ISSS608 Visualisation Analytics Project Minutes

Group 4 members:
farah.foo.2020@mitb.smu.edu.sg
karishmay.2020@mitb.smu.edu.sg
yanrutan.2020@mitb.smu.edu.sg
Minutes taker: Farah

Date and Time: 28 Feb 2022 8pm

Attendees: all

#	Key points discussed and actions	Action Owner	TCD
1	All agree to use MS word for recording of minutes.		
2	**Brainstorm on possible topic for project:** 1. Yanru suggested the project be on the stock market, which can consist of horizon chart or individual stock time series, and time series prediction. Source can be from yahoo finance. Can split the timeseries into trend and seasonality. 1. Farah and Karishma agree the project can be on stock market. 2. Karishma provided a similar reference https://anhhoangbui.shinyapps.io/ISSS608VAStock/ 2. Yanru also suggested can explore on supply chain data and inventory optimization 3. Karishma also suggested can explore on network analysis which prof has yet to cover in week 8 https://isss608-ay2021-22t2.netlify.app/outline/lesson08_outline https://snap.stanford.edu/data/		
3	To schedule a time with prof on Saturday or Friday via google sheet	Karishma	5 Mar 22
4	To set up the chat on whatsapp as our 1st minutes	Farah	1 Mar 22
5	Take 1 day to look for possible data sets To reconvene tomorrow 7pm	all	1 Mar 22

Date and Time: 1 March 2022 7pm -7.30pm

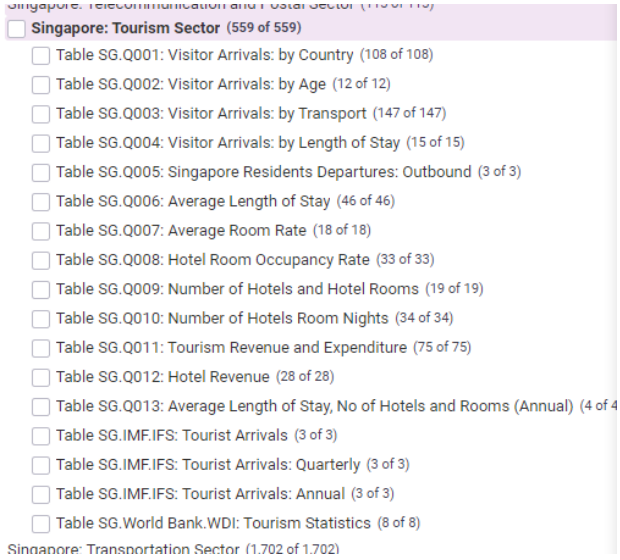
Attendees: all

#	Key points discussed and actions	Action Owner	TCD
1	Karishma has booked prof time on 5 Mar 3.30pm.		
2	**Possible project ideas were discussed by all:** **a) Public Housing / rental data set:** Time series Geospatial map stuff Price prediction Can be obtained here: https://www.ura.gov.sg/realEstateIIWeb/resiRental/search.action **b) Stock market:** Time series Pattern **c) Population:** Age-sex pyramid Geospatial how age varies in the area Time series on population change in the area **d) Commerce supply chain:** Prices of goods Network analysis **e) Airline inflight passenger into Singapore:** Time series Geospatial Animation on countries / population Before and after covid Cannot do prediction		
3	To think about which topic to select and regroup	All	5 Mar 22

Date and Time: 5 March 2022 12pm – 3pm

Attendees: all

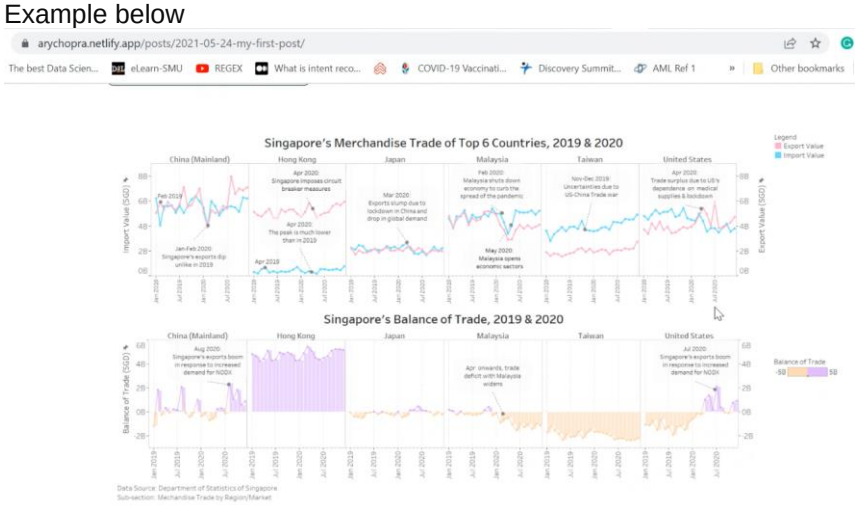
Discussion has continued on whatsapp between 1-4th March

#	Key points discussed and actions	Action Owner	TCD
1	Yanru will set up the Netlify and github.	Yanru	12 Mar 22
2	**Brainstorm for project ideas and concurrently researching:** Expanding on the idea of inbound travellers dataset, all agree that we should have 1 section for descriptive analytics and 2nd section for statistical. Karishma shared this link for inspiration https://www.unwto.org/country-profile-inbound-tourism, to do a map of the world and show number of tourists - But since data is yearly then the country map will only have yearly display and not more detailed. Yanru added we can consider hotels also if there is data. Karishma checked the hotel data is only for 6 month period,and yearly aggregated for 5 years. Available data sets are: - Datagov: 1. Monthly Visitor arrivals into Singapore - Datagov: 2. Monthly Age-Sex Arrivals into Singapore (not per country) - Datagov: 3. Monthly International visitors by length of stay (not per country) - CEIC: there's visitor arrival by country then also by transport like sea, air or land - CEIC: total hotel revenues - CEIC: tourism receipts and expenditure by country https://insights-ceicdata-com.libproxy.smu.edu.sg/Untitled-insight/myseries  Yanru suggested possible expansions: - Comparison of travellers from different countries,, break down by time series into monthly cycles and seasonality and trend - horizon plot for maybe a selected number of countries + Trend during Covid period - time series prediction based on pre-Covid period Clustering for spending is not possible as data is only on yearly basis.		
3	To check with prof on Saturday if the the proposed scope is too large: Mainly scope: Exploratory, Horizon plot, timeseries, 1 on inbound travellers,		

#	Key points discussed and actions	Action Owner	TCD
	1 on tourism spending To choose common countries in all data sets, and same time set. Visualization: timeseries Cycle plot , Box plot, horizon chart, time-series co-variation (with selection) Allow filtering by country and age - Slope graph for tourism revenue? - if got time then calendar heatmap - Time series auto correlation - https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775		
3	To store data set in teams for sharing and reference	all	5 Mar 22

Date and Time: 5 March 2022 3pm – 4pm.

Attendees: all

#	Key points discussed and actions	Action Owner	TCD
1	Post Consultation with prof on our plan in #1, group discussion to choose the topic the group wanted for the project. **1) Tourist data** If forecasting model, then do 2010 to 2019, but actual patterns will be very different from actual, so will not have much value add and useful context, so compare against actual 2020 and 2021 and conclude how much covid has impact the tourist arrivals. Tourist arrival data have clear seasonal pattern if you have long term period. To target for Cycle plot & Horizon graph **2) Trading data** More interesting to look into the internal trading /trade balance patterns of Singapore across the last 3 to 4 years. There will be some insights that beat expectation. Get trade data from Singstat website on 2018, 2019, 2020, 2021. https://www.singstat.gov.sg/find-data/search-by-theme/trade-and-investment/merchandise-trade/latest-data Data will be on import, export, non oil import and export, with the main trading partners of Singapore. Trading data will not be seasonal. a) Country - Look at trade balance to see how trading changed. Can use Slope graph for point to point comparison, not much stat analysis that you can incorporate into this Choose either B) or C) b) Country – descriptive / exploratory graphs Example below  c) Commodities - Look at trade balance to see how trading changed. Can use Slope graph for point to point comparison, not much stat analysis that you can incorporate into this d) Explore time series clustering (time series running, to detect time series patterns and covariate patterns) to compare before covid and during covid. Prof has not tested so not sure if there are good results but it should work. Refer here for the R package: https://github.com/asardaes/dtwclust **3) Stock Market** Time series will be clearly shown if you use stock market data		
	All agreed to go with #2, focusing on 2a, 2b, 2d. To refer to the data set here https://www.singstat.gov.sg/find-data/search-by-theme/trade-and-	All	6 Mar 22

	investment/merchandise-trade/latest-data and think of the storyboard / additional graphs we want to show. To regroup later / tomorrow for the project proposal.		

Date and Time: 6 March 2022 11am – 1pm.

Attendees: all

#	Key points discussed and actions	Action Owner	TCD
1	Based on the previous consultation with Prof, Farah drafted the project proposal write up and shared the draft for Karishma and Yanru to add in the analytics sections Farah also shared the drafted story board using Tableau as a mock up of what the group will try to achieve. Karishma will review the contents and updated into netlify and will submit the proposal before the 2359 deadline today.	Karishma	6 Mar 22
2	To start work on the farmiliarising with the data and exploring charts in R	All.	12 Mar 22

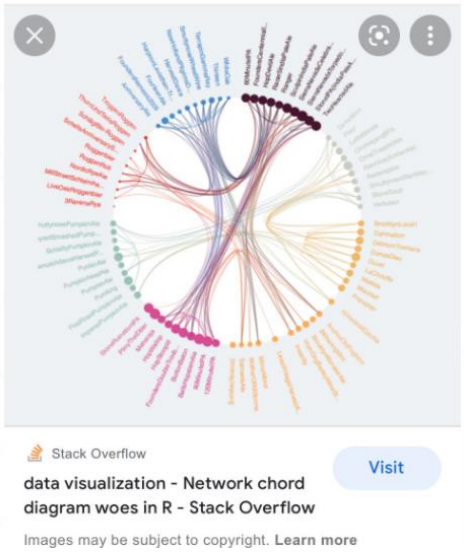
Date and Time: 12 March 2022 7pm -7.30pm

Attendees: all

#	Key points discussed and actions	Action Owner	TCD
1	Karishma shared the import and export data set prepared in R. Karishma also set up the github repository for the project. All to accept the invite and test out the access.	All	12 Mar 22
2	In the split of work, - Yanru volunteered to try the time series clustering and will keep in mind if forecasting can be done. - Karishma suggested a trade map / network map to visualize Singapore's trade to and fro other trading partners. - Farah will explore some graphs for the descriptive analytics To regroup next week	All	19 Mar 22

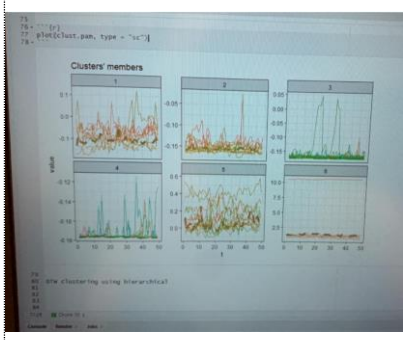
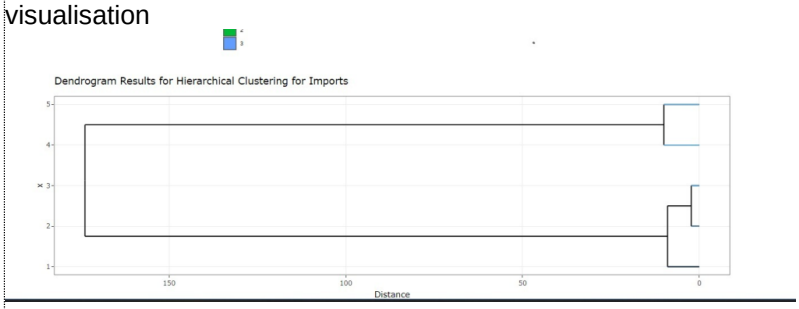
Date and Time: 19 March 2022 3.30pm - 4pm

Attendees: all

#	Key points discussed and actions	Action Owner	TCD
1	Farah shared the basic chart r shiny showing trade values by month.  Yanru suggested to do a lineplot with slider to control time frame, slider should be monthly and year-month. Farah will try to change it to a lineplot.	Farah	26 Mar
2	Yanru feedback that for the timeseries clustering, data format need to be prepared where year-month is in columns across. Will try to prepare and plot out Karishma recommended to have the option to choose by countries, however Yanru reminded that a min number of countries is required to do clustering, so user selection should be based on region. Karishma agree.	Yanru	26 Mar
3	Karishma shared the idea od creating coord plot for trade flows to see by geographical region. There will only be 1 inflow since data is from Singapore perspective  Karishma will explore.	Karishma	26 Mar

Date and Time: 26 March 2022 3.30pm - 4pm

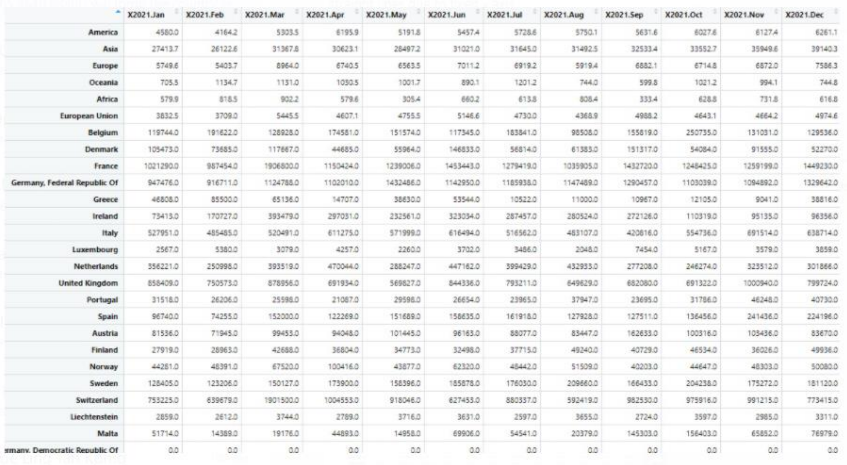
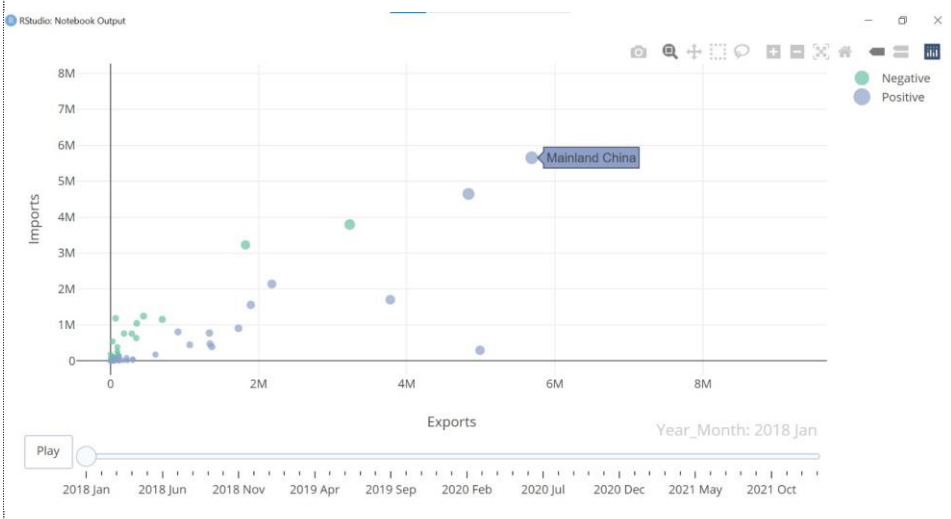
Attendees: all

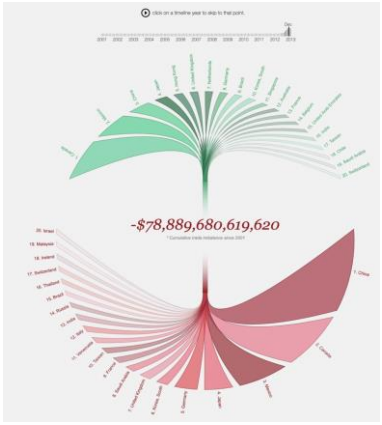
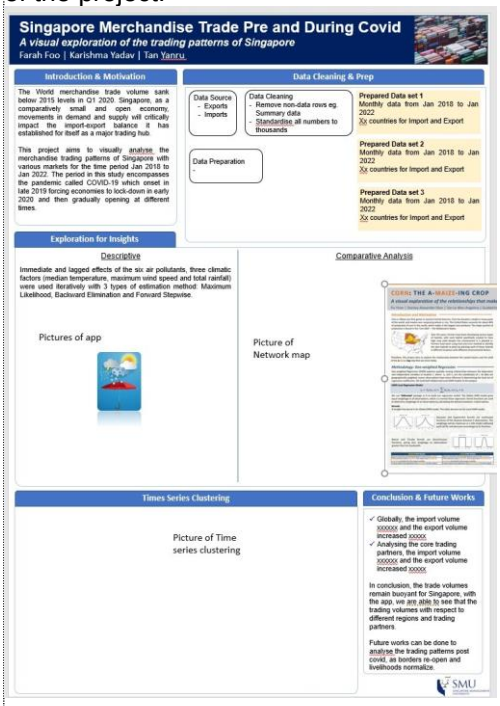
#	Key points discussed and actions	Action Owner	TCD
1	Yanru updated she managed to run a dtw clustering, but work in progress to interpret the results. Considering whether to put number of clusters and countries as a selection for user. 	Yanru	29 Mar
2	Yanru have also explored hierarchical visualization and will continue to do so for more visualisation 	Yanru	29 Mar
3	Farah highlighted challenges in putting in Shiny coding. Yanru advised to get the plots out in ggplot first, then only integrate to Shiny later Farah cited challenges in putting into line graph and shared what was prepared instead. This graph shows that the trade values for the 4 selected trading partners never move drastically pre and during covid years.  Karishma suggested monthwise variation and Yanru suggested to consider animated bubble plot. Farah explained the month plot is not easy to see the impact in trading volume as per below.	Farah	29 Mar

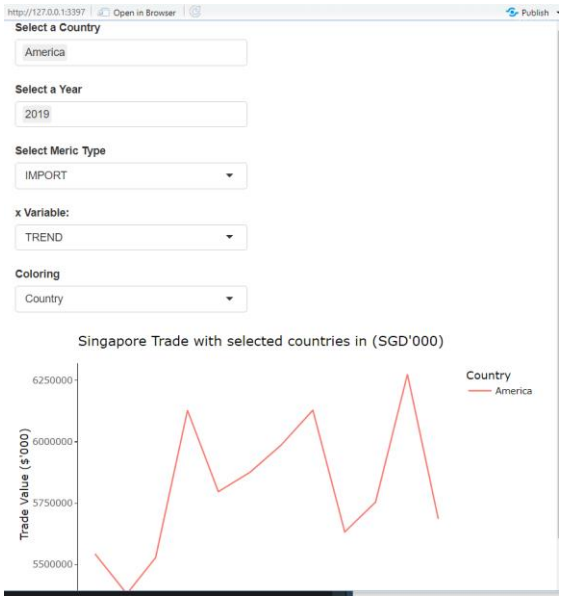
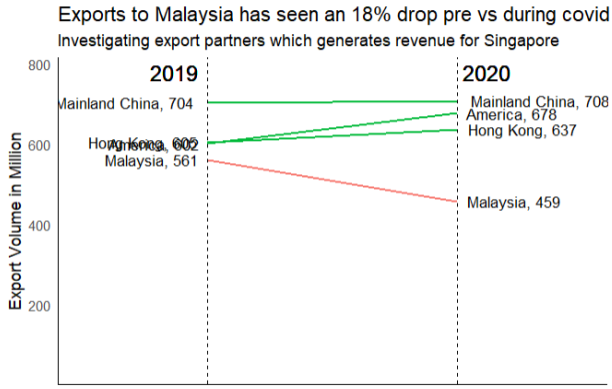
#	Key points discussed and actions	Action Owner	TCD
	 Karishma suggested to adjust the y-axis to show the variations more clearly and use line plots instead of bars.		
4	Karishma updated that she needs time for the other projects and exams. Karishma has set up the design of the app with horizontal tabs across.		

Date and Time: 29 March 2022 3.30pm - 4pm

Attendees: all

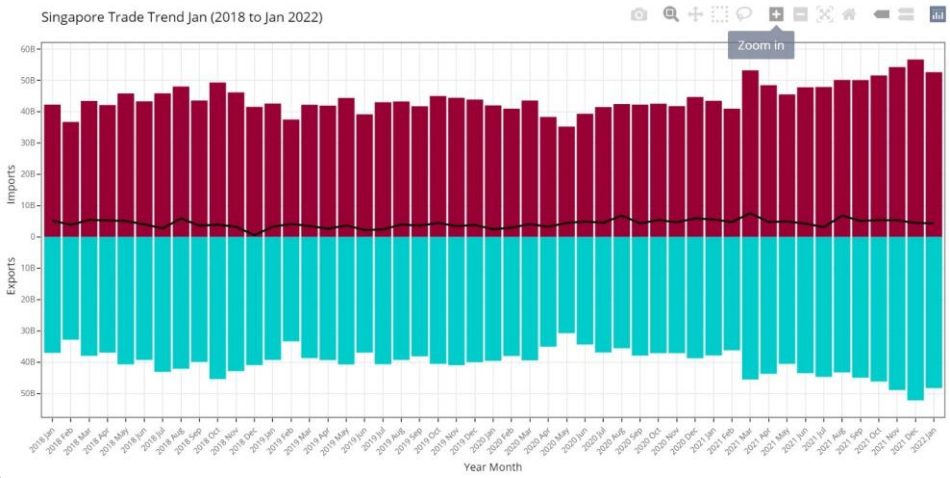
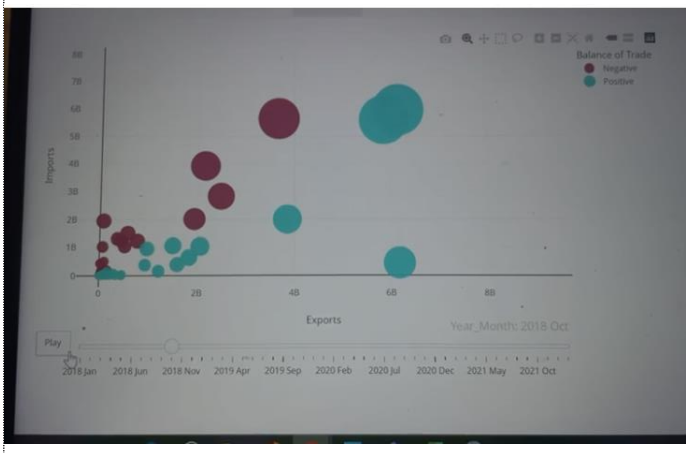
#	Key points discussed and actions	Action Owner	TCD
1	Yanru shared her data transformation using pivot wide.  Yanru updated that the country names in the excel file do not match to the country names in the geospatial libraries, hence more data cleaning is required to present the cluster results into a world map. Will work on this today. (post meeting): Yanru shared the progress after changing country names to fit to the world map. the names of the country will also appear upon mouse hover  Yanru updated will start fitting the constructed graphs into shiny to ensure it is working.	Yanru	2 Apr 2022
	Karishma shared the scatter plot created to show the negative and positive for balance of trade.  	Karishma	2 Apr 2022

#	Key points discussed and actions	Action Owner	TCD
	Karishma recommended using this visualization for the homepage tab http://wits.worldbank.org/visualization/positiveandnegativemerchandisetradebalance.html, and may try to achieve this visualization http://www.brightpointinc.com/united-states-trade-deficit/  Will share progress on whatsapp		
2	Farah has constructed the poster framework and design in prep to put in the contents of the project.  Farah shared her construct of the graph and fitted into Shiny app format, using geom_line with selection input such s country, year, type of data.	Farah	31 Mar 2022

#	Key points discussed and actions	Action Owner	TCD
	 Another dumbbell point to point chart for seeing the changes but is not in app form  Will send the code to Karishma and Yanru for merging into the r shiny app. Farah mentioned that due to time constraints, she will not be exploring any more graphs and start of focus on the poster.		
	Karishma will help to consolidate Farah's codes into the R Shiny. Yanru reminded Karishma to copy all the functions at the top before calling the R shiny app	Karishma	31 Mar 22

Date and Time: 31 March 2022 3.30pm - 4pm

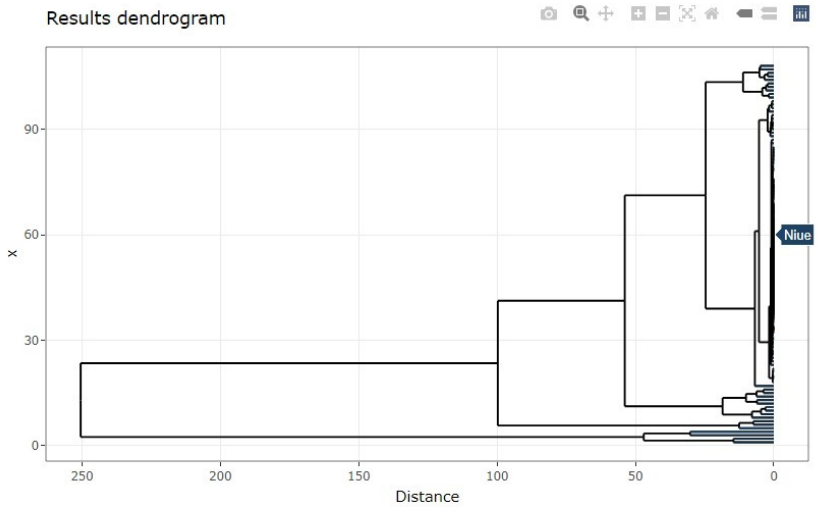
Attendees: all

#	Key points discussed and actions	Action Owner	TCD
1	Karishma has incorporated the time series clustering into the app  Yanru mentioned that she wil explore if tool tips can be added to the chart since the graph looks a bit indistinguishable now, but priority should be to fit in the world map first then only beautify the plots.	Yanru	2 Apr 2022
	Karishma also shared an overall trend created for use on the home page and an animated bubble plot for import and exports. Karishma will continue to build more graphs for user interaction.   		

#	Key points discussed and actions	Action Owner	TCD

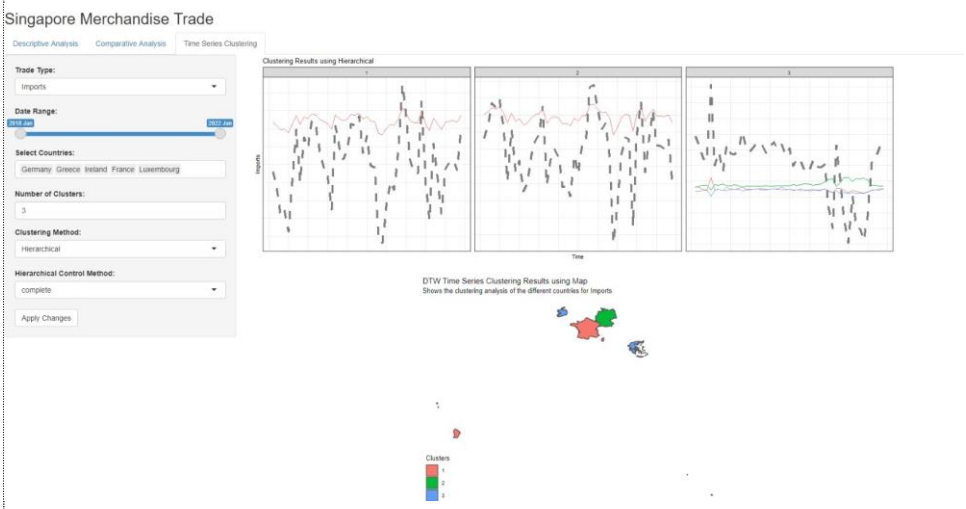
Yanru share the dendrogram with tooltips.

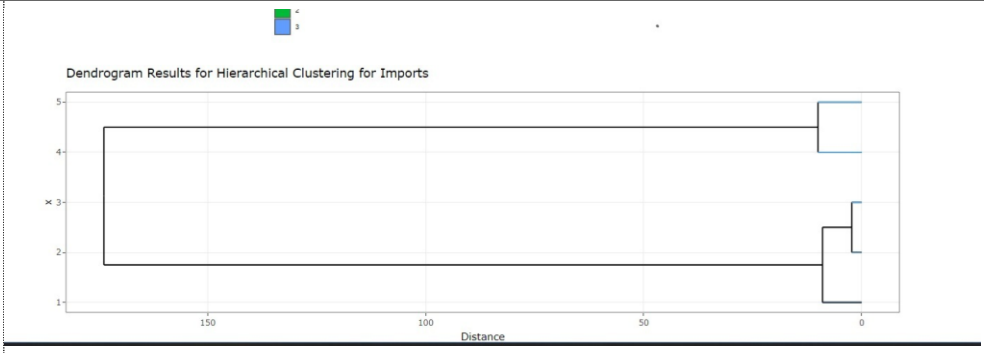
- The base is super shot and labels cannot be added as it will be cramped together especially for imports.
- The clustering is also not good as there are 6 clusters and majority only in 1 cluster.



Karishma recommended to see if the country selection option can be incorporated, however Yanru reminded that a minimum number of countries is required for clustering, hence the country filter was not included.

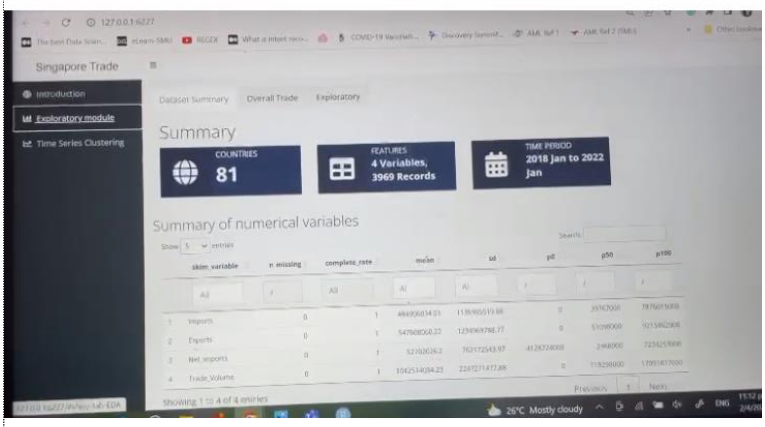
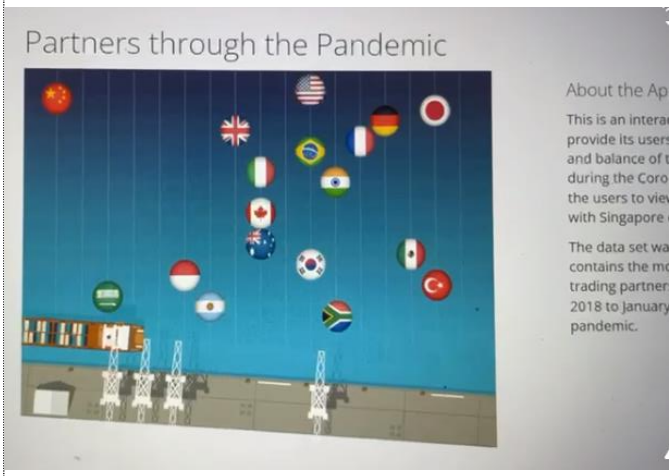
Yanru shared below how the graphs look like if only 5 countries were selected. Yanru will add a function to disable the “apply changes” button if number of countries selected is less than 10.

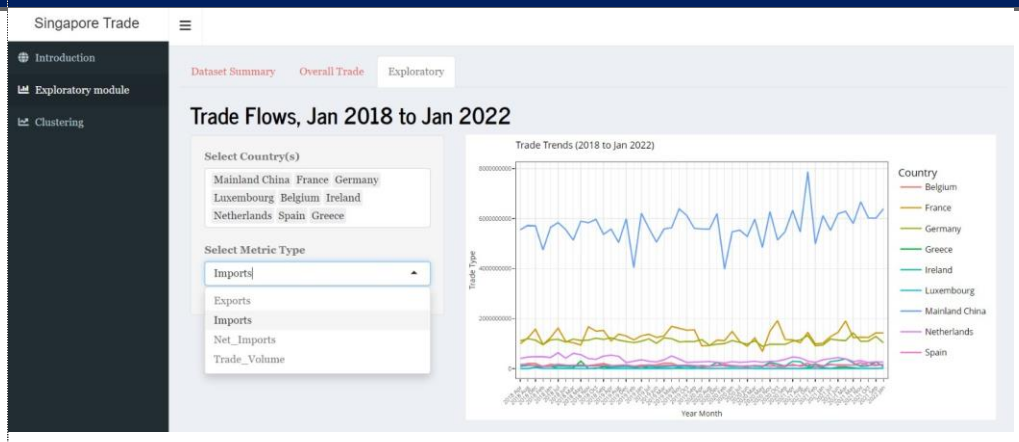


#	Key points discussed and actions	Action Owner	TCD
			
	Karishma requested help to make the clustering page complete. Yanru will update with changes and send to Karishma to replace.	Yanru	1 Apr 22
	Karishma updated that she will make all the plots so we can use it for the poster. Karishma will also work on the website.	Karishma	2 Apr 22
	Yanru will start on the user manual for her part. Karishma recommended to use this link as a guide https://ourshinypet.netlify.app/files/ShinyPET_userguide.pdf/	Yanru	2 Apr 22

Date and Time: 1 Apr 2022 9-10pm

Attendees: all

#	Key points discussed and actions	Action Owner	TCD
1	Karishma shared a video of how the app looks like now and a few options of the gif. All prefer the 3rd picture.    		
	Karishma has added Farah's part but modified		

#	Key points discussed and actions	Action Owner	TCD
			
	Farah has reminded all to keep the pictures coming for the poster		
	Farah suggested to name the shinyapp. Suggestions were made during the meeting: Singtrade, Trade genie, SIMEX (Singapore Import Export), Lion Trade, Trade window, Trade spyglass All agreed on SingTrade Compass.		
	To regroup tomorrow to check on the final app.		

Date and Time: 2 Apr 2022 8pm - 12pm

Attendees: all

#	Key points discussed and actions	Action Owner	TCD
1	Karishma has uploaded the app onto the shinyapps page and requested us to check https://1109karishma.shinyapps.io/singtradecompass/		
	Karishma also created the netlify page and input all the links required while waiting for the final poster and user manual. - Karishma and Yanru are working on the user manual. - Yanru reminded that we still need to write an abstract for the main page of 350 words. - Farah is updating the poster with latest app on the shiny page.		
	During the meeting, photo was taken individually to be sent to prof. All members have reviewed the app, netlify, poster and minutes and agree to be submitted as final		