Tax-ed: A Visual Analytics Dashboard for Exploring and Analysing Tax Revenue in Singapore

Bryan LAU, Taffy Joan CHEOW, Kei Wen TAM

Abstract — With the move towards a open data society, more effort has been placed for public statistical data in Singapore such as tax data to be readily accessible to the public. However, majority of such data is published in a summarized manner or broken down individually in separate data files. While summarized infographics are published as well, the raw data files provided provide limited analytical functions to fulfil the needs of casual data users. To address this gap, we contribute, Tax-ed, a visualization dashboard that aims to allow casual data users to explore and analyse Singapore's tax data with both breadth and depth of all tax data available in Singapore from different government agencies. Through the use of dynamically linking a tree map, bar and line chart, we demonstrate the potential of Tax-ed for users to gain both macro and micro view of the tax revenue structure in Singapore.

Index Terms—Tax revenue, Singapore, tree maps, line charts, bar charts, pie charts, stacked area charts

1 Introduction

As compulsory contributions made to the government by citizens, corporates and firms residing in the nation, taxes form a major part of government funding to sustain the day to day operations that promote social and economic goals that strive to build and develop a nation. In Singapore, taxes alone form 85.5% of total government revenue in 2017[1], with tax revenue amount to S$47 billion, a 5% increase from the total amount collected in the previous fiscal year[2].

In Singapore, tax revenue comprises of various components such as personal income tax, corporate tax and property tax. The composition of tax revenue can be largely swayed by shifting economic structure and lifestyles among the population. Taking in considerations such shifting patterns, tax policies set up by the government such as tax progressive rate and exemptions then serve as a strategic method to sustain its revenue while maintaining competitive and equal tax rates for taxpayers for a stable economy[3].

Given the complex nature of a tax system, analytics has become a key tool for the government to protect their tax revenue by uncovering hidden links in major tax revenue and identifying areas of tax evasion [4]. An analysis and comparison of tax revenue and its revenue across years provides greater insights for casual public data users in studying the economics relating to the shift in tax revenue compositions and the extent of impact resulted from the implementation of tax policies. Furthermore, for taxpayers, an understanding of Singapore's tax revenue structure allows them to understand their role and contributions as taxpayers in a larger context.

While there have been an increasing number of visualizations for tax structure, these visualizations tend to be centred on taxes in other countries. Currently, the Singapore government continues to publish annual data on tax in the form of Comma-delimited files (CSV) and infographics to summarize key findings[5]. Infographics have the tendencies to confine its findings in the analysis of major tax categories such as income tax due to the lack of emphasis on the breakdown of each tax category. Furthermore, not all tax data is published by the Inland Revenue Authority of Singapore as some taxes such as betting tax is not governed by IRAS. However, such taxes are still recognised as tax in the annual government operating revenue.

Upon identifying these gaps, we present an interactive visualization dashboard that aims to aid in the exploration of Singapore tax revenue data in both breadth and depth across 10 years of data, thus enabling users to validate preconceived expectations and uncover insightful findings, trends and patterns. Tax data from the year 2006 to 2016 was selected as data retrieved from these years were reported most consistently.

This paper reports our research and development efforts in building a comprehensive yet simple interactive visualization tool suitable for all kinds of users, ranging from those who do not understand the tax system to public data users involved in the field of taxation. Following a general introduction in Section 1, the paper continues in Section 2 with an elaboration on the motivation and objective of this research. Section 3 provides reviews on past analytical visualizations techniques. Subsequently, the design and selection process of Tax-ed is discussed in Section 4 and 5 respectively. Lastly, the paper concludes with an elaboration on the future direction of the research.

2 Motivation and Objectives

Our research and project stems from the lack of general interactive visualizations that allows users to understand the Singapore tax system. While it is a common practice among government bodies to release individual tax revenue data and infographics annually, there has not been a consolidated effort to pool all forms of tax revenue across the years into a single platform while preserving the depth of the data. The lack of such platforms pose as a challenge for users to discover and explore tax data in a simple yet insightful manner. To be specific, the visualization dashboard aims to provide users the capability to achieve the following:

1) To be able to visualize the composition of tax revenue from both a macro and micro view

2) To provide a graphical representation of the changes in tax revenue across the years

3) To provide details-on-demand for all users

3 Discussion and Analytical Mapping

As there are limited visualizations pertaining to Singapore tax revenue, we expanded the scope of our references to include visualizations about the Singapore budget that encompasses government revenue.

In previous related works, pie charts were commonly used to visualize the part-whole relationship of tax categories as part of the whole tax revenue collected.

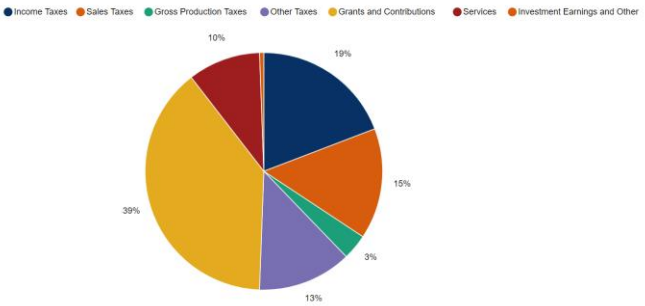

Fig 1: Pie chart of tax revenue[6]

In the following examples above, users were able to extract the percentage of tax collected in each tax category. However, as the number of categories increase, the slices in a pie chart becomes narrower and less visible. As a result, it becomes challenging for users to differentiate the categories. Furthermore, having a circle shape provide less surface for text due to its curved edges. As compared to regular shapes such as rectangle and squares, pie charts pose a challenge in fitting graphs into a dashboard. This is because white spaces in between graphs would result in inefficient use of space on the dashboard.

To overcome this, a stacked area chart can be used. This type of chart has a regular shape for more efficient use of space. With reference to a visualization built for SG Budget 2017[7], a stacked area chart can be used to show the composition of the total tax revenue while depicting the trend of total tax amount collected over time. In the following example, we can observe the composition of tax revenue collected which has generally risen over the years.

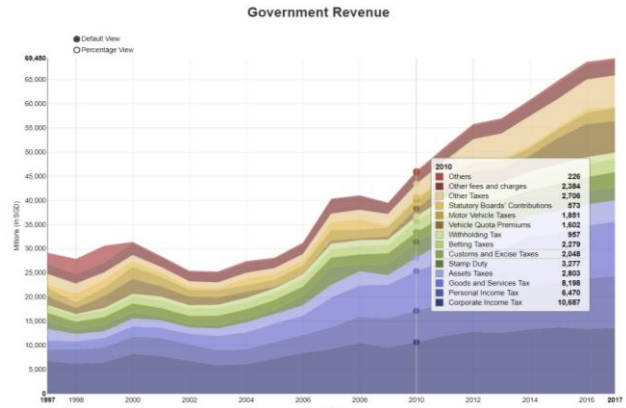

Fig 2: Stacked area chart with drop down lines showing the breakdown of tax area[7]

Despite its dual functions, the stacked area chart can be misleading to users. The stacked nature of the chart can lead to misinterpretations that more tax revenue was collected from the components stacked on top as compared to those stacked below, which might not be the case. The graph can also result in misinterpretation that all components follow an increasing trend due to the undefined sections especially in areas where similar colour tones are used.

Hence to address the gaps identified in both pie chart and stacked area chart, we consider using a tree map. A tree map is able to show the proportion of each category to the total amount. By proportioning the size of the area it covers, a tree map is designed to better represent a value relative to its neighbouring values within a data seta compared to a pie chart [8]. In addition, a tree map is able to hold multi-dimensional data by having a diverging colour scale to represent growth and decline of categories.

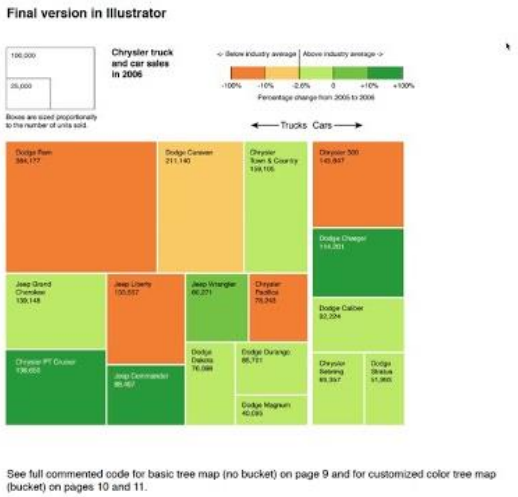

Fig 3: Tree map with diverging colour scale[9]

An interactive visualization dashboard can be created with a number of methods ranging from public software such as Tableau. However, as a software catered to use of the masses, visualizations in Tableau are restricted in terms of styling. Rendering large datasets via Tableau can also result in slower refresh rates when enabling filter or click functions.

4 User Interface Design

Referencing to the taxonomy of data as suggested by Shneiderman, B., the Tax-ed dashboard is built around the approach that allows users to have an overview first, before zooming in the data and filtering accordingly thus getting details on demand [10]. Based on the advantages discussed in the previous section, it was decided that the 3 main visual components of the dashboard would be a tree map, line and bar chart.

When users first access the dashboard, they will first have overview of the total tax revenue composition based on tax categories in the tree map. The tree map is also accompanied with a diverging colour scale to represent year-on-year growth (blue) or decline (red) (Fig. 4).

Fig. 4: Tree map

The dashboard allows for two types of analysis - single year and cumulative years. A single year analysis breaks down the tax revenue for the specific year selected (Fig. 5). On the other hand, a cumulative analysis provides flexibility to the application by enabling users to study tax composition taking into account multiple years and compare the amount collected based on the start and end year specified, whereby the start year serves as the base year (Fig. 6). The change in years selected will trigger a change in all visuals dynamically.

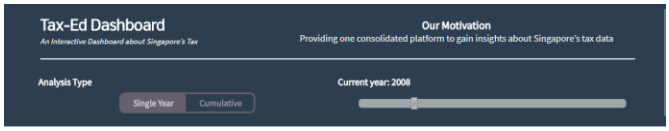

Fig. 5: Dashboard header with indications

Fig. 6: Cumulative slider

Clicking on each tax category of the tree map will generate a bar chart on the right side of the dashboard for users to zoom into a micro view of the different tax types in each category (Fig. 7, 8 & 9).

Fig. 7: View of bar and line chart when zooming in from tree map in a **single** analysis

Fig. 8: View of bar and line chart when zooming in from tree map in a **cumulative** analysis

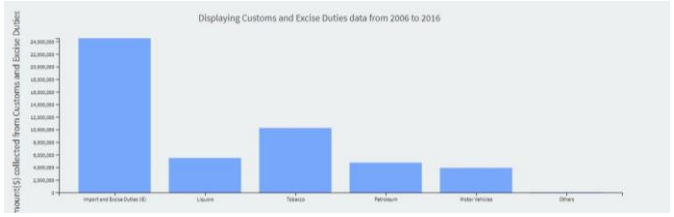

Fig. 9: Bar chart

Lastly, the line chart provides a visual component for a time-series trend analysis of the percentage change, both positive and negative, observed in different tax types (Fig. 10). Greater emphasis can be brought to each line through highlighting when lines are hovered across. Furthermore, users are able to filter tax types accordingly by clicking on the legend text for a customizable visualization.

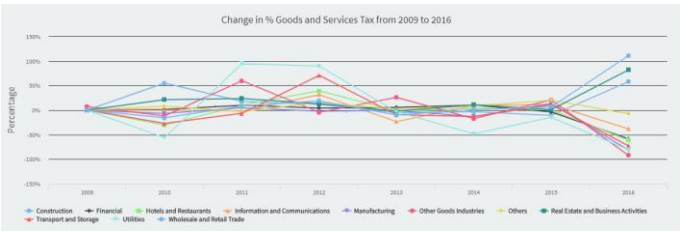

Fig. 10: Line Chart

5 DATA CLEANING

Prior to the building of the visualization, it was essential to pool data from the identified sources and structure the data for analysis. One of the challenges faced was the format in which data extracted from IRAS was presented (Fig. 11).

Assessed Income Group	Number of Taxpayers		Assessable Income[1]		Net Tax Assessed[2]	
S$	Tax Resident	Non-Tax Resident[3]	Tax Resident	Non-Tax Resident[3]	Tax Resident	Non-Tax Resident[3]
			S$'000	S$'000	S$'000	S$'000
20,000 & below	-	18,708	-	99,583	-	15,815
20,001 - 25,000	23,642	718	547,350	16,093	1,647	2,696
25,001 - 30,000	58,949	577	1,644,865	15,793	7,253	2,649
30,001 - 40,000	162,886	772	5,660,830	26,727	40,228	4,490

Fig. 11: Data extracted from Taxable Individuals by Assessable Income Group, Year 2004

Furthermore, such structure was not aligned with data extracted from Datagov.sg or CEIC Database (Fig.12 & 13).

Sector	Revenue	Year
Manufacturing	1717000	2006
Manufacturing	1582000	2007
Manufacturing	1593000	2008
Manufacturing	1554000	2009
Manufacturing	1452000	2010
Manufacturing	1524000	2011
Manufacturing	1495000	2012
Manufacturing	1492000	2013

Fig. 12: Data extracted from Goods and Services Tax by Economic Sector from CEIC Database

financial_year	actual_revised_estimated	class	type	amount ($ '000)	percent_of_gdp
1997	Actual	Tax	Withholding Tax	337000	0.002
1998	Actual	Tax	Withholding Tax	316000	0.002
1999	Actual	Tax	Withholding Tax	467000	0.003
2000	Actual	Tax	Withholding Tax	486000	0.003
2001	Actual	Tax	Withholding Tax	815000	0.005
2002	Actual	Tax	Withholding Tax	626000	0.004

Fig. 13: Data extracted from Government Total Expenditure, Datagov.sg

As such to align the structure of the data for analysis, data was pooled together, standardized and structured into a hierarchical form to the form 4 fields - taxCategory, taxType, taxCollected, year and rate (Fig. 14).

taxCategory	taxType	taxCollected	year	rate

Fig.14: Fields for reconstructed dataset

Tax category represented the parent category such as corporate income tax. In cases where the amount collected in each category and year could be broken down further, the revenue was divided based on the tax types within each category. Lastly, a calculated field, rate, was added to calculate the percentage change from the previous year. As

the data set started with a base year of 2006, data in this year had a rate of 0.

In the development of the dashboard, data presented onto the tree map was too skewed hence resulting in poor visibility of smaller values. Thus, when designing our tree map, a square root was applied to all values in the data set for better visibility of the data while preserving the integrity of trends and patterns observed in the data.

6 Application

6.1 System Architecture

Tax-ed was designed to be flexible and dynamic for in-depth analysis on tax data across the years. Thus, it was developed based on an incredibly expressive and powerful language, JavaScript. JavaScript is not only capable in producing very reactive visualisation, but also in creating animations - all without re-rendering the webpage to provide the users with a seamless experience.

Tax-ed has also been made web based to enable wider access to end users. It also supports responsive design which enables the visualisation to be viewed on different devices.

6.2 Data Visualization

This visual dashboard was mainly created using Data Driven Document (D3.js), a JavaScript library written by Mike Bostock[11]. D3.js manipulates documents based on data using SVG (Scalable Vector Graphics), HTML (Hypertext Markup Language) and CSS (Cascading Style Sheets). This specialized library offers both reliability and flexibility in building a visual analytics application. Not only does it allow reusable code that is compatible to modern browsers, its capabilities supports large data sets, dynamic animation efficiently thus avoid latency in the application[12].

HighCharts was also used in building this visualization dashboard. Developed by the Norway-based company, Highsoft[13], HighCharts was released in 2009 and is also a charting library developed purely in JavaScript. The main advantage provided by HighCharts over D3's line graphing capabilities lies within its ability to automatically generate the path for every line based on the arrays themselves. In contrast, we would have to write multiple functions in D3 to specify how each line were to be drawn before being able to be visualize it on screen. Hence, the ease-of-use over D3's line graphing capabilities led to a choice of utilizing HighCharts over D3.js for the line graphs.

Since the dashboard aims to process large datasets and support interactive and animated features, the use of D3 JavaScript as the main language for the overall dashboard was a practical decision for a seamless and dynamic user interface experience. Furthermore, with the use of D3 JavaScript, a reusable code structure in the backend would allow for value adding features to be added on in the future.

7 Insights

7.1 Top contributors to tax revenue

A general observation of the cumulative tree map from year 2006 to 2016 reveals that corporate income, personal income and goods and services tax are the top 3 largest contributors in generating tax revenue.

7.2 Corporates taxes play a major role in tax revenue

We observed that corporate taxes form the largest portion in tax revenue. This is mainly driven from the large contributions of tax revenue from companies in the income band of $5,000,001 thousand and above. More than 50% of corporate tax revenue collected are contributed from companies in that income band.

7.3 Increasing growth rate of revenue collected from Motor Vehicle taxes

An analysis of the changing pattern in motor vehicle tax collected reveals that there has been an increasing rate of growth. Contrary to the recent government efforts to reduce cars on the roads in Singapore, an increasing growth rate may imply the ineffectiveness of tax rate on motor vehicle consumption.

7.4 Decline in tax revenue collected in 2013

In the year 2013, we observe a major decline in revenue collected as compared to other years. 7 out of 12 tax categories showed a negative year-on-year growth. A closer look into these categories reveals reductions in tax revenue collected to be as much as 20% in various tax types.

8 Conclusion

As demonstrated, the potential of Tax-ed interactive dashboard underlies its ability to aid users to explore and analyse tax revenue data in Singapore. In the future, the dashboard has the potential to be extended for other forms of government revenue such as fees and charges. In addition, it can be enhanced to include tax expenses as well thus providing a comprehensive exploratory visual tool for in the study of Singapore's national budget.

9 Acknowledgments

The authors wish to thank Professor KAM Tin Seong for his guidance and advice throughout this research.

10 References

[1] Ministry of Finance (March 24, 2016). Government Budget and Fiscal Position, Annual, *Datagov.sg*. Retrieved from http://www.singaporebudget.gov.sg/budget_2016/Home.aspx

[2] Tan. A (Aug 24, 2017). IRAS collected S$47b in tax revenue in FY2017; up nearly 5% from a year ago, *The Business Times.* Retrieved from http://www.businesstimes.com.sg/government-economy/iras-collected-s47b-in-tax-revenue-in-fy2017-up-nearly-5-from-a-year-ago

[3] Mehrara, M., & Farahani, Y. G. (2016). The study of the effects of tax evasion and tax revenues on economic stabilities in OECD countries. *World Scientific News*, *33*, 43.

[4] SAS (n.d.). Increase compliance and tax revenue with advanced analytics. Retrieved from https://www.sas.com/content/dam/SAS/en_us/doc/solutionbrief/government-increase-compliance-tax-revenue-108048.pdf

[5] Inland Revenue Authority of Singapore (Aug 24, 2017). Tax Statistics. Retrieved from https://www.iras.gov.sg/IRASHome/Publications/Statistics-and-Papers/Tax-Statistics/

[6] Data.ok.gov. Primary Government Sources of Revenue Pie Chart. Available at https://data.ok.gov/visualization/ve_chart/1001/. Accessed 23 Nov 2017.

[7] Viz.sg. A data-centric look: Singapore Budget 2017. Available at http://viz.sg/viz/budget2017/. Accessed 25 Nov 2017.

[8] Chittenden, T. (14 Feb, 2016). I Like Pie, But I Hate Pie Charts. *BlumShapiro Consulting.* Retrieved from http://blog.blumshapiro.com/blog/2016/02/14/i-like-pie-but-i-hate-pie-charts/

[9] Treemap. Available at https://i.ytimg.com/vi/Lfv-hOXFgOk/hqdefault.jpg. Accessed 25 Nov 2017.

[10] Shneiderman, B. (2005) “The eyes have it: A task by data type taxonomy for information visualization” IEEE Conference on Visual Languages (VL96), pp. 336-343.

[11] Bostock, M. (2012) Data-Driven Document. Available at d3js.org. Accessed 25 Nov 2017.

[12] Bostock, M. Ogievetsky, V. and Heer, J. (2011) “D3:Data-Driven Documents” IEEE Trans Vis and Comp Graphics, 17(12), pp. 2301- 2309 Dec 2011.

[13] Highcharts. (n.d.). In *Wikipedia*. Retrieved November 25, 2017, from https://en.wikipedia.org/wiki/Highcharts