JOURNAL OF DIGITAL SYSTEM DEVELOPMENT

https://e-journal.uum.edu.my/index.php/jdsd

How to cite this article:
Wahid, J., Ting, C. N., Ho, J., Tee, B. Y.C, & Irfan, M. A. (2024). Analysing Library Usage Patterns: A Visual Exploration of Book Loan and Room Reservation Trends. *Journal of Digital System Developments, 2*(1), 154-172. https://doi.org/10.32890/jdsd2024.2.1.10

ANALYSING LIBRARY USAGE PATTERNS: A VISUAL EXPLORATION OF BOOK LOAN AND ROOM RESERVATION TRENDS

[1]Juliana Wahid, [2]Ting Cheng Nam, [3]Jonas Ho, [4]Brian Tee Yu Cheng & [5]Muhammad Amir Irfan

[1,2,3,4&5]Universiti Utara Malaysia, Malaysia
[1]Corresponding author: w.juliana@uum.edu.my

Received: 25/3/2024 Revised: 17/4/2024 Accepted: 23/4/2024 Published: 30/4/2024

ABSTRACT

Library usage refers to how users interact with library resources and spaces, including borrowing books, reserving study rooms, navigating library spaces, and using digital resources. Analysing library usage patterns helps libraries optimise resources, improve services, and better meet the needs of their users. This study aims to visually explore the patterns in book loan activities and room reservation behaviours using the R programming language to identify trends that provide preliminary data for decision-making when establishing library operational policies. Annual book loan and room reservation data from 2017 to 2023 from Perpustakaan Sultanah Bahiyah, Library of Universiti Utara Malaysia, were collected and converted into a compatible R-analysis format. The analysis includes exploratory data analysis and visualisation using base R functions, base R graphics, and ggplot2 functionality. The findings, such as annual book loan rates, trends in book loan rates between undergraduate and postgraduate students, frequency of room reservation discussions across years, and the relationship between book loans and room reservations over time, serve as foundational data to support decision-making for the efficient operation of libraries and to inform strategies for optimising library services and resources to meet user needs better.

Keywords: Data analysis, R programming, Library, book loan, room discussion booking.

INTRODUCTION

The ability to manage, comprehend, and derive insights from vast datasets is important in our rapidly evolving world. This skill is not exclusive to businesses; researchers and policymakers rely on it. As data

generation accelerates, it is essential to address its challenges. Proficiency in handling data is vital for fostering innovation, making informed decisions, and shaping the future of various sectors.

Academic libraries are pivotal in this landscape, as they manage and utilise data on vast collections of resources, including books, journals, multimedia materials, digital repositories and room bookings to enhance services and support their institutions' academic missions. Among the various datasets managed by academic libraries, book loans and room discussion booking data hold particular significance. Book loan data encompasses information on the circulation of library materials, including loan periods, borrower demographics, and circulation trends. Similarly, room discussion booking data consists of records about the reservation and usage of library spaces for academic discussions, group meetings, and collaborative activities. These datasets offer valuable insights into user behavior, resource utilisation patterns, and the effectiveness of library services and facilities.

Data analysis plays a crucial role in academic libraries by providing valuable insights that contribute to optimising services by assessing the usage patterns of resources, spaces, and services. By analysing data on book loans, room bookings, and resource utilisation, libraries can identify areas of high demand or underutilisation. This insight enables them to reallocate resources effectively, adjust service hours, and enhance the availability of popular materials and spaces, thereby optimising service delivery to better meet the needs of users. Through comprehensive data analysis, libraries can optimise operations and engage users effectively which indirectly contributing to fostering the intellectual growth and advancement essential for national development (Pauline Adeniran, 2011).

Academic libraries may face challenges in selecting and implementing appropriate analysis techniques and tools for analysing large volumes of data. Data analysis requires specialised skills and expertise, and libraries may encounter limitations in terms of personnel resources, training opportunities, and access to advanced analytical tools and software. Additionally, libraries must stay abreast of emerging trends and best practices in data analysis to effectively leverage analytical insights for decision-making purposes.

In this era of abundant data, the R programming language has emerged as a powerful tool for analysing large datasets (Mestiri, 2019). Renowned for its versatility in statistical computing and graphics, R enables users to visualize data effectively, aiding in pattern recognition and trend analysis. From this point of view, this study aims to conduct visual analysis of Perpustakaan Sultanah Bahiyah, Library of Universiti Utara Malaysia (UUM) usage pattern in terms of annual book loans and room discussion bookings, using the R programming language. The insights gained will identify areas for improvement and implement specific interventions to enhance overall user experience and satisfaction.

LITERATURE REVIEW

This literature review first examines studies that leverage data visualisation techniques to analyse and interpret library usage and services. Finch and Flenner (2016) generated data visualizations to compare sections of the library book collection with expenditures in those sections and student enrollment. They employed visualisation tools, including Plotly, Microsoft Excel, the Python programming language, the JavaScript library D3.js, and Tableau. Lund (2020) utilized the R programming language to access and examine tweets related to library services from prominent librarians, as well as to analyze posts concerning major book publishers. The analysis involved the utilisation of various packages including 'twitter', 'graphics', 'purr', 'stringr', 'tm', and 'syuzhet'.

The limited availability of literature on data visualization techniques specifically focused on analyzing and interpreting library usage and services suggests that this area of study is relatively underexplored or less documented compared to other topics within library science or data analytics. The next section of this review introduces R and highlights its implementation in other domains, providing valuable knowledge and inspiration that can be leveraged to enhance data analysis, visualization, and decision-making practices within library usage and service.

R was originally introduced by Robert Gentleman and Ross Ihaka, members of the statistical department at the University of Auckland in New Zealand, back in 1996. From 1997 onwards, the "R Core Team" took charge of its ongoing development. Notably, R is designed to function seamlessly across various operating systems, including Unix, Linux, Windows, and MacOS (Mestiri, 2019). The R programming language is a powerful and versatile tool widely used for statistical computing and data analysis tasks in various fields including healthcare (Islam et al., 2023; Konopka et al., 2018; Markazi-Moghaddam et al., 2020), amateur radio spot analysis (Chris, 2020), floral and geographical studies (Sichilongo et al., 2020), sentiment analysis of academic library tweets (Lund, 2020), and nutritional analysis (Vignesh & Nagaraj, 2022).

Islam et al. (2023) conducted exploratory data analysis using R Studio, focusing on a real-time dataset containing vital information regarding patients diagnosed with cardiovascular disease. The R packages used were ggplot2, tidyr, and dplyr. The data was efficiently processed and visually represented, aiding in drawing meaningful conclusions. Konopka et al. (2018) implemented a comprehensive toolkit tailored for thorough exploratory analysis of clinical datasets and conducted a case study analysis on a dataset focusing on patients. The aim was to explore the connections between health, genetics, and social status in the elderly population of Poland. The exploratory analysis primarily focused on numerical attributes and followed these steps: 1) normalization, 2) Principal Component Analysis (PCA), 3) detection and removal of outliers, and 4) clustering. The prcomp function from the base R package was employed for performing principal component analysis.

Markazi-Moghaddam et al. (2020) investigated the time intervals associated with the operating theater of a general hospital through a cross-sectional study utilizing data from patients undergoing elective surgery. The analysis commenced with data summarization and the identification of outliers and anomalies. The study examined whether significant differences existed between surgery categories concerning the duration of stay in the operating theater. In instances where the prerequisites for parametric tests were not met, the study employed a robust alternative to ANOVA (utilizing the R package WRS – Wilcox' Robust Statistics) by employing 20% trimmed means. Furthermore, robust post-hoc tests were conducted to ascertain differences between the means of all potential pairs of surgery categories. Addressing significant variations in group sizes among the surgery categories, a balanced dataset was created to assess potential improvements in results. The R package UBL was utilized for multi-class balancing.

Chris (2020) conducted a study on Amateur radio spots utilizing records from the Reverse Beacon Network (RBN) to identify patterns in various variables. A "spot" is defined as a situation where there exists a propagation path between a transmitter and a receptor location at a specific time and frequency. The distance between the transmitter and receptor is computed using the distHaversine function in R, which calculates the shortest distance between two points on a sphere based on their latitude and longitude. A tool was developed to filter data for a particular unique transmitter city and visually represent activities at that specific location. This tool was constructed using functions and tools from libraries such as maps, ggplot2, tidyverse, and plyr.

Sichilongo et al. (2020) employed a combination of the Metab R package along with other tools such as Automated Mass Spectral Deconvolution and Identification System (AMDIS), as well as MINITAB® statistical analysis software. Their objective was to classify the floral and geographical origins of three randomly selected commercially produced and three unprocessed natural organic honeys from Zambia and Botswana. They utilized gas chromatography mass spectrometry (GC–MS) untargeted metabolomics of volatile components for this purpose. The Metab R package was utilized to adjust peak intensities, present peak areas, eliminate false positives, perform normalization by internal standard and biomass, as well as conduct statistical tests such as analysis of variance (ANOVA) and t-tests on data generated by AMDIS for metabolomics, both targeted and non-targeted. Vignesh & Nagaraj (2022) applied Exploratory Data Analysis (EDA) to a nutritional dataset, employing visualization packages in R such as bar charts, histograms, box plots, scatterplots, barcodes, and violin plots.

In summary, the aim of this study to implement R programming for analysing and interpreting library usage of book loans and room discussion bookings is significant due to the existing gap in the literature and the practical relevance to library operations.

METHODOLOGY

The methodology section of this study outlines the procedures employed to conduct a visual analysis of the annual book loan and room discussion reservation data from Perpustakaan Sultanah Bahiyah, Library of Universiti Utara Malaysia. The procedures consist of data collection, preparation, exploratory data analysis (EDA), data visualisation, interpretation of findings, and discussion of implications for decision-making.

Data collection involved gathering annual book loan and room discussion reservation data from Perpustakaan Sultanah Bahiyah, Library of Universiti Utara Malaysia, spanning from 2017 to 2023. Subsequently, the collected data underwent preparation, in which the data was converted from Excel format to a format that was suitable for analysis in R.

After data preparation, exploratory data analysis (EDA) was conducted to understand the characteristics and patterns present in the data. This involved calculating summary statistics and exploring relationships between variables. Based on these relationships, research questions were formulated to investigate the observed patterns further.

1. How do annual book loan rates vary across different academic years within the population?
2. How do trends in annual book loan rates differ between undergraduate and postgraduate students within the library?
3. Are there noticeable differences in the frequency of room discussion bookings for various academic years?
4. Are there significant fluctuations in room types of reservation frequencies between different academic years?
5. How does the number of books borrowed from the library relate to the frequency of room discussion reservations over time?

The next step is to utilise the R programming language and its associated libraries, namely base R graphics and *ggplot2*, to create visualisations that aim to address the research questions. *ggplot2* is a powerful and widely used package in R for creating graphics and data visualisations (Wickham, 2016). It provides a flexible and intuitive syntax for building various plots, including scatter plots, bar plots, line

plots, histograms, and more. The use of visual representations such as bar plots, line plots and scatter plots facilitates a clearer comprehension of each research question.

Upon the generation of visualisations, the findings were also interpreted, focusing on key metrics such as annual book loan rates, trends across different borrower categories, frequency of room discussion reservations over time, and any correlations between book loans and room reservations. These interpretations were contextualized within the framework of library operations and decision-making processes.

DATA PREPARATION

The datasets that were collected consist of two Excel file formats (.xlsx) named *Pinjaman Buku Mengikut Kategori Pengguna 2017 – 2023* and *Tempahan Bilik Karel dan Perbincangan*. However, these files are not directly compatible with R. These files were converted to a table format, such as Comma-separated values (CSV). Next, both of CSV files were imported using the *read.csv()* function in the R environment, then a variable was assigned to each of this CSV file namely as *loanbook* for *Pinjaman Buku Mengikut Kategori Pengguna 2017 – 2023.csv* and *appointment* for *Tempahan Bilik Karel dan Perbincangan.csv*.

EXPLORATORY DATA ANALYSIS

The structure and class of the dataset were observed with the *str()* function, while the glimpse of the whole dataset can be viewed using *head()* and *tail(()* function which shows the first and last 6 rows of the dataset. Figure 1 shows the structure and several rows of the dataset. It is shows that *loanbook* has 84 objects (rows) and 15 variables (columns) with its data type consists of *Tahun (integer)*, *Bulan (Factor)*, *Ijazah Pertama (integer)*, *Pelajar Jarak Jauh (integer)*, *Master Full Time (integer)*, *Master Part Time (integer)*, *PhD Full Time (integer)*, *PhD Part Time (integer)*, *Ijazah Pertama UUMKL (integer)*, *Master UUMKL (integer)*, *PhD UUMKL (integer)*, *Pinjaman Antara Perpustakaan (integer)*, *Akademik (integer)*, *Pentadbiran (integer)* and *Pesara (integer)*. Meanwhile, *appointment* has 84 objects (rows) and 4 variables (columns) with its data type consisting of *Tahun (integer)*, *Bulan (character)*, *TEMPAHAN_BILIK_PERBINCANGAN (integer)* and *TEMPAHAN_KAREL_TERTUTUP (integer)*.

Figure 1

R Result on Viewing the Structure and Several Rows Data of Dataset

```
> str(loanbook)
'data.frame':   84 obs. of  15 variables:
 $ Tahun                        : int  2017 2017 2017 2017 2017 2017 2017 2017 2017 2017 ...
 $ Bulan                        : Factor w/ 12 levels "Apr","Dis","Feb",..: 4 3 7 1 8 6 5 10 12 11 ...
 $ Ijazah.Pertama               : int  148 3976 5795 3074 1929 432 393 383 3530 4681 ...
 $ Pelajar.Jarak.Jauh           : int  25 44 26 2 32 6 46 13 28 8 ...
 $ Master.Full.Time             : int  149 580 388 330 315 110 109 53 454 382 ...
 $ Master.Part.Time             : int  68 281 323 252 169 72 62 20 148 261 ...
 $ PhD.Full.Time                : int  264 303 335 200 208 174 359 213 268 276 ...
 $ PhD.Part.Time                : int  51 38 62 72 69 32 86 109 53 100 ...
 $ Ijazah.Pertama.UUMKL         : int  0 0 12 2 1 0 0 0 0 0 ...
 $ Master.UUMKL                 : int  8 18 16 12 11 2 1 0 20 25 ...
 $ PhD.UUMKL                    : int  0 0 4 0 2 0 1 0 5 5 ...
 $ Pinjaman.Antara.Perpustakaan: int  25 15 33 23 20 9 17 14 7 8 ...
 $ Akademik                     : int  345 458 305 300 347 197 471 537 295 345 ...
 $ Pentadbiran                  : int  57 58 103 110 102 92 93 79 106 93 ...
 $ Pesara                       : int  0 0 0 0 0 0 0 0 0 0 ...
> str(appointment)
'data.frame':   84 obs. of  4 variables:
 $ Tahun                        : int  2017 2017 2017 2017 2017 2017 2017 2017 2017 2017 ...
 $ Bulan                        : chr  "Jan" "Feb" "Mac" "Apr" ...
 $ TEMPAHAN_BILIK_PERBINCANGAN: int  72 53 127 118 274 38 12 10 127 324 ...
 $ TEMPAHAN_KAREL_TERTUTUP    : int  40 61 54 53 180 146 135 160 163 242 ...
> head(loanbook)
  Tahun Bulan Ijazah.Pertama Pelajar.Jarak.Jauh Master.Full.Time Master.Part.Time PhD.Full.Time PhD.Part.Time Ijazah.Pertama.UUMKL Master.UUMKL PhD.UUMKL Pinjaman.Antara.Perpustakaan Akademik Pentadbiran Pesara
1  2017   Jan            148                25             149              68           264            51                    0             8          0                           25       345          57      0
2  2017   Feb           3976                44             580             281           303            38                    0            18          0                           15       458          58      0
3  2017   Mac           5795                26             388             323           335            62                   12            16          4                           33       305         103      0
4  2017   Apr           3074                 2             330             252           200            72                    2            12          0                           23       300         110      0
5  2017   Mei           1929                32             315             169           208            69                    1            11          2                           20       347         102      0
6  2017   Jun            432                 6             110              72           174            32                    0             2          0                            9       197          92      0
> tail(loanbook)
   Tahun Bulan Ijazah.Pertama Pelajar.Jarak.Jauh Master.Full.Time Master.Part.Time PhD.Full.Time PhD.Part.Time Ijazah.Pertama.UUMKL Master.UUMKL PhD.UUMKL Pinjaman.Antara.Perpustakaan Akademik Pentadbiran Pesara
79  2023   Jul            327                 5              21               2            64             7                    0             2          0                            2        85          32      0
80  2023  Ogos            196                 0              17               4            50             3                    0             0          0                            1        70          33      0
81  2023  Sept              2                 0              97               8           114            12                    0             0          0                            2       235          22      2
82  2023   Okt            882                 0              94               8            31            16                    0             2          0                            1       115         100      0
83  2023   Nov           1195                 4              25               9            61            31                    0             0          0                            4        71          70      1
84  2023   Dis           1808                10              36               4           200            78                    0            16          4                            6       604          79      0
> head(appointment)
  Tahun Bulan TEMPAHAN_BILIK_PERBINCANGAN TEMPAHAN_KAREL_TERTUTUP
1  2017   Jan                          72                      40
2  2017   Feb                          53                      61
3  2017   Mac                         127                      54
4  2017   Apr                         118                      53
5  2017   Mei                         274                     180
6  2017   Jun                          38                     146
> tail(appointment)
   Tahun Bulan TEMPAHAN_BILIK_PERBINCANGAN TEMPAHAN_KAREL_TERTUTUP
79  2023   Jul                         177                      34
80  2023  Ogos                          88                       3
81  2023  Sept                           2                       2
82  2023   Okt                          50                      10
83  2023   Nov                         170                      18
84  2023   Dis                          45                       2
```

Based on the research questions determined earlier, next step, we examined various factors in the dataset and visualized them using both base R graphics and ggplot2 functionality to see how they relate to each other. The plots obtained in the study help us to see the patterns and trends in the data more clearly. The description of this process will be detailed in the next section.

DATA VISUALIZATION AND ANALYSIS

This section will address the research questions stated in the methodology section using base R function, base R graphics and ggplot2 functionality.

1. How do annual book loan rates vary across different academic years within the population?

To get the result, firstly we total up all the value for each column in *loanbook* dataset based on *Tahun (Year)* variable except column *Bulan* using codes as shown in Figure 2. The *aggregate()* function is used to calculate the sum of numeric variables in a dataset *loanbook* grouped by the variable *Tahun (Year)*. The dot (.) means to apply the aggregation function to all numeric columns in the dataset, while ~ *Tahun (Year)* specifies that we want to aggregate the data based on the variable *Tahun (Year)*. The code *data = loanbook[, -2]* specifies the dataset to be used for aggregation, which is *loanbook*. However, *[, -2]* part suggests that the second column of the *loanbook* dataset is excluded from the aggregation. This typically means that the second column contains non-numeric data or it is a column that we do not want to include in the aggregation. The code *sum* specifies the aggregation function to be used, which is sum in this case. It tells R to calculate the sum of each numeric variable for each group defined by the *Tahun (Year)* variable and stores the result in a new data frame called *total_by_year*. In the second line, the code *total_by_year* will display the content of the *total_by_year* variable, showing the result of the aggregation with the total number of user type for each year from 2017 until 2023 (Figure 3).

Figure 2

R Codes for Total Up Loanbook Dataset Based on Tahun (Year) Variable

```
total_by_year <- aggregate(. ~ Tahun, data = loanbook[, -2], sum)
total_by_year
```

Figure 3

Result for Total Up Loanbook Dataset Based on Tahun (Year) Variable

```
> total_by_year <- aggregate(. ~ Tahun, data = loanbook[, -2], sum)
> total_by_year
  Tahun Ijazah.Pertama Pelajar.Jarak.Jauh Master.Full.Time Master.Part.Time PhD.Full.Time PhD.Part.Time Ijazah.Pertama.UUMKL Master.UUMKL PhD.UUMKL Pinjaman.Antara.Perpustakaan Akademik Pentadbiran Pesara
1  2017         29806              257             3367             2047            3014            767                  15           146         21                          205      4008        1027      0
2  2018         19864              158             2363             1146            2075            439                   0            92         44                          184      3088         614      0
3  2019         17171               54             2343              893            1785            515                   0            28         13                          141      3027         667      0
4  2020          4715               65              678              174             482            144                   0            12         10                           66       907         320      0
5  2021           813               85              137               75             143            103                   0             0         15                           92       757         308      2
6  2022          4432               35              347               75             391            187                   0             5          0                           79      1410         388      1
7  2023          6943               43              663               65             841            196                   0            27          4                           34      1601         647      4
```

Next, we created a bar plot with years on the x-axis and the total number of book loans for each type of user on the y-axis. Each type of user (*Ijazah Pertama*, *Pelajar Jarak Jauh*, *Master Full Time*, *Master Part Time*, *PhD Full Time*, *PhD Part Time*, *Ijazah Pertama UUMKL*, *Master UUMKL*, *PhD UUMKL*, *Pinjaman Antara Perpustakaan*, *Akademik*, *Pentadbiran* and *Pesara*) will have its own bar, showing the corresponding book loan rate. Overall, the codes as shown in Figure 4, generate a bar plot showing the total numbers for each year, with each type of user variable represented by a separate bar and different colours. Figure 5 shows the result of the code which is the bar plot.

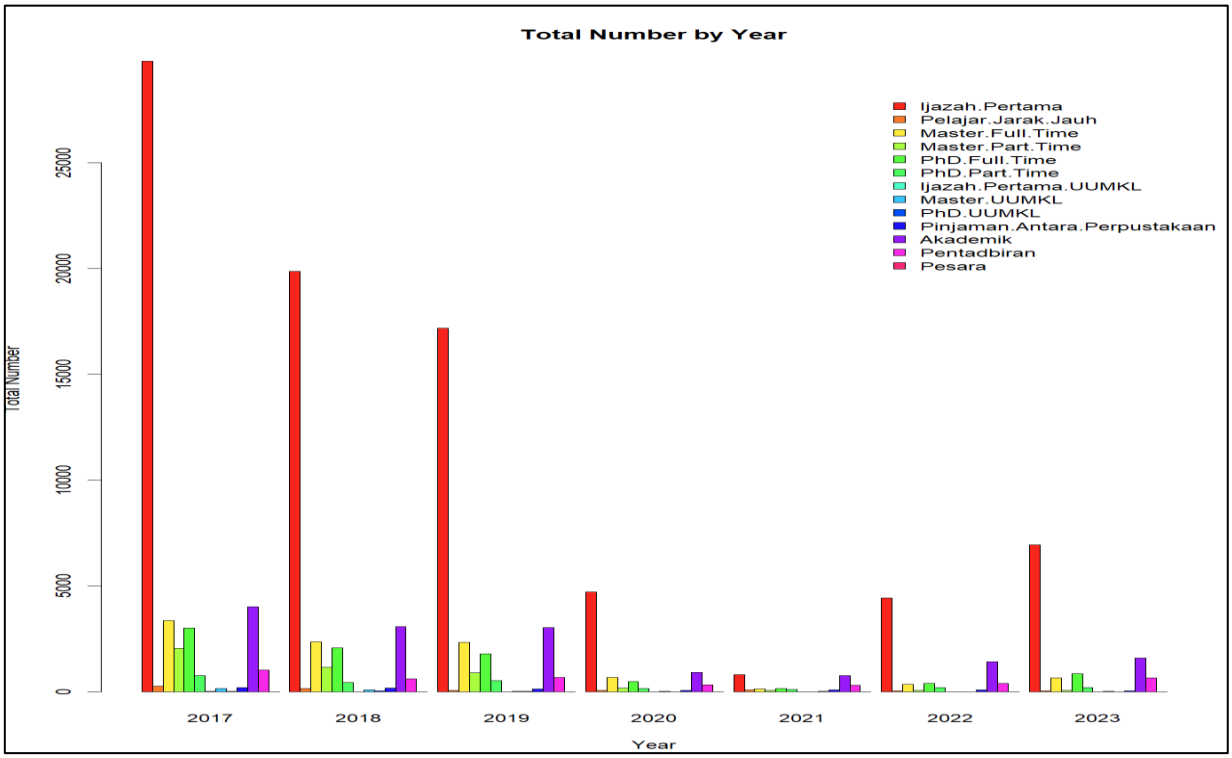
Figure 4

R Codes for Creating Bar Plot with Number of book loan based on Year and User Types

```
barplot(t(as.matrix(total_by_year[, -1])), beside = TRUE,
        col = rainbow(ncol(total_by_year) - 1),
        names.arg = total_by_year$Tahun,
        main = "Total Number by Year",
        xlab = "Year",
        ylab = "Total Number",
        legend.text = colnames(total_by_year[, -1]),
        args.legend = list(x = "topright", bty = "n", inset = c(0.05, 0.05))
)
```

Figure 5

Bar Plot with Number of book loan based on Year and User Types



In Figure 5, the borrowing activity shows a consistent decrease from year 2017 to 2021 for all types of users. These could be due to shifts in user behavior, includes increase in borrowing of e-books compared to physical books and preference for online resources or remote services. In addition, the COVID-19 pandemic has had a profound impact on how people engage with libraries. With restrictions on physical access to libraries and a heightened emphasis on social distancing, there has been a significant increase in demand for digital resources. Users are relying more on e-books, online databases, and digital collections for research, leisure reading, and academic purposes. However, there is slightly increasing trends in borrowings activity from 2022 until 2023. As UUM reopens after the pandemic and public health restrictions ease, users may seek out in-person activities, including visiting libraries and borrowing books. The reopening of the library provides users with access to physical collections, browsing opportunities, and in-person assistance from library staff, which may lead to an uptick in borrowing activity. While digital resources and online services played a significant role during the pandemic, there remains a distinct appeal to the physical experience of browsing shelves and discovering new books in libraries.

Based on the insights gleaned from the bar plot, UUM library administrators might prioritize acquiring more electronic resources or investing in technologies that support e-book lending, online databases, virtual reference services, or remote access to digital collections. In addition, UUM Library can offer various services beyond borrowing, such as research assistance, workshops, events, and maker spaces. Tracking the attendance or participation rates in these activities can help identify emerging interests or needs among library users. For instance, a surge in attendance at technology workshops may indicate growing interest in digital literacy skills, prompting the library to expand its technology-related programs. UUM Library can respond by enhancing their online presence, providing responsive online assistance, and fostering virtual communities. Next, UUM Library can also invest in user-friendly self-service kiosks, mobile apps, or online platforms that empower users to independently manage their library accounts and access resources.

2. How do trends in annual book loan rates differ between undergraduate and postgraduate students within the library?

Undergraduate (UG) and postgraduate (PG) students often represent two major user segments in academic libraries. Understanding how book loan rates differ between these groups provides valuable insights into the borrowing behaviors and needs of the primary users of the library resources. To obtain the result for this research question, a line plot with months of the year on the x-axis and the number of book loans on the y-axis was created. Each line on the graph represents UG and PG categories, showing how their borrowing behaviors vary throughout the year.

In order to create the line plot, firstly, a *ggplot2* library was loaded. Next, the *loanbook* dataset was separated into data frames based on year which named as *loanbook2017, loanbook2018, loanbook2019, loanbook2020, loanbook2021, loanbook2022* and *loanbook2023*. Subsequently, several other data frames were created respectively based on these data frames that contain four variables such as *Tahun (Year), Bulan, UG* and *PG*. *UG* contains total number of borrowings for UG that sum the values in columns 2, 3 and 8 corresponding to UG variables such as *Ijazah Pertama, Pelajar Jarak Jauh* and *Ijazah Pertama UUMKL* respectively. *PG* contains total number of borrowings for PG that sum the values in columns 4, 5, 6, 7, 9 and 10 corresponding to PG variables to *Master Full Time, Master Part Time, PhD Full Time, PhD Part Time, Master UUMKL* and *PhD UUMKL* respectively. These data frames were named as *UGPG2017, UGPG2018, UGPG2019, UGPG2020, UGPG2021, UGPG2022* and *UGPG2023*.

Next, each of *UGPG2017, UGPG2018,* and so on were reshaped from wide to long format and assigned to new variable. This is to allows for easier comparison and visualization of multiple variables over time. The *Bulan* variable was converted to a factor with the desired order of levels. This ensures that the months are plotted in the correct order on the x-axis of the plot. To create a line plot, ggplot() function was used, where *Bulan* (Month) is plotted on the x-axis, *Frequency_of_Borrowings* is plotted on the y-axis, and different colors are used to distinguish between UG and PG student borrowings (*color = Student_Types*). The *theme_minimal()* function sets a minimal theme for the plot. Figure 6 shows sample codes for the year 2017 data on overall steps mentioned in this and previous paragraph. Figure 7 shows the line plots for frequency of books loan for the year 2017 until 2023 between UG and PG students.

Overall, as shown in Figure 7, the trends between UG and PG students show fluctuations in borrowing activity throughout the year 2017 until 2023, with both groups experiencing periods of increase and decrease. UG students generally borrow more books than PG students across the months each year, with some exceptions such as July of 2017, July of 2018, July of 2019, January of 2021, September of 2022, and September of 2023. These trends may reflect differences in academic schedules, study habits, or resource needs between undergraduate and postgraduate students.

Figure 6

R Codes for Creating Line Plot with Number of book loan based on Year 2017 and User Types

```
library(ggplot2)
loanbook2017<-loanbook[loanbook$Tahun==2017,c(1:11)]

UGPG2017<-data.frame(Tahun=loanbook2017$Tahun,Bulan=loanbook2017$Bulan,
UG=rowSums(loanbook2017[,c(3,4,9)]), PG=rowSums(loanbook2017[,c(5,6,7,8,10,11)]))

# Define the desired order of months
month_order <- c("Jan", "Feb", "Mac", "Apr", "Mei", "Jun",
                 "Jul", "Ogos", "Sept", "Okt", "Nov", "Dis")
```

```
# Reshape the data from wide to long format
booking_data_long2017 <- tidyr::pivot_longer(UGPG2017, cols = c(3, 4),
                names_to = "Student_Types",values_to = "Frequency_of_Borrowings")

# Convert the 'Bulan' variable to a factor with the desired order of levels
booking_data_long2017$Bulan   <-   factor(booking_data_long2017$Bulan,   levels   =
month_order)

ggplot(booking_data_long2017, aes(x = Bulan, y = Frequency_of_Borrowings, color =
Student_Types, group = Student_Types)) + geom_line() +
  labs(title = "Frequency of Book Loans in 2017",
       x = "Month",
       y = "Frequency of Borrowings",
       color = "Student Types") + theme_minimal()
```

Figure 7

Line Plots for Frequency of Books Loan between UG and PG students for year 2017 until 2023



UUM library administrators could consider reallocating resources based on the borrowing trends. Since UG students generally borrow more books, the library may need to ensure sufficient stock and availability of materials that cater to their needs. This might involve expanding collections in areas relevant to UG courses or popular topics. Recognizing the exceptions where PG borrowing peaks, the library administrators could implement targeted promotional campaigns to encourage more borrowing among PG students during those months. This could involve showcasing relevant resources, offering incentives, or organizing events tailored to PG students' interests or academic requirements. To accommodate fluctuations in borrowing activity, the library administrators may need to adjust service hours, staffing levels, or loan policies. For example, during peak borrowing periods, they could extend operating hours or provide additional assistance to manage increased demand. Since borrowing activity reflects academic schedules and study habits, the administrators could also review the allocation of study spaces within the library. They might consider adjusting seating arrangements or providing designated areas that cater to the specific needs of UG and PG students during peak usage times.

3. Are there noticeable differences in the frequency of room discussion bookings for various academic years?

To get the result from this research question, a bar plot like the one described in the first research question above was created using *appointment* dataset. Similar codes are also used to total up all the value for each column in *appointment* dataset based on *Tahun (Year)* variable that excludes column *Bulan* and stores the result in a new data frame called *totalApp_by_year* as shown in Figure 8. In the second line, the code *totalApp_by_year* will display the content of the *totalApp_by_year* variable, showing the result of the

aggregation with the total number of rooms discussion (*BILIK PERBINCANGAN* and *KAREL TERTUTUP*) reservations for each year from 2017 until 2023 as shown in Figure 9.

Figure 8

R Codes for Total up appointment Dataset Based on Tahun (Year) Variable

```
totalApp_by_year <- aggregate(. ~ Tahun, data = appointment[, -2], sum)
totalApp_by_year
```

Figure 9

Result for Total up appointment Dataset Based on Tahun (Year) Variable

```
> totalApp_by_year <- aggregate(. ~ Tahun, data = appointment[, -2], sum)
> totalApp_by_year
  Tahun TEMPAHAN_BILIK_PERBINCANGAN TEMPAHAN_KAREL_TERTUTUP
1  2017                        2130                    1685
2  2018                        1265                    1242
3  2019                        2861                    1659
4  2020                         381                     513
5  2021                           0                       0
6  2022                         697                      55
7  2023                        1255                     174
```

Next, we created a bar plot with years on the x-axis and the total number of room reservations on the y-axis. Each room type discussion (*BILIK PERBINCANGAN* and *KAREL TERTUTUP*) will have its own bar, showing the corresponding room reservation rate. Overall, the codes as shown in Figure 10, generate a bar plot showing the total numbers for each year, with each type of room represented by a separate bar and different colours. Figure 11 shows the result of the code which is the bar plot.

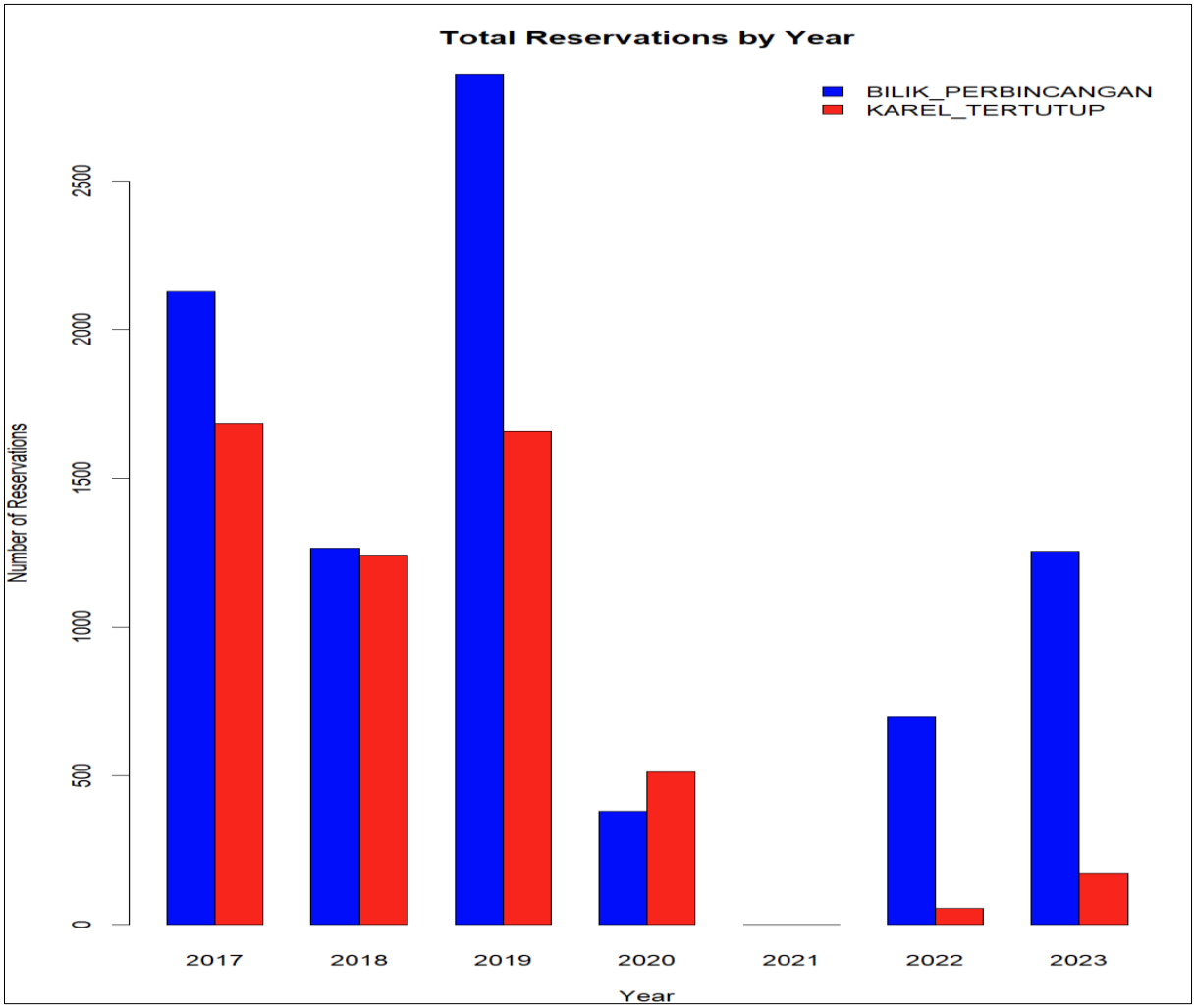
Figure 10

R Codes for Total up appointment Dataset Based on Tahun (Year) Variable

```
barplot(t(as.matrix(totalApp_by_year[, -1])), beside = TRUE, col = c("blue",
"red"),
        names.arg = totalApp_by_year$Tahun,
        xlab = "Year",
        ylab = "Number of Reservations",
        main = "Total Reservations by Year",
        legend.text = c("BILIK_PERBINCANGAN", "KAREL_TERTUTUP"),
        args.legend = list(x = "topright", bty = "n")
)
```

Figure 11

Bar Plot with Number of Discussion Room Reservation based on Year



By examining the heights of the bars for each room type across different years in Figure 12, it was consistent of user preferences over BILIK PERBINCANGAN from year 2017 until 2019. This is perhaps due to the increased emphasis on collaborative learning approaches in educational institutions that may drive higher demand for discussion rooms where students can engage in group study sessions, project work, or collaborative research. In addition, students may require dedicated spaces for team meetings, brainstorming sessions, or project discussions, leading to a preference for discussion rooms equipped with collaborative tools and technology. However, even though the number of both types of room reservation decreased in the year 2020 due to the influence of COVID-19, there is a steady increase reservation for KAREL TERTUTUP. The reason perhaps is that some users may prefer KAREL TERTUTUP for individual study, research, or focused work, as these spaces offer privacy and minimal distractions. During exam periods, students may seek out KAREL TERTUTUP as quiet study spaces conducive to concentration and exam preparation, leading to increased demand for these facilities. The movement control order due to COVID-19 have influenced the closing of university contribute to null reservation for both type of rooms in the year 2021. The reopening of UUM library provides users with access to physical services which indicate number of reservation or both type of rooms in the years of 2022 and 2023. Even though after the COVID-19 pandemic, it was consistent of user preferences and needs over BILIK PERBINCANGAN compared to KAREL TERTUTUP. Based on these insights, UUM

library may consider expanding the number of available BILIK PERBINCANGAN or adjusting booking policies to better accommodate user's demand.

4. Are there significant fluctuations in room types of reservation frequencies between different academic years?

To address this research question, a line plot with months of the year on the x-axis and the frequency of room reservation on the y-axis was created. Each line on the graph represents frequency of room reservation of a different room type, showing how room reservation frequencies fluctuate over time.

In order to create the line plot, similar to steps in addressing second research question earlier, firstly, a *ggplot2* library was loaded. Next, the *appointment* dataset was separated into data frames based on year which named as *appointment2017*, *appointment2018*, *appointment2019*, *appointment2020*, *appointment2021*, *appointment2022* and *appointment2023*. Next, each of *appointment2017*, *appointment2018*, and so on were reshaped from wide to long format and assigned to new variable. The *Bulan* variable was converted to a factor with the desired order of levels. This ensures that the months are plotted in the correct order on the x-axis of the plot. To create a line plot, ggplot() function was used, where *Bulan* (Month) is plotted on the x-axis, *Frequency_of_Reservations* is plotted on the y-axis, and different colors are used to distinguish between UG and PG student borrowings (*color = Room_Types*). The *theme_minimal()* function sets a minimal theme for the plot. Figure 12 shows sample codes for the year 2017 data on overall steps for creating the line plot. Figure 13 shows the line plots for frequency of rooms reservation for the year 2017 until 2023 of different room types.

Figure 12

R Codes for Creating Line Plot with Number of Room Reservation based on Year 2017 and Room Types

```
# Define the desired order of months
month_order <- c("Jan", "Feb", "Mac", "Apr", "Mei", "Jun",
                 "Jul", "Ogos", "Sept", "Okt", "Nov", "Dis")
appt_data_long2017 <- tidyr::pivot_longer(appointment2017, cols = c(3, 4),
        names_to = "Room_Types", values_to = "Frequency_of_Reservations")

# Convert the 'Bulan' variable to a factor with the desired order of levels
appt_data_long2017$Bulan <- factor(appt_data_long2017$Bulan, levels = month_order)

ggplot(appt_data_long2017, aes(x = Bulan, y = Frequency_of_Reservations, color =
Room_Types, group = Room_Types)) + geom_line() +
  labs(title = "Frequency of Room Reservations in 2017",
       x = "Month",
       y = "Frequency of Reservations",
       color = "Room Types") + theme_minimal()
```

Figure 13

Line Plots for Frequency of Room Reservation between BILIK PERBINCANGAN and KAREL TERTUTUP for year 2017 until 2023



In general, the patterns observed in reservations for both "BILIK PERBINCANGAN" and "KAREL TERTUTUP" exhibit fluctuations throughout the years 2017 to 2023. Both categories experience varying levels of activity, with some periods showing increases while others show decreases. Exceptionally, no reservation at all for both room types for the year 2021 due to restrictions on physical access to libraries.

UUM library administrators could implement a more flexible reservation system that accommodates fluctuations in room bookings. This could involve allowing for shorter notice periods for reservations or implementing dynamic pricing strategies to incentivize bookings during periods of low demand. To mitigate the impact of fluctuations in room reservations, the library administrators could enhance communication with users to raise awareness of room availability and encourage bookings during quieter periods. This could include promoting the benefits of using library facilities for group discussions or private study sessions. During periods of low room reservations, the library administrators could explore alternative uses for these spaces. For example, they could repurpose the rooms for events, workshops, or collaborative activities that align with the library's objectives and user needs. Recognizing the impact of physical access restrictions on room reservations, the library administrators could invest in virtual collaboration tools and platforms. This would enable users to engage in group discussions or collaborative work remotely, providing an alternative solution during periods when physical access to library spaces is limited. The administrators may need to review existing reservation policies and procedures to ensure they remain responsive to changing user needs and circumstances. This could

involve revising cancellation policies, extending reservation durations, or introducing tiered reservation systems to accommodate varying levels of demand. Learning from the experience of no reservations in 2021 due to physical access restrictions, the administrators could develop contingency plans and protocols to manage similar situations in the future. This could involve establishing guidelines for remote access to library resources and services, ensuring continuity of support for users during challenging circumstances.

5. How does the number of books borrowed from the library relate to the frequency of room discussion reservations over time?

To respond to this research question, a data frame named *loanAppt* was created by combining information from *loanbook* and *appointment* dataset. The information of *loanbook* dataset consists of the total number of book loans made by summing up columns 3 to 11, while the information of *appointment* dataset consists of the total number of room reservations by summing up columns 3 and 4. In addition, the column of *Tahun* and *Bulan* of *loanbook* dataset were also included in the *loanAppt* data frame. Figure 14 shows the codes for this step. In the second line, the code *loanAppt* will display the content of the *loanAppt* variable, as shown in Figure 15, showing the result of the data frame with 4 columns and 84 rows. Column 3 consists of the sum of overall sum of loan books for UG and PG students, while column 4 contain the total of both room reservations (BILIK PERBINCANGAN and KAREL TERTUTUP).

Figure 14

R Codes for Creating loanAppt Data Frame

```
loanAppt<-data.frame(Tahun=loanbook$Tahun, Bulan=loanbook$Bulan,
StudentLoan=rowSums(loanbook[,c(3:11)]), RoomRes=rowSums(appointment[,c(3,4)]))
loanAppt
```

Figure 15

R Result of loanAppt Data Frame

```
> loanAppt<-data.frame(Tahun=loanbook$Tahun, Bulan=loanbook$Bulan, StudentLoan=rowSums(loanbook[,c(3:11)]), RoomRes=rowSums(appointment[,c(3,4)]))
> loanAppt
   Tahun Bulan StudentLoan RoomRes
1   2017   Jan         713     112
2   2017   Feb        5240     114
3   2017   Mac        6961     181
4   2017   Apr        3944     171
5   2017   Mei        2736     454
6   2017   Jun         828     184
7   2017   Jul        1057     147
8   2017  Ogos         791     170
9   2017  Sept        4506     290
10  2017   Okt        5738     566
11  2017   Nov        4481     649
12  2017   Dis        2445     777
13  2018   Jan         823     170
14  2018   Feb        1723     116
15  2018   Mac        5742     448
16  2018   Apr        3272     199
17  2018   Mei        1665     335
18  2018   Jun         344      73
19  2018   Jul         321      67
20  2018  Ogos         343      69
21  2018  Sept        3054     196
22  2018   Okt        3839     438
23  2018   Nov        3361       0
24  2018   Dis        1694     396

 ..............................
77  2023   Mei         893     204
78  2023   Jun         506     177
79  2023   Jul         428     211
80  2023  Ogos         270      91
81  2023  Sept         233       4
82  2023   Okt        1033      60
83  2023   Nov        1325     188
84  2023   Dis        2156      47
```

Next, we calculated the covariate (covariance) and correlation to assess the strength and direction of the relationship between the number of books borrowed and the frequency of room reservations as shown in Figure 16. A positive covariance of 191636.7 suggests that as the number of books borrowed increases,

the frequency of room reservations also tends to increase, while value 0.57 of correlation indicates a positive linear relationship between these two variables.

Figure 16

R Code and Result of Covariance and Correlation between Number of Books Borrowed with Frequency of Room Reservations

```
cov(loanAppt[3],loanAppt[4])
               RoomRes
StudentLoan 191636.7
cor(loanAppt[3],loanAppt[4])
               RoomRes
StudentLoan 0.570006
```

To address this research question, we also carried out simple linear regression models to describe the effect of total number of books borrowed might have on the frequency of room reservations. Figure 17 shows the codes and results on performing the simple linear regression using *lm()* function and viewing the report summary of the regression model using *summary()* function. Overall, the result in Figure 17 provides information about the coefficients, significance of the predictors, goodness of fit, and overall significance of the model. It suggests that the frequency of room reservations has a statistically significant effect on the total number of books borrowed, as indicated by the low p-value (1.519e-08).

Figure 17

R Code and Result of Simple Linear Regression Models between Number of Books Borrowed with Frequency of Room Reservations

```
model <- lm(loanAppt$StudentLoan ~ loanAppt$RoomRes, data = loanAppt )
summary(model)

Call:
lm(formula = loanAppt$StudentLoan ~ loanAppt$RoomRes, data = loanAppt)

Residuals:
    Min      1Q  Median      3Q     Max
-3177.3  -555.4  -471.2    30.2  5581.7

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       601.548    182.914   3.289  0.00148 **
loanAppt$RoomRes    4.297      0.684   6.282 1.52e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1316 on 82 degrees of freedom
Multiple R-squared:  0.3249,      Adjusted R-squared:  0.3167
F-statistic: 39.46 on 1 and 82 DF,  p-value: 1.519e-08
```

In addition, we plot the simple linear regression models using *plot()* function and *abline()* function that will draw the regression line as shown in Figure 18. The result of scatter plot and regression line (Figure 19) shows an upward line slope which suggests a positive correlation between the number of rooms reserved and the number of books borrowed.

Figure 18

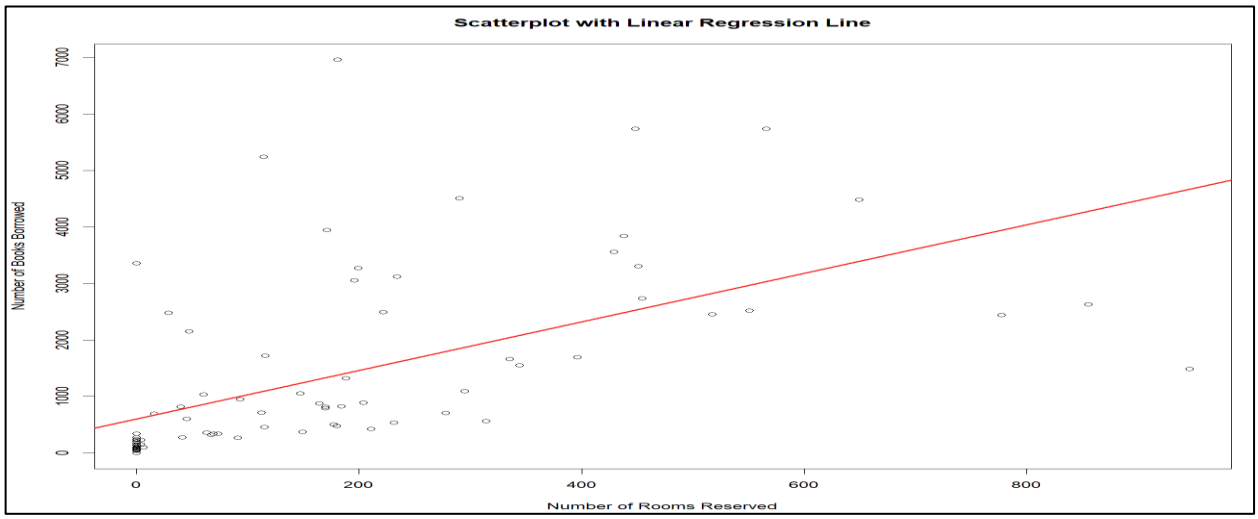
R Code for Plotting Simple Linear Regression Models

```
plot(loanAppt$RoomRes, loanAppt$StudentLoan, main = "Scatterplot with Linear
Regression Line", xlab = "Number of Rooms Reserved", ylab = "Number of Books Borrowed")

abline(model, col = "red", lwd=2)
```

Figure 19

Scatterplot and Regression Line between Number of Books Borrowed with Frequency of Room Reservations



The library administrators could consider allocating resources based on the observed correlation. If there is a significant increase in the number of books borrowed corresponding to an increase in room reservations, the library may need to ensure a sufficient stock of materials to meet the anticipated demand. Recognizing the positive correlation, the administrators could optimize the utilization of library spaces to accommodate both study and collaborative activities. This could involve prioritizing room reservations for group discussions or study sessions during peak borrowing periods to enhance the overall user experience. The administrators could implement targeted promotional activities to encourage room reservations and borrowing activities concurrently. For example, they could offer incentives or discounts for users who reserve rooms for group study sessions or provide recommendations for relevant resources available for borrowing. Continuously monitoring and analyzing the correlation between room reservations and book borrowings allows the administrators to make informed decisions regarding resource allocation, service provision, and strategic planning. Regular assessment of these patterns enables the administrators to adapt and optimize library services to better meet the evolving needs of library users. Leveraging the positive correlation, the administrators could enhance the overall user experience by providing integrated services that facilitate both room reservations and access to relevant library resources. This could involve developing user-friendly platforms or mobile applications that streamline the process of booking rooms and borrowing books. Recognizing the potential for collaboration between room reservations and borrowing activities, the administrators could explore partnerships with academic departments or student organizations to promote joint initiatives that utilize library spaces and resources effectively.

CONCLUSION

In conclusion, through a thorough examination of annual book loan rates across different academic years and between undergraduate and postgraduate students, as well as an analysis of room discussion booking frequencies over time, this study has provided valuable insights into the dynamics of library usage within academic settings. The findings reveal variations in books borrowing and room reservation patterns across different academic years, indicating potential shifts in student preferences, academic demands, and resource utilization trends. Additionally, the study highlights correlations between the number of books borrowed and the frequency of room discussion reservations, suggesting interdependencies between these two aspects of library usage. Moving forward, further research avenues offer opportunities to deepen our understanding of these dynamics and inform strategies for optimizing library services and facilities to better meet the evolving needs of users. Future direction could include developing predictive models to forecast future fluctuations in room types of reservation frequencies between different academic years, conducting a behavioral analysis to understand how the number of books borrowed from the library relates to the frequency of room discussion reservations over time and finally, investigating the impact of technology, such as online reservation systems or digital resource access, on borrowing behavior and room reservation frequencies over time.

ACKNOWLEDGMENT

The authors gratefully acknowledge the data contribution from Perpustakaan Sultanah Bahiyah, Library of Universiti Utara Malaysia. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Chris, N. (2020). *Exploratory Data Analysis and Point Process Modeling of Amateur Radio Spots*. http://www.reversebeacon.net/,

Finch, J. L., & Flenner, A. R. (2016). Using data visualization to examine an academic library collection. *College and Research Libraries, 77*(6), 765–778. https://doi.org/10.5860/crl.77.6.765

Islam, R., Tanweer, S., Nafis, T., Hussain, I., Qurashi, S. N., & Azimi, J. (2023). Unveiling The Hidden Secrets: Exploring Cardiovascular Disease With Data Analysis Using R Programming Language. *Community Practitioner, 20*(10), 263–277. https://doi.org/10.5281/zenodo.10065888

Konopka, B. M., Lwow, F., Owczarz, M., & Łaczmański, Ł. (2018). Exploratory data analysis of a clinical study group: Development of a procedure for exploring multidimensional data. *PLoS ONE, 13*(8). https://doi.org/10.1371/journal.pone.0201950

Lund, B. D. (2020). Assessing library topics using sentiment analysis in R: a discussion and code sample. *Public Services Quarterly, 16*(2), 112–123. https://doi.org/10.1080/15228959.2020.1731402

Markazi-Moghaddam, N., Jame, S. Z. B., & Tofighi, E. (2020). Evaluating patient flow in the operating theater: An exploratory data analysis of length of stay components. *Informatics in Medicine Unlocked, 19*. https://doi.org/10.1016/j.imu.2020.100354

Mestiri, S. (2019). *How to use the R software*. MPRA Paper No. 119428. https://mpra.ub.uni-muenchen.de/119428/

Pauline Adeniran. (2011). User satisfaction with academic libraries services: Academic staff and students perspectives. *International Journal of Library and Information Science, 3*(10). https://doi.org/10.5897/ijlis11.045

Sichilongo, K., Padiso, T., & Turner, Q. (2020). AMDIS-Metab R data manipulation for the geographical and floral differentiation of selected honeys from Zambia and Botswana based on volatile chemical

compositions using SPME–GC–MS. *European Food Research and Technology, 246*(8), 1679–1690. https://doi.org/10.1007/s00217-020-03523-x

Vignesh, K., & Nagaraj, P. (2022). Analysing the Nutritional Facts in Mc. Donald's Menu Items Using Exploratory Data Analysis in R (pp. 573–583). https://doi.org/10.1007/978-3-031-07012-9_48

Wickham, H. (2016). Getting Started with ggplot2. *In: ggplot2. Use R!*. Springer, Cham. https://doi.org/10.1007/978-3-319-24277-4_2