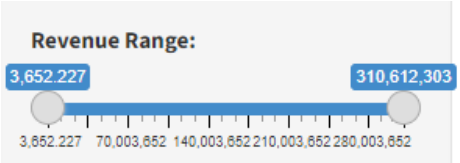
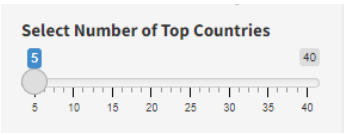
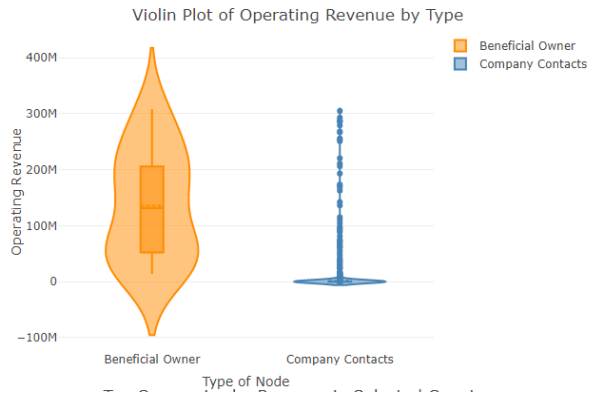
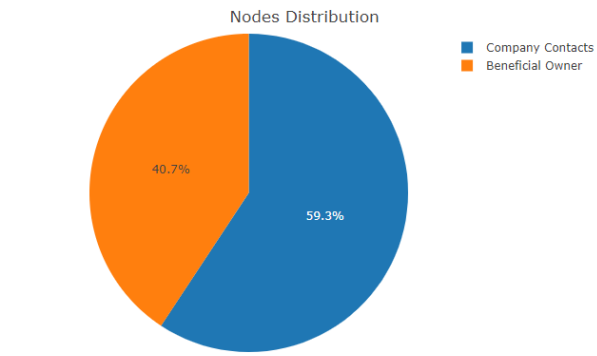
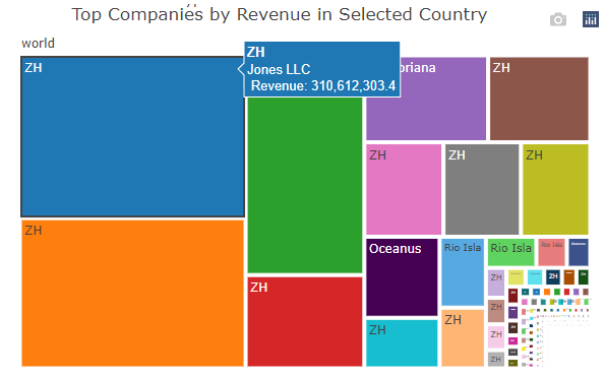
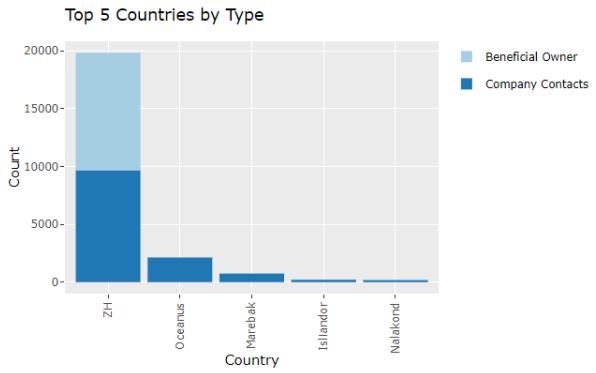
Group 5 Rshiny Application User Guide

Back to Netlify

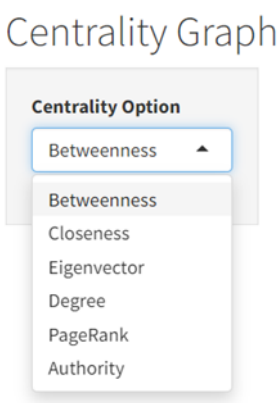
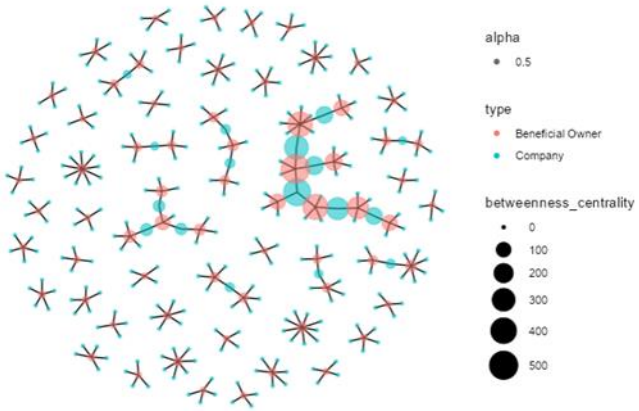
1. EDA

Screenshot	Description
Input	
	Drop-down Selection User can choose the country of interest, else they can see all countries as an overview. This selection impacts the violin plot, pie chart, and treemap.
	Slider User can choose the range of revenue that they would like to zoom into. This selection impacts the violin plot and pie chart.
	Slider User can choose the number of top countries that the user would like to observe on. This selection impacts the treemap and histogram.
Output	
	Violin Plot The violin plot allows users to look at anomalous points and by hovering the mouse on the points that falls out of the violin plot, user will be able to see the “id” associated with that anomalous point.
	Pie Chart For every node, there can be two types, either it is a Company Contact or Beneficial Owner. The breakdown follows the country selection chosen so that the user can see the breakdown of Company Contacts vs Beneficial Owners in a given selection.
	Treemap Depending on the country selection, and the number of top countries selected, the treemap will show in terms of revenue proportion of each id. The display will be labelled by the country. By hovering over the box, you will be able to see the respective country, id, and company revenue. This allows the user to glimpse into the big players at an overview, and also within the country selected.

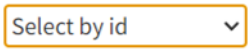
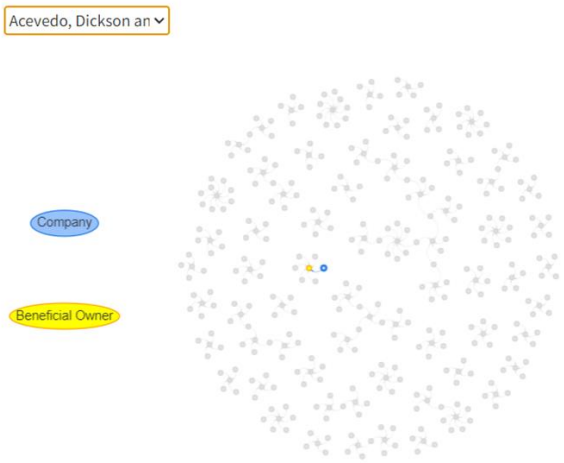
Screenshot	Description
Output 	<u>Histogram</u> Each bar in the histogram represents a country, and the height of the bar indicates the count of companies. The different colors in each bar represent the different types of companies.

2. Network Graph

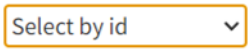
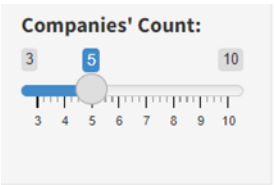
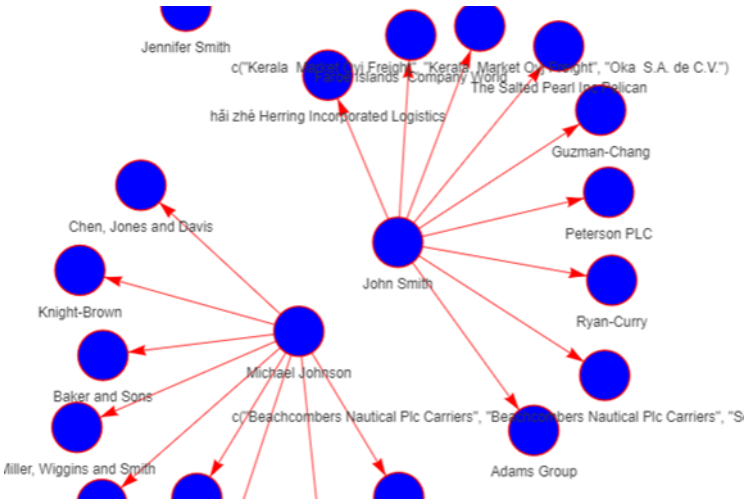
i. Centrality Graph

Screenshot	Description
<u>Input</u> 	<u>Drop-down Selection</u> User can choose the Centrality Options. Options includes “Betweenness”, “Closeness”, ”Eigenvector”, “Degree”, PageRank”, “Authority”.
<u>Output</u> 	<u>Network Graph</u> Through this graph, users are able to identify the nodes with the highest score of based on the centrality selection. The nodes are also labelled with the id, and the bigger the id, the larger the centrality score

ii. Network Connection

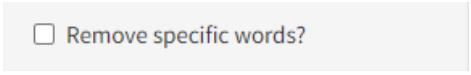
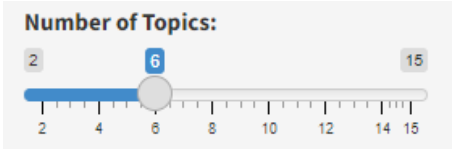
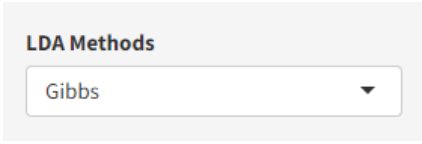
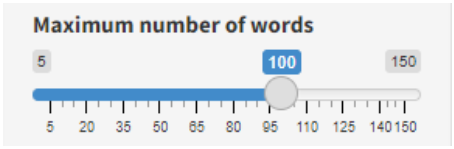
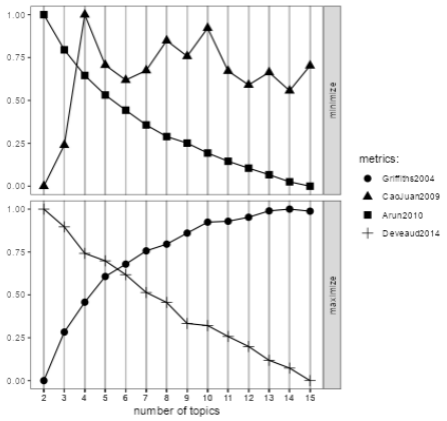
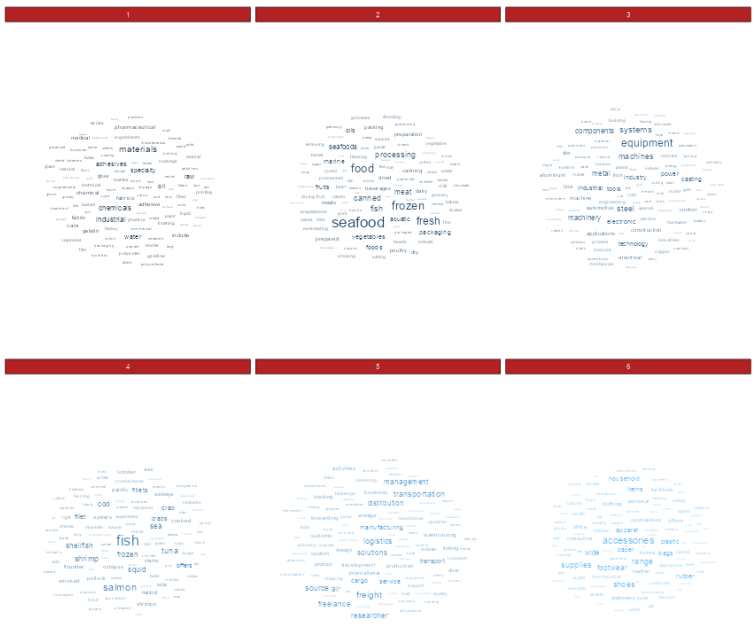
Screenshot	Description
Input	
	Drop-down Selection User can choose the id, and the graph will change the focus to the id selected.
Output	
	Network Graph By zooming into the graph, the names of id will appear to have a better view on what other connections are being connected to the selected id.

iii. Company Count

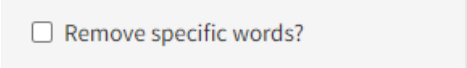
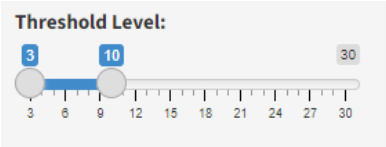
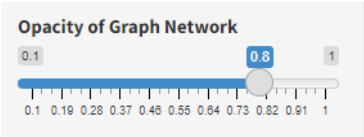
Screenshot	Description
Input	
	Drop-down Selection User can choose the id, and the graph will change the focus to the id selected.
	Slider The user can filter the minimum number of companies so that the users can view only those nodes that fulfil the criteria.
Output	
	Network Graph By zooming into the graph, the names of id will appear to have a better view on what other connections are being connected to the selected id.

3. Text Analysis

i. unigram

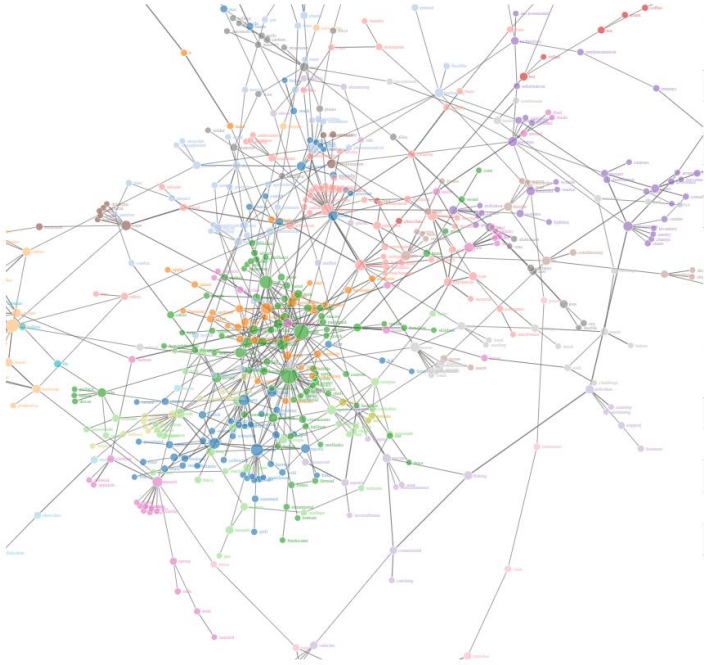
Screenshot	Description
Inputs	
	<u>Checkbox</u> Click on checkbox to include other stopwords that user feels should also be removed. (Max 10 additional words)
	<u>Slider</u> User can choose the number of topics to be produced. Output will be generated as a wordcloud for each topic.
	<u>Drop-down Selection</u> User can choose the type of LDA method. Options includes “Gibbs” and “VEM”.
	<u>Slider</u> User can choose the max number of words to be produced in the wordcloud.
Outputs	
	<u>LDA Optimizer Graph</u> Graph to show ideal number of topics via intersection point. In this example, the ideal number of topics will be either 5 or 6.
	<u>Wordcloud</u> Wordcloud will be rendered as part of the output to show the associated words for each topic. The size of the word is related to the frequency of occurrence. i.e. big words = high occurrence

ii. bigram

Screenshot	Description
Inputs	
	<u>Checkbox</u> Click on checkbox to include other stopwords that user feels should also be removed. (Max 10 additional words)
	<u>Slider</u> User can choose the threshold level for the network graph. This allows the user to zoom in to the range of weight of each word that they want to focus on.
	<u>Slider</u> User can choose the opacity level for the network graph.
	<u>Drop-down Selection</u> User can choose the type of clustering method. Options includes "Louvain", "Edge Betweenness", "Walktrap", "Infomap".
	<u>Drop-down Selection</u> User can choose the Centrality method. Options includes "Degree", "Betweenness", "Closeness".

Screenshot	Description

Outputs



Network Graph

This shows the connection between the words in product_services description. The connection between words shows the relationship of occurrence between the two connected words.

Colors depict the clusters that will be churned based on the cluster option chosen.

Depending on the centrality option chosen, the size of the nodes will change.

Bigger nodes means higher centrality score.

Show 5 entries Search:

	word1	word2	weight
1	frozen	seafood	74
2	frozen	fish	50
3	freight	forwarding	41
4	fresh	shellfish	37
5	marine	fats	36

Showing 1 to 5 of 20,872 entries

Previous 1 2 3 4 5 … 4,175 Next

Datatables

The first datatable output shows the bigram breakdown as well as the weight between the two occurring words.

This allows user to see the breakdown of bigram words included in the network, and choose additional stopwords should they feel some of the words are redundant.

Show 5 entries Search:

	word	degree	closeness	betweenness
1	black	17	0.0000606759298586251	592
2	building	10	0.0000505612296491051	0
3	canned	68	0.0000945000945000945	10356.4626984127
4	casual	23	0.0000722334585379948	0
5	cooked	26	0.0000895575855274942	12256.4886554622

Showing 1 to 5 of 594 entries

Previous 1 2 3 4 5 … 119 Next

The second datatable shows the centrality scores for each node present in the network graph. This allows user to toggle according to the centrality score to see which words have higher/lower centrality scores.