VISCOMPAT: A COMPREHENSIVE VISUALISATION ON COMMUTER PATTERNS FOR WORKING ADULTS AND STUDENTS ACROSS REGIONS IN SINGAPORE REGION USING SBS BUS DATA

Jouta Lim, Joachim Fu, James Chua

Abstract— The adoption of smart card technologies tracking systems for transportation analysis allows public transport planners to make use of the big data captured daily. This data displays the travelling patterns of commuters by the different modes of transportation. Bus transportation has the most expansive network that travels to almost all parts of Singapore. This research therefore allows this planners to have a comprehensive overview of the commuter travel patterns. However, due to the vast amount of data provided, this reseach will focus on a specific commuter group which is the working adults and students travelling in the morning peak hours. VISCOMPAT, a visualisation tool, will be used to demonstrate the framework and methodology.

Index Terms—Travel patterns, Regions, Commuting to work/school, residential location, urban transport, Visual Analytics, Dextra.

---

1 INTRODUCTION

Bus transportation in Singapore has been expansive in its network. Being able to traverse to all different regions far and wide with an approximately a million rides recorded a day. Commuters travel by bus to their workplace or school or for other personal reasons. Living in a small country where land is scarce, it gives an advantage for people to travel to different parts of Singapore. The Land Transport Authority provides the records of commuters' smart card and travelling patterns. However, the challenge in extracting huge amounts of data proved to be difficult for consolidation and conducting of analysis. Currently, there are many analytical tools that have been created to analyse such data but is static and lack interaction. As such, the focus will be to explore new techniques to allow transport planners to transform the data into an interactive and dynamic display.

2 MOTIVATION AND OBJECTIVES

Locals travel in the morning to reach work or school on time. Ideally, the distance between home and workplace or school should be short which provides ease and accessibility for commuting. However, not everyone has the luxury by far. People travelling across regions take longer times and may need to set off earlier in the morning. Despite efforts by companies to provide flexible working hours or work from home schemes, this also means that workers must replace commuting time with work. Though most of the current businesses are situated in the Central Business District, there are plans for urban development for businesses in the western part of Singapore. The inclusion of Changi Business Park as seen in the east may also have workers who live in the west. Even travelling to schools may be a hassle due to the need for cross-region travelling.

The main objective of this project is to find out the ratio between cross region and within region commuter patterns with boarding, alighting stops and duration types as the secondary objectives. Users are primarily the urban planners who will be provided an interactive, dynamic overview of how buses traverse to reach the targeted destination. The motivation for this work is to explore the relationship between daily bus commuters and the geographical location of workplace so that urban planners can devise new policies with reference to the commuting patterns.
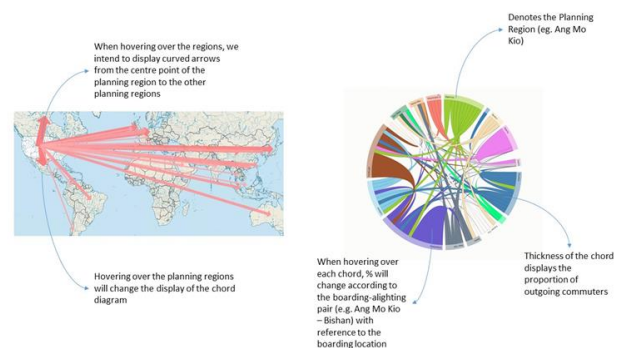
3 DATA TRANSFORMATION

This study was performed on over 5 million records in a period of a week from the 15th to 19th of Feb 2016 provided by the Land Transport Authority. Two datasets were provided. One containing the commuter card data with the boarding and alighting stop station and boarding and alighting times. The other contained the planning area of each bus stop by the Urban Development Authority. The bus commuters chosen consisting of students and working adults are sieved out using the period from 5am to 9am in the morning. The

two types of travel are cross region and within region. The regions are demarcated by the geospatial data (X and Y axis). The URA dataset has three different demarcations: Region, Planning Area and Subzone. The planning area best suits the visualisations as it is not broken down to too many subjects which may lead to poor aesthetics. Furthermore, the Planning Areas were set by the URA to facilitate urban planning, making it best suited for us to address our main objective. The bus stops have its respective names of which after joining the two datasets using JMP Pro, the alighting and boarding planning areas are retrieved. Cross region and within region travels are denoted by the different and same boarding and the alighting planning areas of the commuter respectively. The planning areas are demarcated in the map below.

4 VISCOMPAT AS A BUS TRANSPORT ANALYSIS TOOL

After transforming the dataset, the utilisation of VISCOMPAT as a visualisation tool will help the users to interact with the actionable data. The goal of the visualisation is depicted to be similar to the

illustration above. With the map as the filtering tool for users to click on each planning area, highlighting the planning area on the map. At the same time, the chord diagram will be brushed to show the flow to and from the selected planning region. Alternatively, a dropdown list has also been set up for people to find a specific planning area with ease, which will brush both the map and the chord diagram to focus on the area of interest. The following sections will explain in detail the reasons behind the choice of each diagram and how it is created and linked together.

4.1 Chord Diagram

The chord diagram has compact visualisation and has nice aesthetics. Its interactivity reduces data overload and has easy integration into any web based system. However, with the amount of data given, there are about 45 subjects which means about 45*44/2=990 chords. Nevertheless, there are some functions in the following applications that allows the user to zoom into a chord. The aim of the chord diagram is to illustrate

4.1.1 Power BI

Measure by ALIGHTING_STOP_STN and BOARDING_STOP_STN
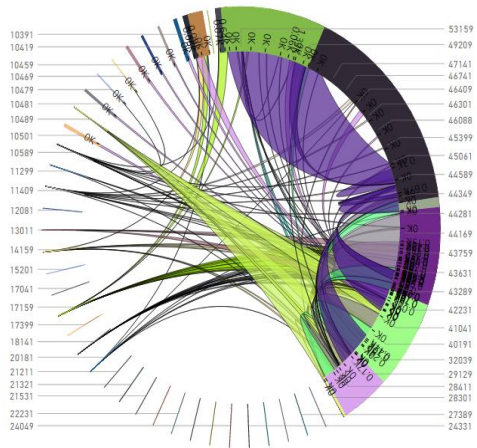
Web design developers have been sharing numerous design templates for non-technical users to implement the visualisation with ease. With Power BI, the chord template as displayed above shows the boarding and alighting bus stop pair with the size of the chord to denote the proportion of outgoing commuters with respect to the boarding stop. However, the limitations of using Power BI are (1) there is only one template (2) the lack of customisation of colours and the size of the surrounding arc which demarcates the respective planning areas (3) not all the bus stops are shown which makes the visualisation incomplete and not comprehensive. It will be difficult for users to select the arcs especially those that are rather thin which constitutes to poor aesthetics which leads to lack of interactivity.

4.1.2 D3.js
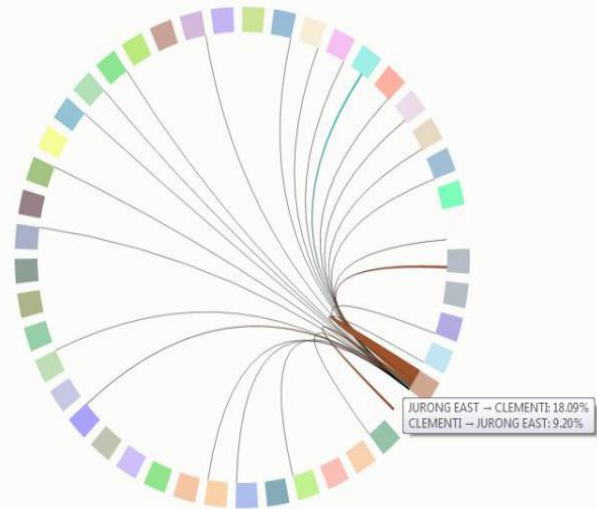
In contrast to the previous case, D3.js allows for greater degree of customisability and control to achieve what is needed to be presented to the users. As shown in the diagram above, the chord diagram serves as an appropriate visualisation chart for the dataset. Unlike Power BI, the proportion of the surrounding arcs are equally divided so that users can still hover over a significant region for viewing. In addition, there is a tooltip that is displayed as shown in the diagram above for users to gain detailed information regarding the percentage flow of commuters from the boarding planning area (Jurong East) to the alighting planning area (Clementi) and vice versa.

The chord diagram was referenced to Andrew's Chord example online. With reference to the colour the random colour generator online, each planning area is attached to a colour. The colour scheme table was manually created as shown below.

```
name,color
ANG MO KIO,#9ACD32
BEDOK,#377DB8
BISHAN,#F5DEB3
BOON LAY,#EE82EE
BUKIT BATOK,#40E0D0
BUKIT MERAH,#FF6347
BUKIT PANJANG,#D8BFD8
BUKIT TIMAH,#D2B48C
CENTRAL WATER CATCHMENT,#4682B4
```

When building the chord diagram, the boarding-alighting pairs data table as shown below was manually created. A consolidation of the boarding planning areas on the y axis and the alighting planning areas on the x axis was manually done on excel and saved in a .csv format as shown below.

A	ANG MO KIO	BEDOK	BISHAN	BOON LAY	BUKIT BAT	BUKIT MEI	BUKIT PAN	BUKIT TIM	CENTRAL	CHANGI	CHOA CHU	CLEMENTI	DOWNTO'	GEYLANG	HOUGANG	JURONG E	JURONG V
ANG MO KIO	165910	4251	26495	0	122	727	0	1755	1325	242	0	333	0	3635	13025	0	0
BEDOK	1533	251469	303	11	825	2620	3	169	0	5930	28	121	717	40217	4476	200	25
BISHAN	10833	1088	61711	0	34	588	0	1548	467	93	0	414	0	753	1725	82	0
BOON LAY	0	0	0	326	0	0	0	176	0	0	0	0	0	0	0	11	1862
BUKIT BATOK	153	488	356	215	92322	1917	2341	10731	479	0	4127	11276	0	290	0	11738	5847
BUKIT MERAH	246	734	256	17	479	113420	43	2087	152	15	80	2507	125	1236	84	1621	187
BUKIT PANJANG	14	1063	13	41	8159	2214	36371	7045	13	0	6467	2823	0	744	0	2641	1665
BUKIT TIMAH	338	63	1109	38	3634	1120	356	27053	894	0	286	4779	0	139	234	767	661
CENTRAL WATER CATCHMENT	427	16	893	0	46	192	193	500	190	0	64	21	0	5	57	17	18
CHANGI	62	2803	25	0	0	16	0	0	0	7250	0	0	0	225	214	0	0
CHOA CHU KANG	0	51	0	0	4360	417	6943	1803	1168	0	120297	688	0	49	0	1213	3366
CLEMENTI	181	88	397	561	4212	5314	552	11367	289	0	188	87734	5	168	119	15015	2165
DOWNTOWN CORE	0	6	0	0	0	17	0	0	0	0	0	0	0	24	0	0	0
GEYLANG	1934	20416	253	44	137	3025	172	1293	96	721	71	327	511	89144	7220	75	69
HOUGANG	19397	17382	2947	0	0	176	0	1440	118	2503	0	176	111	13418	117785	1	0
JURONG EAST	0	208	161	1500	6270	4172	144	1971	111	0	1572	27598	0	153	3	81038	16349
JURONG WEST	0	14	86	19412	4415	679	201	2225	100	0	3445	5913	0	66	0	38771	270106
KALLANG	1324	2246	598	10	179	1072	203	980	34	15	113	49	289	12205	408	27	9
LIM CHU KANG	0	0	0	0	0	0	0	10	0	0	0	161	0	0	0	0	0
MANDAI	162	0	0	0	0	0	0	0	130	0	0	0	0	0	0	0	0
MARINE PARADE	199	5348	21	4	0	1179	81	14	0	484	0	51	902	11647	382	29	5

Each row on the data table was then summed up and each boarding-alighting planning area pairs were given a percentage with respect to that sum. After consolidating each percentage, a matrix table is

created and saved in JSON format as shown below.

```
[
[0.609224103,0.015609738,0.097290053,0,0.000447986,0.002669555,0,0.006444387,0.004865421,0.00
.000466346,0.009033158,0.001652407,0.001634047,0,0,0.000374546,0.001053868,0,0.00140271,0.00092
3832,0.015646458],
[0.003950145,0.647970687,0.000780753,2.83442E-05,0.002125812,0.005205018,7.73023E-06,0.000435
44185E-05,0.010299237,0,0,0.028918773,0,0.001857831,0.00742617,0.015061056,0.004656172,0,0,0.
8708,0.000698297,0,0.01359489,0,0,0.002316491,0.0002782288],
[0.110957473,0.011143887,0.632077598,0,0.000348246,0.006022616,0,0.015855457,0.004783268,0.00
0851E-05,0.00060431,0.004137988,0.0419637,0.001782203,0.000338004,0,5.12127E-05,0.001526139,0.
065193789,0,0,0.000194608,0.010846853],
[0,0,0,0.120875046,0,0,0,0,0.004078606,0.690396737,0,0,0,0,0,0,0,0.16981831
[0.000997054,0.003180148,0.002319944,0.00140109,0.601634387,0.012492506,0.014603915,0.0699306
2.60668E-05,0.000436619,0,0,6.5167E-06,0.003636316,0,0,0,0,0.02757214,1.30334E-05,0.0003844485
1919,0.000566953],
```

Upon reading the JSON file and the other data tables on the script, the expected chord diagram is generated.
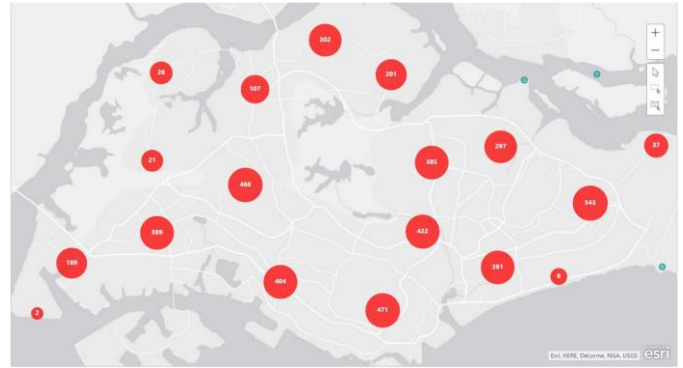
4.2 Geographical Map

A geographical map as a visualization tool is best suited to our data set given its very nature being rooted on geographical locations that are best seen and understood from a map visualization. The map allows users to view and investigate geographical implications on the commuter patterns.

4.1
4.2

4.2.1 Power BI
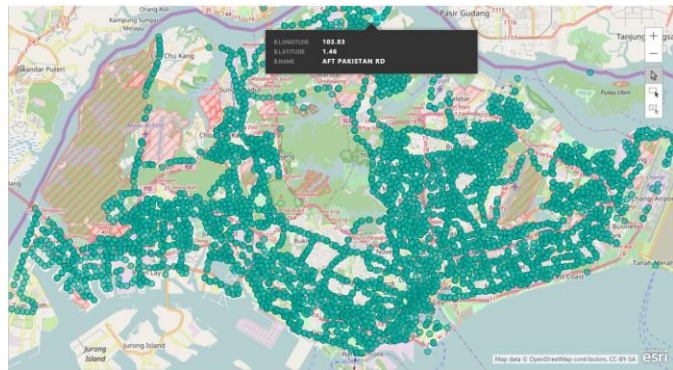
Power BI offers both their inhouse geographical mapping visualization and also an experimental feature offered by ArcGIS. Power BI's own geographical mapping visualization was prone to misrepresenting the coordinates provided by our data, which often led to inaccurate visualizations that were not representative of the dataset. ArcGIS's visualization was able to accurately plot out the latlong values in the dataset, offering further features such as clustering and having different layers of map data available for the base map or for overlaying over our data. However, the visualizations were limited in their capabilities to allow us to geocode the data based on demands, and were restricted by the template format by only allowing certain formats and ways of visualizing the data.


4.2.2 Tableau Desktop

Tableau and its rich database of geographical maps allows for easy plotting of maps with latlong data from the dataset. Its database also contains detailed map subunits such as Singapore's planning areas in the form of map overlays selectable from its settings. However, similar to Power BI, Tableau's geocoding and mapping capabilities were limited with the file formats it was able to read and use, also with the restrictions as to what can be mapped out on the map. Despite its impressive database with map details from official authorities, the inability to use it in computing calculations from the dataset rendered them redundant, to only contribute aesthetically to the visualization.

4.2.3 D3.js
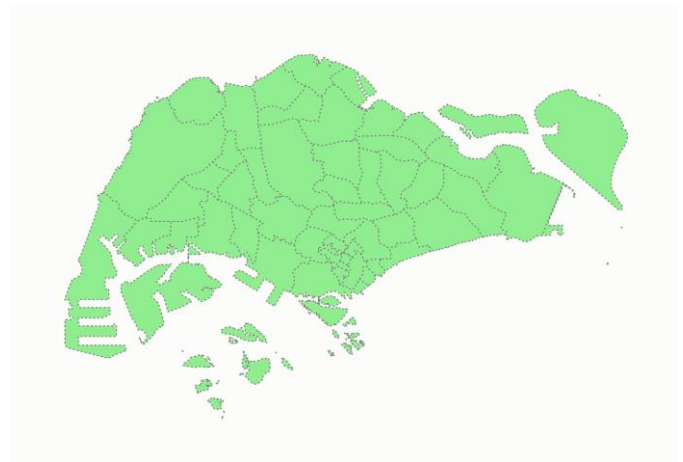
In contrast to the 2 available commercial tools on the market, D3.js provides the flexibility to customize as needed for the visualizations, granting creators the freedom to select and implement exactly as they have wished to. D3.js allows for creation of map

diagrams from the polygon shapefiles available publicly from data.gov.sg, the Singapore's Public Data portal. With the high customizability, the need to plot out all bus stops were no longer needed, with the ability to summarize data by the planning areas to give a clearer and more concise data display to users exploring the data. This provides a better visualization for users without overloading them with data, whilst giving them sufficient details in their investigation.

The map was constructed with reference to Mike Bostock's "Let's Make a Map", a step by step guide using shapefiles with d3.js to make use of its robust graphing libraries to draw out the polygons on web applications that could be easily hosted and accessed by others. Further improvements on the map data were made by following the guide in making conversions from the initial shapefiles to the resulting topojson files used for the application. Topojson files give the highest file size efficiency whilst maintaining the data formats needed, improving the overall efficiency of the application by removing redundant data. D3.js has many libraries catered for different needs, in particular, the TopoJSON library has been utilized in the mapping application. D3.js's ability to read in different types of data files allows for easy integration between different diagrams and present the data in a dashboard format for users to explore.

5   User Interface Design
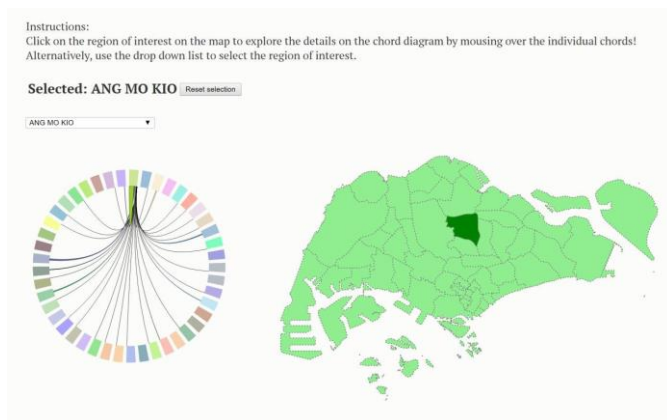
VISCOMPAT serves as a simple yet interactive tool for users to have a comprehensive view on the commuter flow. The UI simply allows the user to click on the map, which serves as a filter for the chord diagram. These two diagrams should fit neatly on a web page and to remain accessible to all users.

Clicking on a planning area on the map highlights the area, to allow users to see which area have they selected. Upon selecting the specific area on the map, the users may move over to the chord diagram and hover over the individual chords for an in-depth view of the commuter flows between the respective planning areas. For users who wish to view a specific area yet are not certain on where does the area lie geographically, they may use the dropdown list provided to select the name of the area they are interested in, which will be reflected on both the map and the chord diagram. In summary, the map acts as an interactive filter allowing users to view the geographical location of the area of interest, whereas the chord diagram provides a detailed view into the data of the area selected. Clicking on Reset selection clears all brushing from the interactive visualizations, showing the chord diagram in its full with the intersections between the various chords. From there, users may hover over the chords of interest to find out the name of region and select it via either the dropdown list or the map.

6   Conclusion

Comparing across all the visualizations tools available, we found that each of them have their strengths and weakness in aiding us in formulating our visualization.

JMP Pro serves as a useful data processing tool able to clean big data files and perform advance data joining functions to allow us to process the data files into formats more efficient and effective for our usage.

Tableau and Power BI being visualization tools with well-defined UI allows us to use them easily for initial data exploration in helping us understand our data files to allow us to formulate our visualization design.

Finally, D3.js gives us the freedom to customize visualizations to provide for an interactive customized visualization suited to both the data and our design needs. Given its strengths in design capability, it was our tool of choice for our visualization tool.

References

[1]   (URL link)
[2]   https://bl.ocks.org/kerryrodden/7090426
[3]   http://www.thegreatcodeadventure.com/d3-sunburst-diagram-tutorial/
[4]   http://www.straitstimes.com/singapore/lta-seeks-more-precise-picture-of-travel-patterns
[5]   http://web.mit.edu/11.521/papers/WorkingSocialDocument_Aug2012_v2.pdf
[6]   http://ac.els-cdn.com/S0968090X14002988/1-s2.0-S0968090X14002988-main.pdf?_tid=026d947e-a996-11e6-8b65-00000aab0f26&acdnat=1479037423_466fe08d5d862c7a7647ccd38213fc00
[7]   https://www.jtlu.org/index.php/jtlu/article/viewFile/740/429
[8]   https://bost.ocks.org/mike/map/
[9]   https://publictransportsg.wordpress.com/2013/10/01/bus-stops/
[10]   http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-1/167/2014/isprsarchives-XL-1-167-2014.pdf
[11]   http://bl.ocks.org/AndrewRP/raw/7468330/
[12]   http://worksingapore.com/articles/live_4.php
[13]   https://www.random.org/colors/hex
[14]   http://bl.ocks.org/curran/55d327542393530662c3
[15]   https://bost.ocks.org/mike/map/
[16]