

An R Shiny Analytics Tool Incorporating Current and Novel Approaches to Visualizing Library Survey Data

Joshua LAM Jie Feng, Karthik NITYANAND, Shreyansh SHIVAM

Abstract — Libraries have obtained demographic, behaviour and user experience data from users through service quality surveys and with increasing reliance on such survey data for evaluation of library performance, we propose the use of an interactive visual analytics tool to allow library staff and managers to explore and analyze library survey demographics and Likert scale ratings data to get richer and deeper insights through various data visualisation and modelling techniques. With the user journey designed for sequential build up of knowledge from various visualisations and models, we hope to improve knowledge discovery by library managers and analysts. We use Singapore Management University's February 2018 Library Survey data as a use case for our application to highlight the variety and depth of insights which can be derived, and propose ways our project may be extended upon.

Index Terms— Library service quality survey, Likert scale data visualisations

1 INTRODUCTION

Traditionally, libraries have used surveys to obtain information regarding their users, in terms of their demographic information, behaviour and experiences in consuming library services [1], and is increasing being used as part of evaluation of the performance of libraries in terms of service delivery [2][3]. Internationally, the Association of Research Libraries have created the LIBQUAL+ survey tool which measures library users' perceptions of service quality in three dimensions: 'Affect of Service', 'Information Control', and 'Library as Place' [4].

In Singapore Management University, which has two physical libraries (the Li-Ka Shing Library and Kwa Geok Choo Law Library), the LIBQUAL+ survey tool was used for its 2015 library service quality survey. Subsequently, implementation of the survey was by Insync™ Surveys [5] in February 2018 and February 2020. Based on the SMU February 2018 Library Survey results [6], the information being collected are of three components: (i) demographic information (consisting of the role of respondents in the university, major area of study/research, the frequency of use of campus, library and online library resources by survey respondents), (ii) the physical library which the respondent visited more frequently, and (iii) the largest proportion of data collected related to user rating of importance and performance of the library in a Likert Scale from 1 (low) to 7 (high) on 26 items in the Library Survey. These 26 items were based on four dimensions: 'Communication', 'Service Delivery', 'Facilities and Equipment', and 'Information Resources'. Respondents were also asked, on a scale of 1 (least likely) to 10 (most likely), how likely they will recommend the library service to other students.

- Dr. Joshua LAM is a postgraduate student under the Master of IT in Business programme (Analytics track) at School of Information Systems, Singapore Management University, Singapore (e-mail: joshua.lam.2018@mitb.smu.edu.sg)
- Mr. Karthik NITYANAND is a postgraduate student under the Master of IT in Business programme (Analytics track) at School of Information Systems, Singapore Management University, Singapore (e-mail: karthikn.2019@mitb.smu.edu.sg)
- Mr. Shreyansh SHIVAM is a postgraduate student under the Master of IT in Business programme (Analytics track) at School of Information Systems, Singapore Management University, Singapore (e-mail: shreyanshs.2019@mitb.smu.edu.sg)

The 2018 and 2020 SMU Library Survey was implemented through an internet portal, and also allowed for free text comments to be provided by respondents. The results from the 2018 library survey were published in a 155-page document [7], with breakdown of survey information and findings in data visualisation such statistical tables, bar graphs, line charts, and gap grids. Aggregate information and information broken down based on major area of study/research were documented on individual pages of the report.

This paper reports our research and development efforts in creating a free and open-source software web application with an interactive user interface, to allow visual analysis of survey data comprised of mainly Likert scale responses. This paper consists of the following sections: we will start with a literature review of current data visualisation practices with library survey data We then discuss the motivations of our project and the objectives to be achieved, describe the principles for user interface design, explain the application and the data visualisation and modelling tools to be used, and finally through use of February 2018 SMU Library Survey data as case study, show instances of meaning derived through the use of various graphs and models. We will also possible extensions of the application in the future.

2 LITERATURE REVIEW

For survey data, it is imperative for display of information regarding the demographic profile of survey respondents. [8] Beyond that, given that our survey data has focus on Likert scale data with importance and performance ratings, relevant data visualisation techniques would be the divergent stacked bar chart, Lollipop chart or dot plots with lines (otherwise known as Dumbbell plots) for sentiment analyses and comparison of importance vs satisfaction scores.[8][9].

3 MOTIVATION AND OBJECTIVES

We have embarked on this research as we believe we can tap deeper into the rich raw data available from the SMU 2018 Library Survey compared to current survey results. This can be done incorporating techniques not previous used in traditional survey data to derive meaningful information for action, and this can be done through use of numerical inputs from respondents to form data visualisation as described in the literature review above, instead of statistical means and gap grids which have been captured in current survey results.

Furthermore, statistical means in aggregated format may not accurately capture the sentiment meaning from ordinal Likert scales.

While the published February 2018 Library Survey report calculates the gap between performance scores and importance scores for different items, each individual sub-category of respondents taking up a single page, leadings to iterations of similar information across different pages which make it difficult to make comparisons. Using an interactive computer interface could allow for clearly visualisation, make use of human computing interaction and increase user engagement with the survey results. The following are the specific analytical outcomes we hope to achieve using the application, with SMU's February 2018 Library Survey as a case study:

- 1) To be able to display descriptive visualisation of the demographic profile of library Survey respondents,
- 2) To create visualisations of Likert scale results through divergent stacked bar charts, to show sentiments of respondents,
- 3) To be able to compare importance and performance score for each survey questionnaire line items to review the performance of the library in different service quality items to guide future strategy,
- 4) To generate new insights from data using novel visualisation techniques with appropriate data transformation,
- 5) To incorporate statistic modelling techniques to support data visualisation,
- 6) To allow for details to be displayed on-demand through the use of interactivity such as tooltips and user selections,
- 7) To optimise human-computing interaction as users engage the application user to view data transformation and facilitate new information discovery.

4 USER INTERFACE DESIGN: STORYBOARD & DESIGN PRINCIPLES

4.1 User Journey - Storyboard

The design of the application is guided by the user journey we intend to bring to the user, or the sequence of acquisition of new information from the library survey results. For the implementation of the applications with SMU Library Survey data, the users will include SMU librarians, management and casual users.

The storyboard aligns with the guiding principle of Schneiderman's Visual Information Seeking Mantra: 'Overview first, zoom and filter, then details on demand' [10].

For the overview, the user will need to begin the understanding of the survey data through knowing an overview and understanding the make-up of the survey respondents. This is accomplished through overview visualisation of hierarchical demographic data of the respondents through bar graphs, donut charts and Sunburst diagrams. This is followed by the accurate representation of Likert scale survey data using divergent stacked bar charts, Dumbbell plots and heat maps, with filters to allow zoom in on specific segments of the data. We then proceed to explore the use of data transformation, visualisation and statistical modelling techniques including novel use of UpSet plots and principal components analysis (PCA). These are to allow users to derive deeper insights from beyond existing survey results.

Altogether, the application user will be able to get a clear understanding of the demographic of survey respondents, their impression of the importance and performance of the library based on different survey item categories, supplemented by UpSet plot, heat map, PCA and ANOVA analyses.

4.2 User Interface - Design Principles

Design of the user interface is to intentionally bring the user through sequential panels of different visualisations to gradually enlighten the application user on the data, of which it is accomplished mainly through separate visual graphics. Information exploration should be fun for the user [10]. The landing panel would be an introduction and overview of the survey data, then brought through A single hue colour with different degrees of saturation is used in the user interface for consistency. In this prototype application, blue and yellow was selected due to its association with Singapore Management University official school colours of deep blue (Websafe Hexadecimal #151C55) and gold (Websafe Hexadecimal #8A704C) to build affinity with members within the SMU community and minimises distractions to the user with pre-attentive processing. Graphics will have non-data ink minimised, and if colours are used, they aid in the differentiation across different groups (e.g. Sunburst diagram).

We employ principles from the Taxonomy of Interactive Dynamics described by Heer and Schneiderman [11] for the visualisations. Panel selection through the user journey will be using a left-hand navigation panel with a top to bottom sequence of information discovery. At the data and view specification, appropriate visual geometries include doughnut, bars, box plots and points largely in representing demographic and aggregate data. Filtering of data is achieved through radio buttons and dropdown boxes provided to the user on the left of panel graphics. This is to allow user engagement and 'zoom in' with data transformation for new information discovery on sub-sets of data. The selections used for different panels are mainly on the library which the respondent predominantly visits, the major area of study/research (e.g. law or economics), and the respondents' role in the school (e.g. student or faculty).

Sorting will be present within data to expose patterns in the data, and values and models derived from the source data [11]. Derivation of new information will be through calculation of mean scores for Dumbbell plots, or calculation of proportions of respondents from existing raw survey data to be represented in UpSet plots, as well gaining insights through reducing dimensionality of datasets through PCA. Interactivity is through tooltips for information display only on demand to prevent clutter on the user interface., while as well as support user knowledge discovery.

5 APPLICATION – SYSTEM AND DATA VISUALIZATION

5.1 System

To tap on the advent of web-enabled visual analytics applications, our research uses the R Shiny application framework to create an interactive Shiny application which will be hosted on the shinyapps.io server. Design principles discussed above guide user interface design captured within the Shiny user interface script (ui.R) and server script (server.R). The user interface script allows parameter refinements to be captured and processed by the server script for data processing and transformation with algorithms to allow for new model building and visualisation to facilitate human computing interaction.

'Guide boxes' are placed at the right hand side to provided essential information to the user to facilitate seamless use of the application.

5.2 Data Visualization Components

Data visualizations are created to support the visual analytics process by combining the visual analytics methods with tight coupling with human interaction for users to gain new knowledge [12]. Figure 1 shows a diagrammatic representative of the visual analytics process:

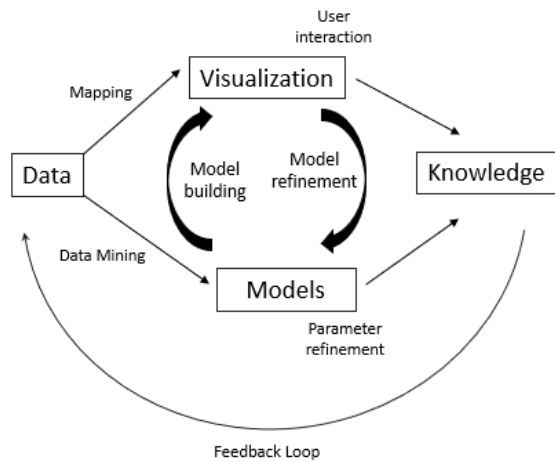


Fig. 1. Diagrammatic representation of visual analytics process. Adapted from VisMaster [12].

5.2.1 Visual Analytics

The following sections describe the visualisation components in the application. For various panels which show data visualisation, there is a left-hand side bar where filter options are available, updated when users move across various tabs for visualisation. The following are filter options for users:

- 1) Library which respondent predominantly visits,
- 2) Major area of study/research, and
- 3) Respondent role.

The following are the visualisations created for the project (described using the grammar of graphics [13]):

- 1) Demographic data have been portrayed using donut charts, Sunburst diagrams and grouped bar charts to show the demographic break down of survey respondents,
- 2) Grouped bar charts show the breakdown of number of survey respondents according to the frequency of: (i) visit to the library, (ii) visit to the campus, and (iii) access to online library resources, denoted by colour aesthetic,
- 3) Divergent stacked bar charts, which are well described for Likert scale data visualisation, is used with bar geometries and red-blue dual colour palette with gradation in saturation, to show the proportion of respondents' ratings from strongly disagree (dark red) to strongly agree (dark blue) across 26 items,
- 4) Dumbbell plots which compare mean performance scores against mean importance scores for various items types in sorted order of item importance, to give clarity on the rated performance on items compared to each item's importance scores. Point geometries were used for mean scores with colour to denote the relative difference between the improvement and performance scores,
- 5) Heat maps show the percentage of respondents who rated an item a score on the Likert scale, represented using a monochromatic colour scale with facets denoted by different major areas of study/research,
- 6) UpSet plot is a visualisation technique described since 2014 to display aggregates of intersection data between different sets, compared to traditional approaches such as the Venn diagram [14]. Using combinatorial matrixes as the centre piece to understand the characteristic of intersection, it also allows us to visualise aggregate element numbers in bar graphs, and also summary statistics derived from elements within each intersection, such as using boxplots with UpsetR

package. We explore use of the UpSet plot in visualising library survey data.

5.2.2 Modelling Techniques

The following statistical modelling techniques are used in the project:

- 1) We use a novel way of using principal components analysis (PCA) to gain insight into the library survey results. Respondent ratings for questions items are used to plot items within the principal components spaces and see the proximity of different items in this space, in order to see how questions are closed related based on respondent inputs.
- 2) One-way analysis of variance (ANOVA) will be conducted to assess for significance of differences in satisfaction level of respondents, according to their role (e.g. faculty, undergraduate students, and postgraduate students).

6 CASE STUDY & DISCUSSION

We would now like to use the interactive visual analytics tool in the context of the February 2018 SMU Library Survey data in specific use cases to highlight the strengths of the interactive applications as compared to the current results generated.

6.1 Demographics: Pie Charts, Sunburst Diagrams and Grouped Bar Charts

As described in the storyboard and user journey, analysts and applications users should be informed on the characteristics of respondents to the library survey. While in current published results this is captured in a form one statistical table showing aggregates and percentages according to respondent role, major area of study and international exchange status, our application captures the information in three sections.

The first is a pie chart showing proportion of respondents according to which library they visit more frequently, to give analysts a sense of relative frequency of library usage.

The second is a sunburst diagram with 2 concentric circles, and the hierarchy based on the general role (e.g. faculty, undergraduate, postgraduate) and specific role (e.g. under general faculty role, there are specific roles assigned to full Professors, Associate Professors and Assistance Professors). Aggregate statistics and percentages are captured in the centre of the sunburst which change based on user scroll-over.

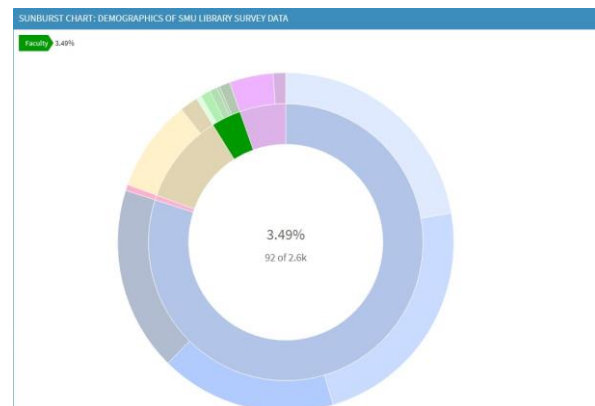


Fig. 2. Sunburst diagram highlighting faculty proportion (green) as 3.49% of total respondents. Majority of respondents are students (blue), standing at 79.9% of total responses of 2600.

The last section is through grouped bar charts which show number of respondents according to frequency, with separate bars to denote type of visit – campus, library or online resource use. This improves upon the separated bar graphs in the original results. From the diagram below, we clearly see for visits to the library, most respondents still visit the library on weekly or daily bases, although most respondents come on campus daily. Another new insight revealed is how use of online library resources still lags physical library use.

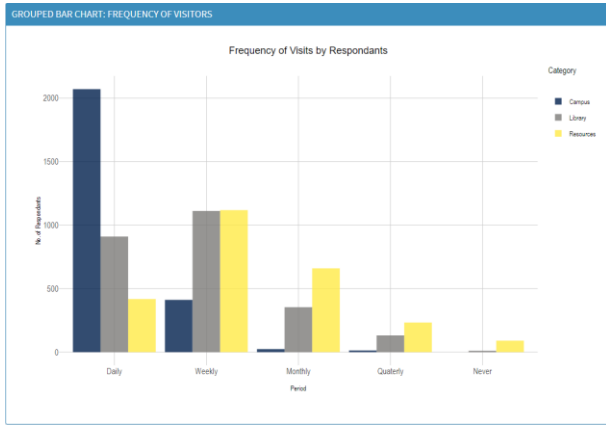


Fig. 3. Grouped bar charts show that the physical campus and library visits are common and online library resources still trail behind physical library use.

6.2 Divergent Stack Bar Chart

The divergent stacked bar chart shows a new dimension of respondent data not captured by current survey results. It is an accurate representation of aggregate sentiments respondents in Likert scales, as numerical values in such scales are ordinal variables. Currently, published results capture the mean values of performance and importance ratings in the reports. Design principles on colour scales have been described in section 5.2.1. The figure below shows the scroll-over tooltip which shows that 2.0% of respondents strongly disagreed (2.0%) that they could find a place to work in groups in the library. However, general sentiment from the colour scale shows most people agree and strongly agree on positive outcomes of service quality across all 4 question dimensions.

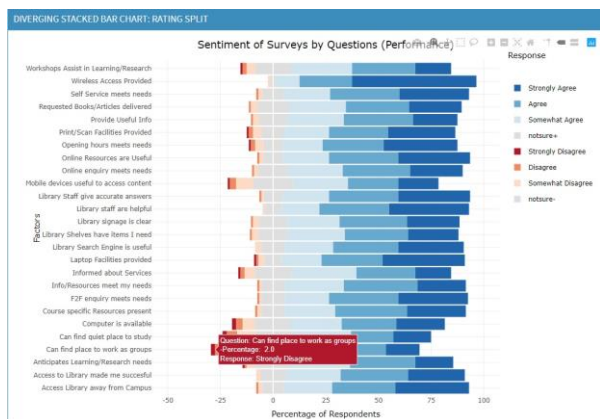


Fig. 4. Divergent stacked bar charts with tooltip functionality being used.

6.3 Dumbbell Plots

We propose Dumbbell plots as superior alternative for survey data results representation compared to gap grids from current survey results.

Gap grids currently used place questions items on a two-dimensional plan, plotting mean importance scores against mean performance scores, both multiplied by 10. Colour coding is used to highlight the gap between importance and performance scores for each questions, where red denotes where the performance score is lower than the importance score. This is because the sorting of items is in accordance to their importance ratings which give application users and analysts a sense of which are the most important items to focus on, and how the library is performing on these uppermost items.

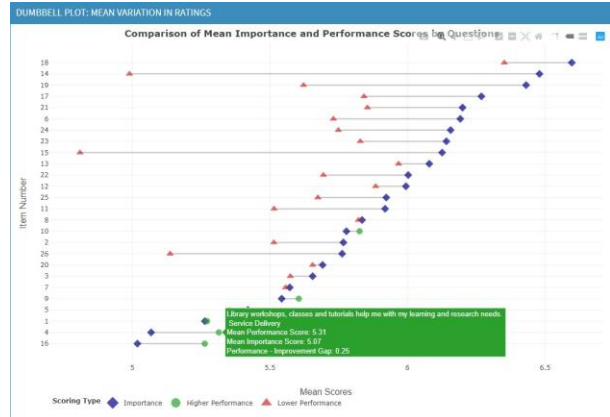


Fig. 5. Dumbbell plots sorted in order of importance score allow users to easily see the most pressing concerns of respondents and how the library fares, by looking at performance scores on the same horizontal axis.

The linear sort arrangement is superior to the gap grid display of questions in 2-dimensional grid as it makes it very difficult for comparison by the user across 2 dimensions. The Dumbbell plot allows comparison of performance and importance by linear gaps, and the colour aesthetic highlights clearly whether the performance score is higher than the importance score (denoted by green), or if the library performance on that item is lagging behind the importance placed on the item (denoted by red).

6.4 Heat Maps

Heat map plots are designed to show the different major areas of study/research, instead of the current published results which is captured in tables which capture only the top 5 importance scores items, performance score items, and gap scores in pages 32-34 of published results [7]. Rollover allows users to see the proportion a monochromatic colour scale to make inferences. Using an example from the figure below, we can see that across the board, larger proportions of accountability respondents gave performance ratings of 6, compared to those in law or economics.

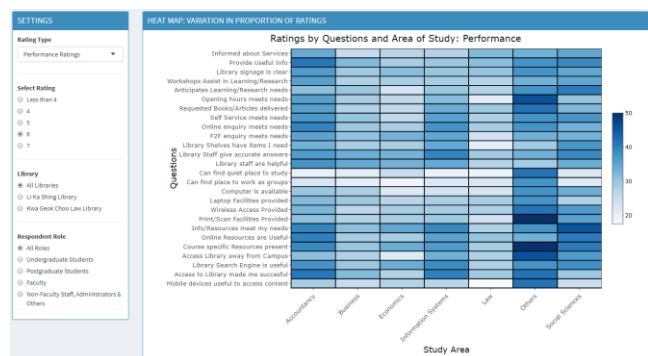


Fig. 6. Heat map of proportion of performance rating of 6 across different major area of study/research, by each question item.

6.5 UpSet Plot

UpSet plots are used to show aggregates of set intersections in large combinatorial matrices. Our project uses the UpSet plot to enlighten analysts of the Library Survey data “Can I get an idea on respondent groups based on the *variation of their responses* across different survey items?”

Design of the user interface through allowing the user to select the types of items to use for UpSet plot analysis through (i) selecting one of the four dimensions: ‘Communication’, ‘Service Delivery’, ‘Facilities and Equipment’, and ‘Information Resources’, and (ii) selecting the ‘Importance’ or ‘Performance’ response for comparison. This is to allow analysis of results in a suitable context.

Due to the explosion of the combinatorial matrix from having too many items during comparison, constraint has been fixed toward sorting different set intersection by descending order, showing a maximum of 20 set intersections. This prevents information overload on the analysts and clearer display of the data over the application user interface page.

In one instance of the UpSet plot generated, respondents have been grouped based on whether they have given Likert scale rating of ‘6’ or more for performance items in the dimension of ‘Communication’. We see that a majority set of 780 respondents have rated all three ‘Communication’ question items 1, 2 and 3 with ‘6’ and above, with the second largest group of 312 respondents who have rated item 3 above ‘6’, but have rated questions items 1 and 2 below ‘6’, and the third largest group of 261 respondents who rated item 2 and 3 above ‘6’, and item 1 below ‘6’. This should prompt emphasis on improving service quality for item 1, followed by item 2. To the left of the combination matrix, we see the breakdown on responses which fulfil criteria for each individual items, and below we see the boxplot summary of recommendation scores of respondents (how likely they will recommend the library to students – scored on scale of 1 to 10) faceted according to set intersections.

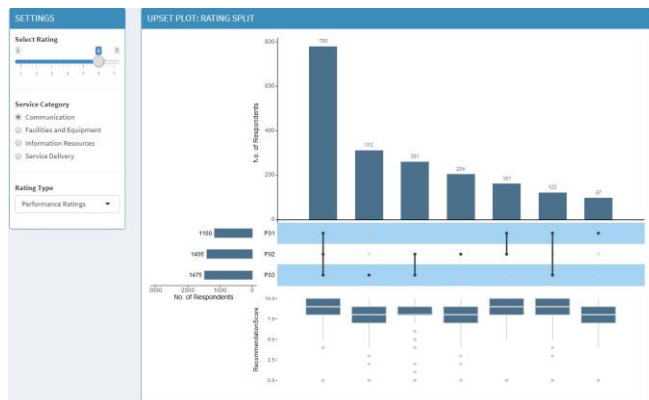


Fig. 7. Upset plot combination matrix with appended boxplot summary of recommendation scores.

6.6 Principal Components Analysis

FactoMineR package was used for principal component analyses. Using all respondent survey inputs across various questionnaire items as input, singular value decomposition method was employed and calculated eigenvalues for individual principal components were displayed in a scree plot to guide analysts in determining which principal components to be used for further visualisation.

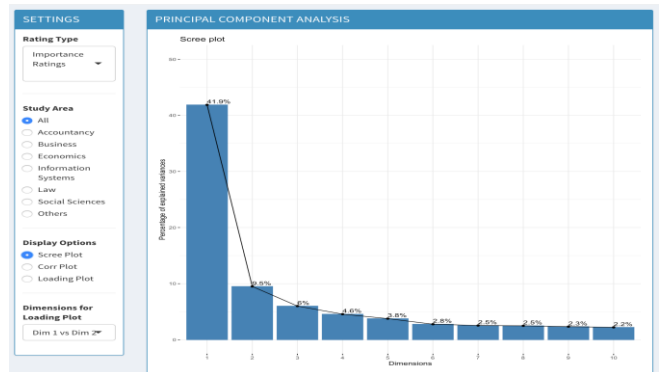


Fig. 8. Scree plot during PCA analysis for selection of principal components.

The loading plot shows the contributions of individual items in the principal components, and such information is supplemented by the factor loading table in PCA, to give the analyst an idea of the weight of items in contributing to variation, and help in finding unique clusters of questions.

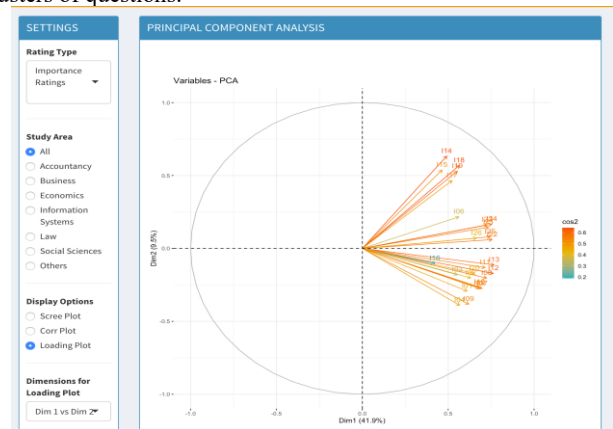


Fig. 19. The loading plot displays the strength of individual items influence the principal component.

6.7 Analysis of Variance (ANOVA)

One-way ANOVA statistical modelling was used to assess if difference in the responses in questions items in four dimensions across 5 groups of respondents were significant. This was after quantile-quantile (QQ) plots were created to check that data adhere to normal distribution for subsequent analysis. In the figures below, statistically significant difference was obtained only in comparison of faculty and respondents deemed in other roles. This will guide analysts with understanding whether calculated differences are statistically significant and guide whether to take action upon the data.

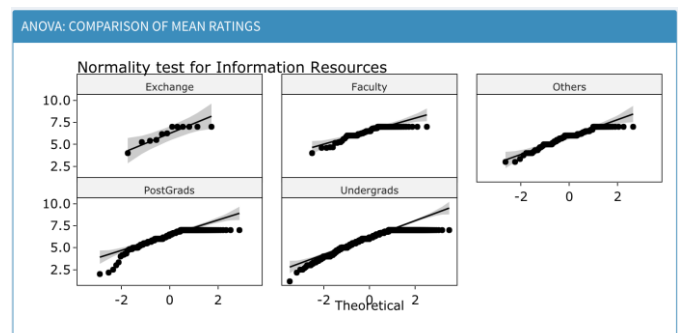


Fig. 10. QQ plots assess the normality of data prior to ANOVA analysis

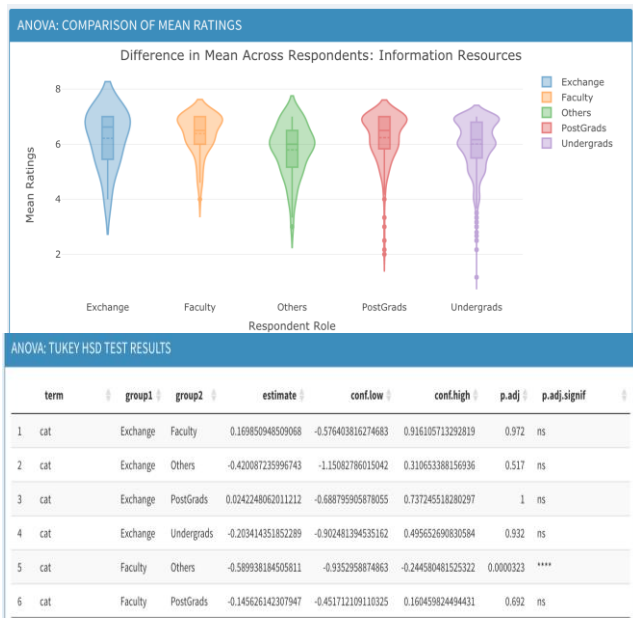


Fig. 11. ANOVA analysis results with statistical significance obtained only in comparison of faculty group with 'other roles' group, for analysis of boxplot and violin plots in mean ratings above.

7 CONCLUSION & EXTENSION

We have shown how we are able to better communicate and derived insights from library survey data using the analytics tool created using R Shiny framework as we compared findings and visualisations with current published results from SMU's February 2018 Library survey. This is achieved through parameter refinement through filters and options afforded to the user, coupled with interactivity such as tooltips to promote human computing interaction, user engagement and knowledge discovery. The design of the user journey and building of knowledge sequential also allows comprehensive review of various aspects of the survey data. The analytics tool can be deployed for use in SMU Library's February 2020 survey, if the data is available.

We propose the following as extensions for our project. First, topic modelling using free text comments can be used to better understand the sentiment of user comments in the February 2018 SMU Library Survey data. Next, we can reframe how we view libraries in the future and create survey questions and dimension which are in line with the future direction of the library [15]. With an addition of add data function to the application, we will be able to use SMU Library's February 2020 survey data for analysis.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Associate Professor KAM Tin Seong, from the School of Information Systems, Singapore Management University for his mentorship and unwavering support during the course of this project.

REFERENCES

- [1] Hiller, S., & Self, J. (2004). From Measurement to Management: Using Data Wisely for Planning and Decision-Making. *Library Trends*, 53(1), 129–155.
- [2] Using key performance indicators to measure library performance. Leo Appleton. *Library Connect*. 4th April 2018. Available:

<https://libraryconnect.elsevier.com/articles/using-key-performance-indicators-measure-library-performance>

- [3] Assessing Library Performance: New Measure, Methods and Models. Julia C. Blixrud. *Proceedings of the IATUL Conferences*. Paper 9. Available: <https://docs.lib.purdue.edu/iatul/2003/papers/9>
- [4] LibQUAL+: Charting Library Service Quality. Available: http://www.libqual.org/about/about_lq/general_info
- [5] InSync Surveys. Available: <https://insync.com.au/>
- [6] Library Service Quality Survey Results for 2018 Available: <https://library.smu.edu.sg/about/about-us/library-service-quality-survey/2018-survey-results>
- [7] Singapore Management University Library Survey Report – Scope: All Respondents, February 2018. Available: https://library.smu.edu.sg/sites/library.smu.edu.sg/files/library/pdf/SMU_Libraries_BeHeardSurvey_FullReport.pdf
- [8] Tableau - White Paper on Visualizing Survey Data. Version 2.0 June 2016. Steve Wexler.
- [9] How to Visualize Importance vs. Satisfaction Survey Responses. *Data Revelations*. May 10th, 2017. Available: <https://www.datarevelations.com/resources/importancesatisfaction/>
- [10] Ben Shneiderman. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages (VL '96)*. IEEE Computer Society, USA, 336.
- [11] Heer, J. & Shneiderman, B. (2012). Interactive dynamics for visual analysis. 10. 30-55. 10.1145/2133416.2146416.
- [12] The Visual Analytics Process. *VisMaster: Visual Analytics – Mastering the Information Age*. Available: <https://www.vismaster.eu/faq/the-visual-analytics-process/>
- [13] Hadley Wickham. A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, vol. 19, no. 1, pp. 3–28, 2010.
- [14] Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. *IEEE transactions on visualization and computer graphics*, 20(12), 1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>
- [15] Performance Measures for 21st Century Libraries. Peter Brophy. *Forth Northumbria*. Available: <https://www.libqual.org/documents/admin/brophy.pdf>

AUTHORS' BIOGRAPHY

Dr. Joshua LAM is a medical doctor in Singapore with a decade of experience in clinical medicine and health IT. He interest spans across healthcare, informatics and data analytics. He graduated with the degree of Bachelor of Medicine and Bachelor of Surgery (MBBS) from Yong Loo Lin School of Medicine, National University of Singapore and is currently pursuing a Master of IT in Business (Analytics) at Singapore Management University.

Mr. Karthik NITYANAND is a data analyst in Singapore with an interest in big data analytics. He graduated with the degree of Bachelor of Engineering (EEE) from Jeppiaar Engineering College, Chennai (Affiliated to Anna University, Chennai) and worked in the SAP domain prior to moving to Singapore to pursue a Master of IT in Business (Analytics) at Singapore Management University.

Mr. Shreyansh SHIVAM is an engineer with 4 years' experience in industrial automation working in the oil and gas industry in Dubai. He graduated with the degree of Bachelor of Engineering majoring in Electronics and Instrumentation, from VTU, India. He is currently pursuing a Master of IT in Business (Analytics) at Singapore Management University.
