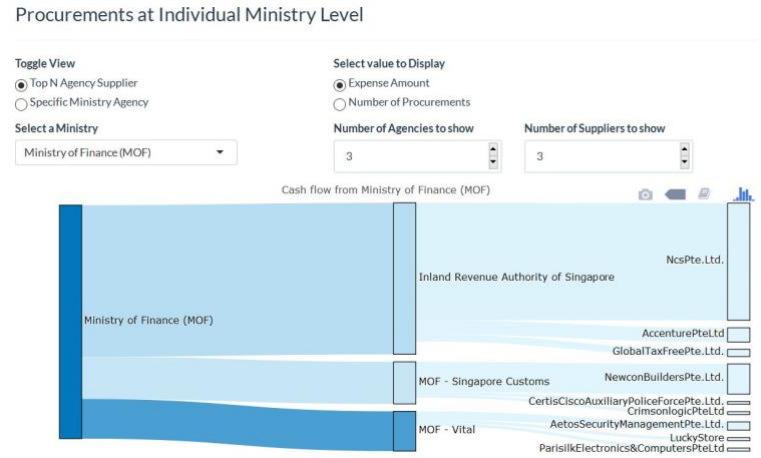
Ministry to Abbrevation Mapping

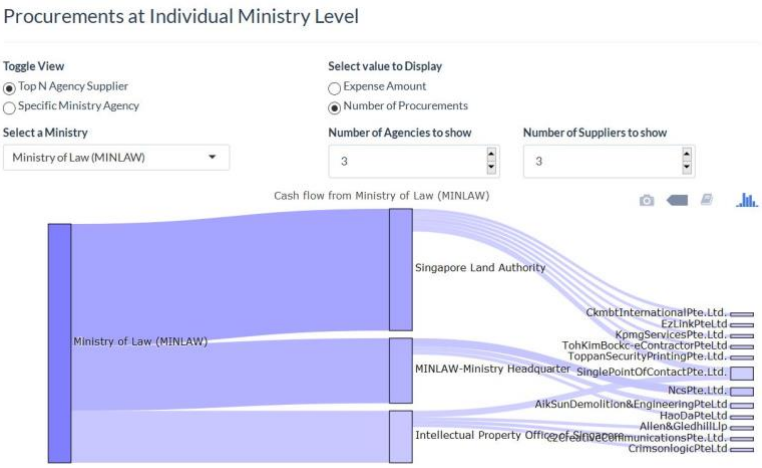
Ministry	Abbreviation
Ministry of Finance (MOF)	MOF
Organs of State	ORGSTATE
Ministry of National Development (MND)	MND
Ministry of Transport (MOT)	MOT
Ministry of Trade and Industry (MTI)	MTI
Ministry of Culture, Community and Youth (MCCY)	MCCY
Ministry of Social and Family Development (MSF)	MSF
Ministry of Education (MOE)	MOE
Ministry of Manpower (MOM)	MOM
Ministry of Home Affairs (MHA)	MHA
Prime Minister's Office (PMO)	PMO
Ministry of Defence (MINDEF)	MINDEF
Ministry of the Environment and Water Resources (MEWR)	MEWR
Ministry of Foreign Affairs (MFA)	MFA
Ministry of Health (MOH)	MOH
Ministry of Information, Communications and the Arts (MICA)	MICA
Ministry of Law (MINLAW)	MINLAW

Dismiss

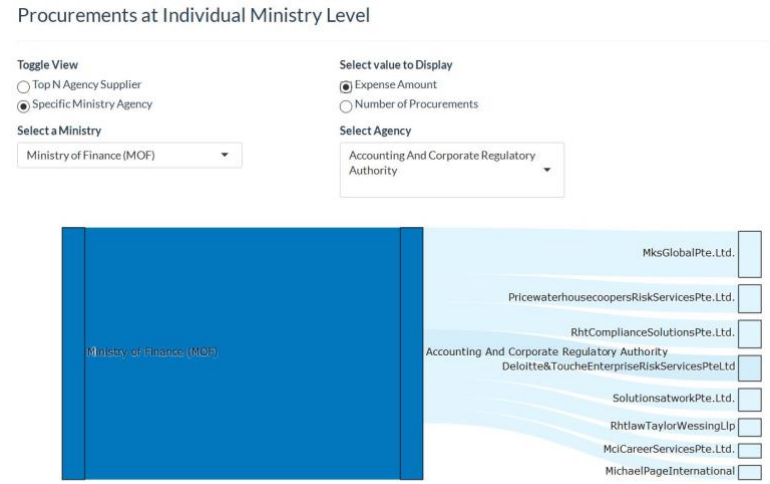
- Procurements at Individual Ministry Level

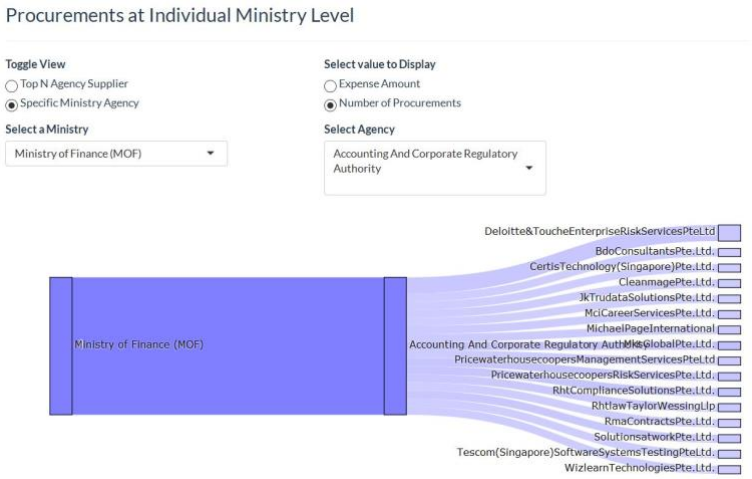
- Here, we visualize the flow of cash and number of orders using a Sankey diagram. The controls given help determine the view.
- By default, it is set to Top N Supplier and Top n Agency and to show the expense amount.
  - We can select a single ministry.
  - We can select the up to Nth top agency to view
  - We can select the up to Nth top supplier to view



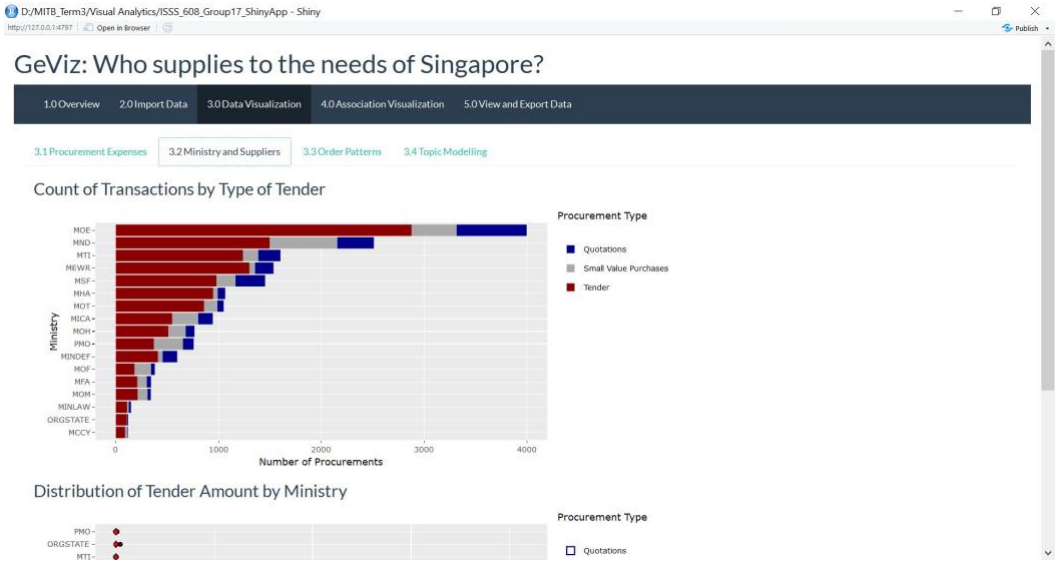


- We can also choose to visualize by specific Ministry and Agency.
  - We can select the ministry
  - We can select the agency



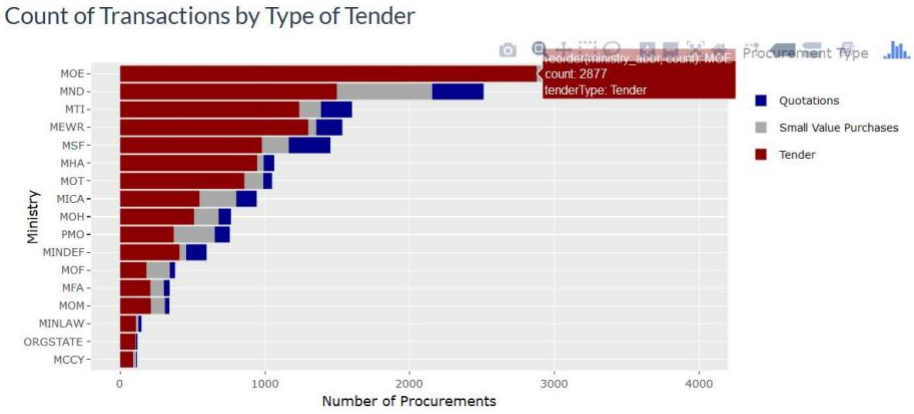


1.1. Ministry and Suppliers

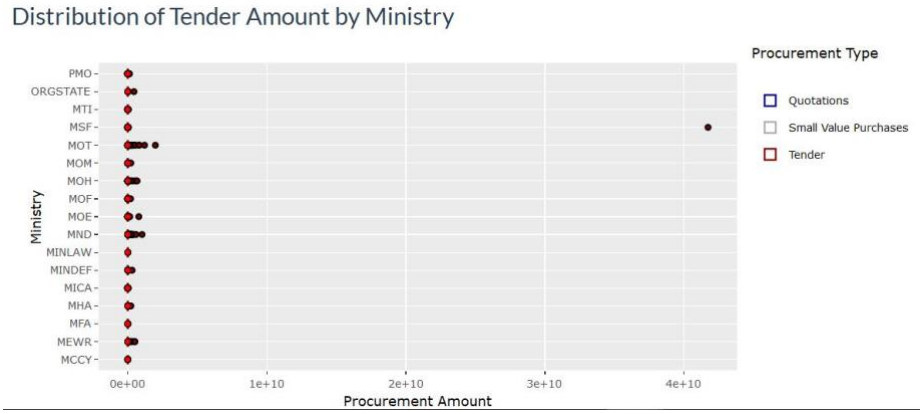


In this screen we have 2 visualizations. They are,

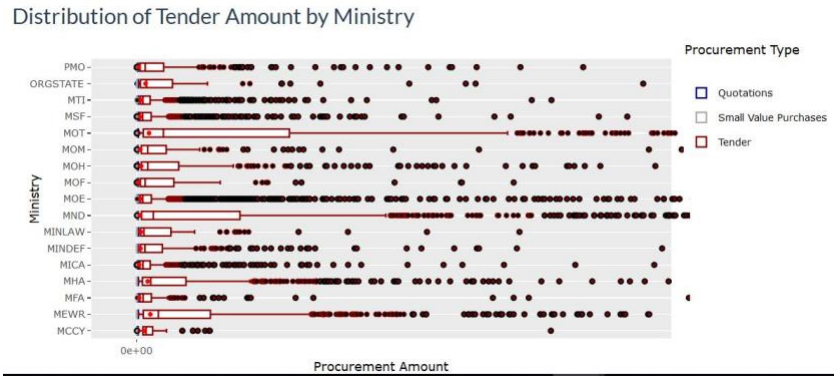
- Count of Transactions by type of Tender



- Distribution of Tender Amount by Ministry
    - This is a box plot of the amounts disbursed in individual tenders.

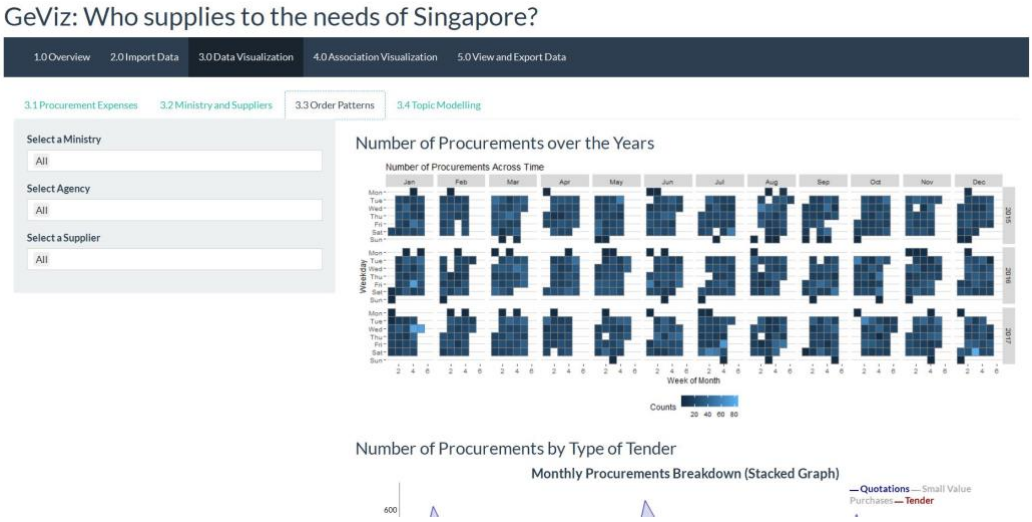


- We can zoom in on the values on the left.



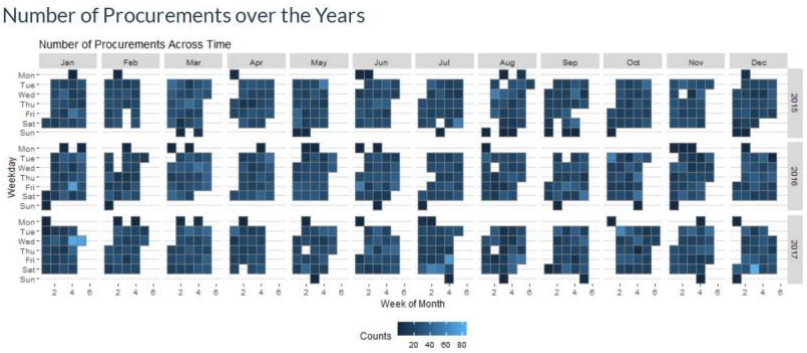
1.2. Order Patterns

In this screen we explore the variations in procurements over time. Mainly their counts on a weekly and monthly level.

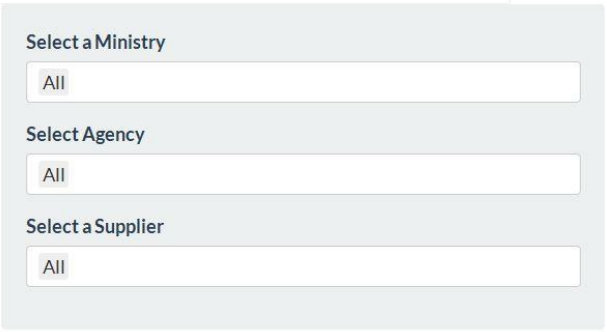


There are 2 visualizations.

- Number of Procurements over the Years
  - This one shows the weekly count of tenders issued.

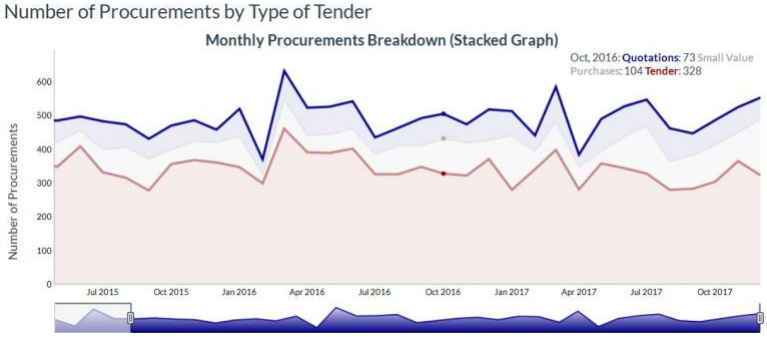


- This can alo be controlled using the panel given



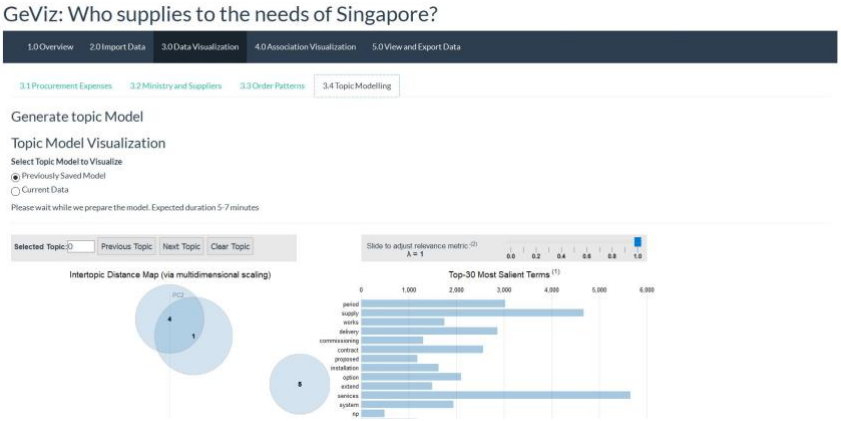
- Using the above panel, you can select
  - Multiple Ministries
  - Multiple Agencies
  - Multiple Suppliers

- Number of Procurements by Type of Tender
  - It shows the number of Procurements split by type on a month level. It is interactive, and you can adjust the slider below the graph for adjusting the view.

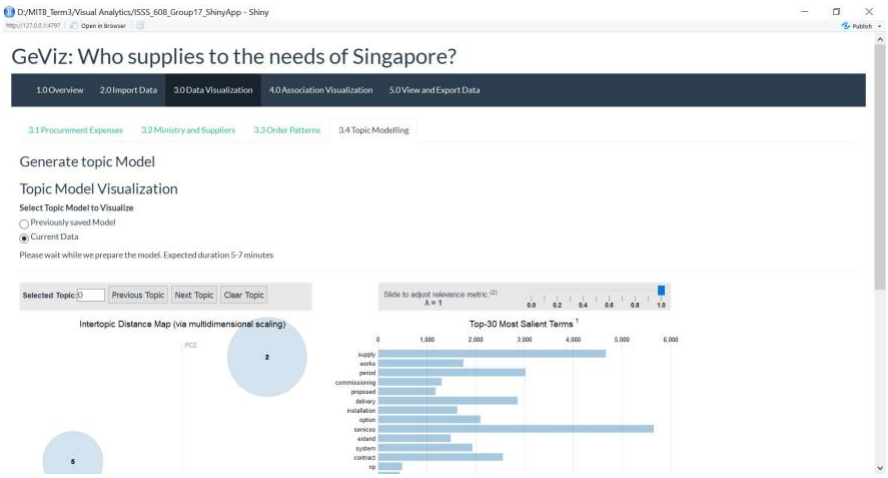


1.3. Topic Modeling

This Screen allows us to build and visualize the LDA based topic model. As the model building takes more than 6 minutes on average, we have used a previously saved model to visualize. When you are using the default data, there is no action required from the user.



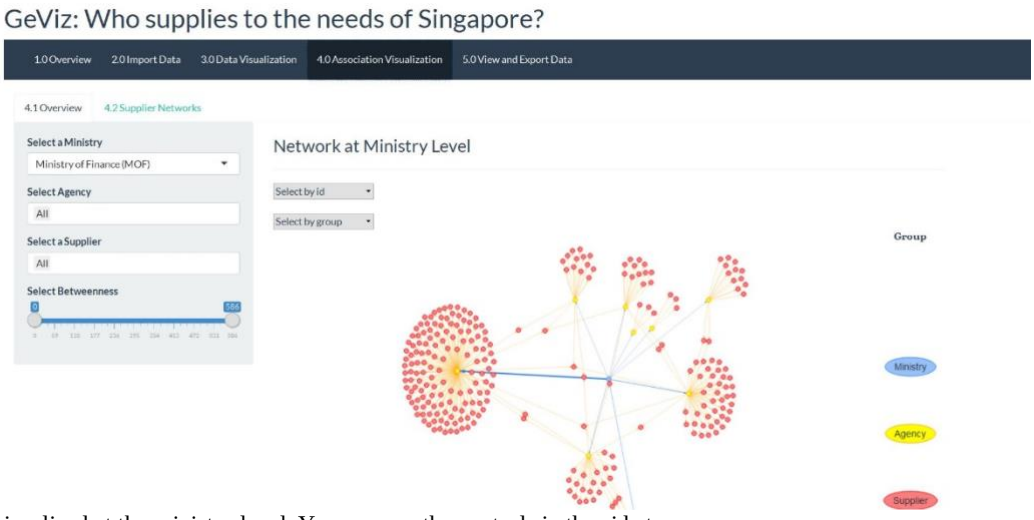
In case we want to build and visualize using the current data, click on “Current Data”. The system will start generating the model. There is a possibility that the display will not change until the model is generated and visualized.



2. Association Visualization

This section uses network visualizations to visualize the connections between ministry, agency and supplier.

2.1. Overview



Here, the data is visualized at the ministry level. You can use the controls in the side to

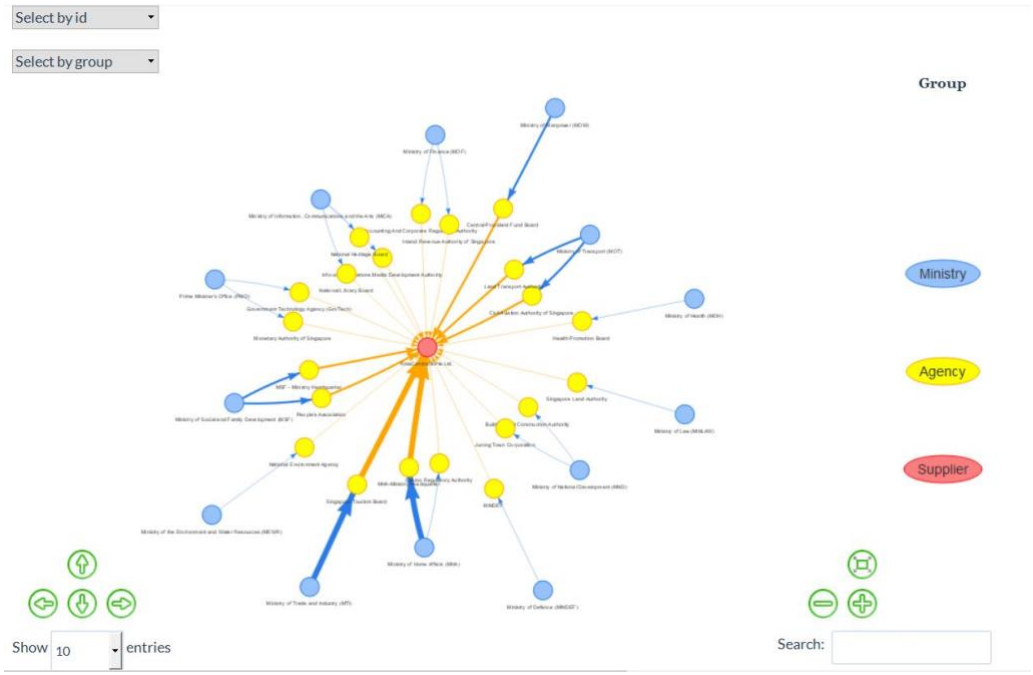
- Select the ministry
- Select Multiple agencies in that ministry
- Select Suppliers associated with that ministry.
- To filter based on their betweenness.

The selected data is visible below the graph.

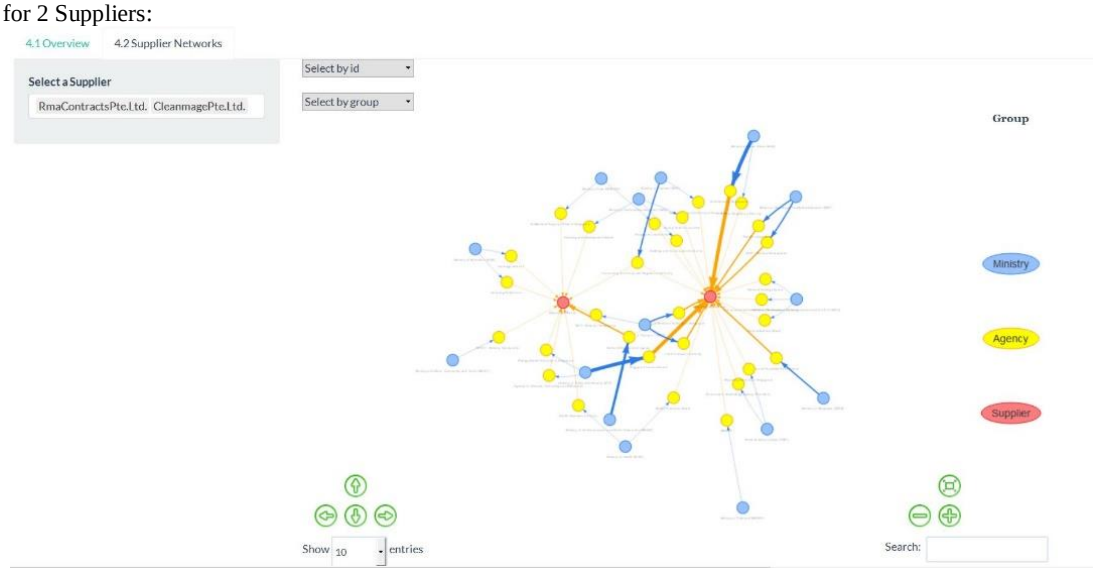
2.2. Supplier Networks

In this page, we visualize the network based on selected suppliers. By using the Controls provided you can add more supplier networks into the graph. The selected data is visible below the graph.

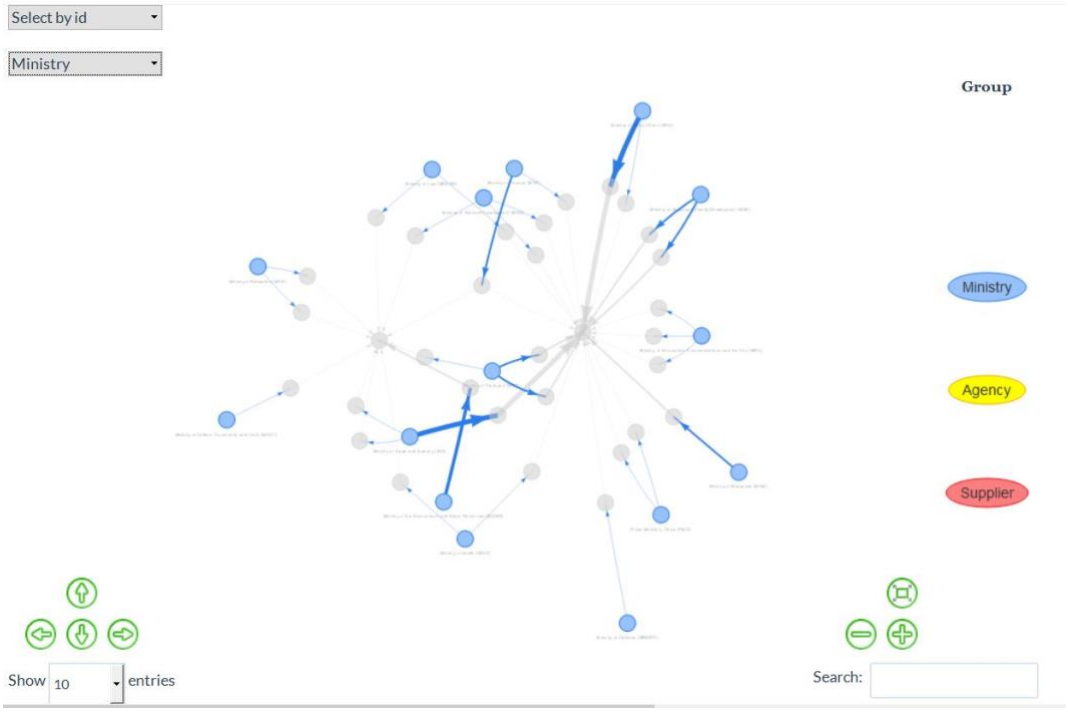
Sample View for Single Supplier:



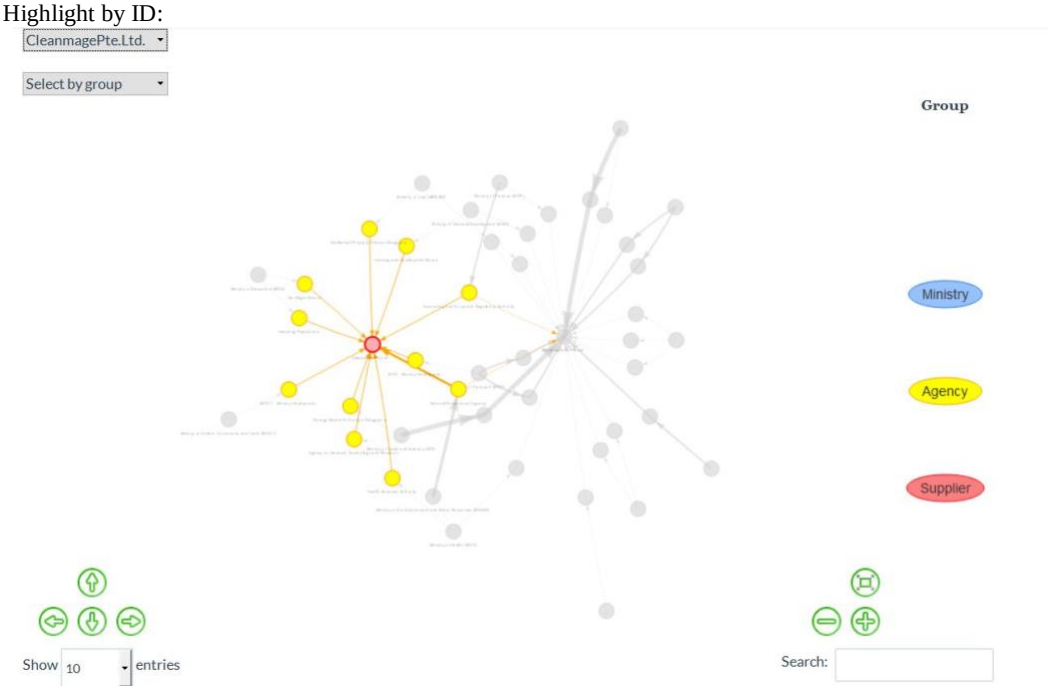
Sample View for 2 Suppliers:



Sample View of Highlight by Ministry:



Sample view of Highlight by ID:



3. View and Export Data

This section is for viewing the data used for analysis and saving them to a location in your PC if required.

You can:

- Select the number of rows to display
- Search for records based on keywords
- Export Data using the export button

By pressing the “Export Data” button, the current data will be saved to
File: [Project Folder Path]/data/Saved/government-procurement-via-gebiz_export.csv