

STAT 455 – DATA COLLECTION AND SAMPLING METHOD

RESEARCH TOPIC:

*‘TO IDENTIFY AND COMPARE THE CAPACITY OF UNIVERSITY OF
CANTERBURY CENTRAL LIBRARY A WEEK BEFORE SEMESTER BREAK
AND A WEEK DURING SEMESTER BREAK’*

HARIZ AQIL BIN ABD GHANI (47725827)

1.0 Introduction

Data collection is the most important part of any research, and it is integrated to the entire study. To undergo an unbiased and reliable research, there are many methods to conduct using thoroughly analysis and conclusion.

For this project, we conducted an observation and collected the data to gain insights and draw the conclusions. From various methods of collecting the data, in this project, we used observation of 10 hours per person when conducting the research.

2.0 Survey Objective

Our primary aim is to ***identify and compare the capacity of the UC library before the semester break and a week during the semester break***, with the purpose of gaining insights into the factors that affect library capacity during different periods. The survey will seek to collect data on the number of occupied spaces in the library, the frequency and duration of visits, the type of resources used, and the overall data collector's experience during each period. The objective is to provide a comprehensive analysis of the data collected, highlighting any significant differences between two periods, and drawing meaningful conclusions.

During initial discussion, we consider research that it is easy to collect, easy to access thus the research must be specific, measurable, achievable, relevant, and time-bound (SMART). A clear and concise objective is vital ensure the observation is focused, and data collection is meaningful.

In this research, the population is every possible student inside UC central library within two period mentioned above (a week before semester break and a week during semester break), and the subpopulation is the number of students at different floors. Therefore, we sampled the data by randomizing floor levels in the library.

3.0 Methodology

We decided to collect observation data, number of occupied spaces in the UC central library and not survey questionnaire. It is easier to gain data and sorting the data using observation count. In the UC central library, there are 11 floors in total which are level 2, level 3, level 4, level 6, level 7, level 8, level 9, level 10, level 11.

In our discussion, we separated the busy floor which are level 2, level 3, and level 4, meanwhile non-busy floor are the silent floors, level 6, level 7, level 8, level 9, level 10, level 11. We used simple random sampling (SRS) method, which we randomly pick one floor from busy floor and one floor from the non-busy floor, at least. 5th floor was excluded in the research as it is not accessible for student, and it is only for UC staff. Another variable that we randomize is the time interval for data collection. Instead of assigning each person to specific time frame, we randomize the time frame, so it is more random.

The observation count was taken over a period of two weeks with the first week is a week before semester break and a week during semester break. To make it simpler in the explanation, the week before semester break is named to P1, whereas the week during the semester break is named to P2. From all the floors in the UC library, we excluded the discussion rooms to make it easier to collect observation count. Supposedly, as the group member of 3, we have the total 30 hours, 15 hours for P1, the first week and 15 hours for P2 which is second week. We assigned evenly 15 hours for 3 people in one week.

Busy	11 - 3 pm
non-busy period	other time exclude busy period

During our data collection process, we distinguished between two main time frames: a busy period, which spanned from 11am until 3pm, and a non-busy period, encompassing all other times during opening hours. To facilitate data recording, we initially divided the day into four distinct categories – morning, noon, evening, and night – before further consolidating them into the broader categories of busy versus non-busy periods. This approach allowed us to accurately capture and compare library occupied space patterns during different periods of the day and draw meaningful insights from the data collected.

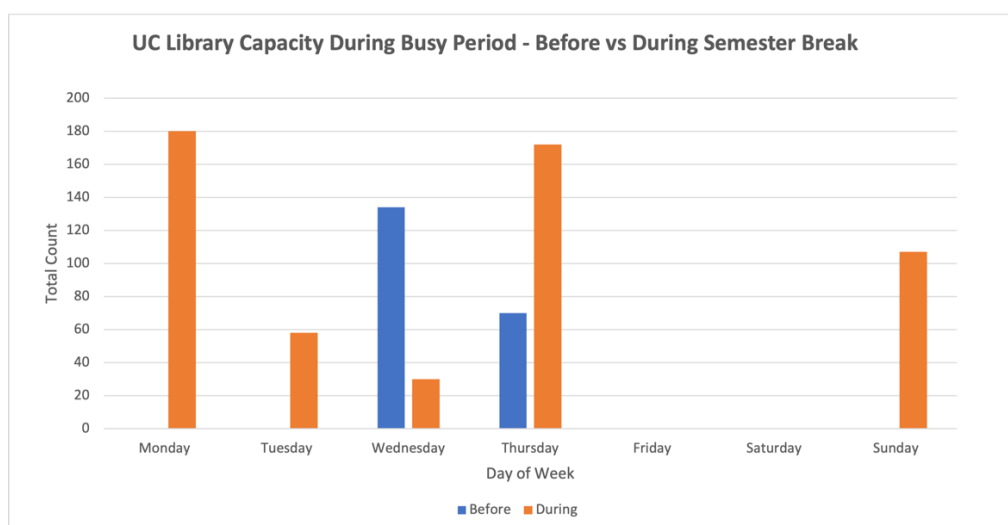
Similarly with other two data collectors in the group, we choose our own time frame within specified day. For example, I was assigned for Thursday during P1. I choose to collect the data on evening time, which is non-busy period. It is however due to weather condition and car park availability in the university since I did not have class that day. Although, this does not affect the way the count was taken but it could be biased since I purposely choose evening time frame due to traffic and weather.

4.0 Data Quality and Data Analysis

Data quality is a crucial aspect of any analysis, as it directly impacts the accuracy and reliability of the conclusions drawn from the data. It encompasses a range of factors, including accuracy, consistency, completeness, and validity of the data, among others. High-quality data is essential for making a valid and reliable conclusion that can inform decision-making and drive positive outcomes.

In contrast, data analysis is the process of examining and interpreting data to identify patterns, trends, and relationships. It involves using various analytical and statistical techniques to extract meaningful insights from the data and draw conclusions. Data analysis is a critical step in any research project, as it helps to provide a deeper understanding of the data and its implications.

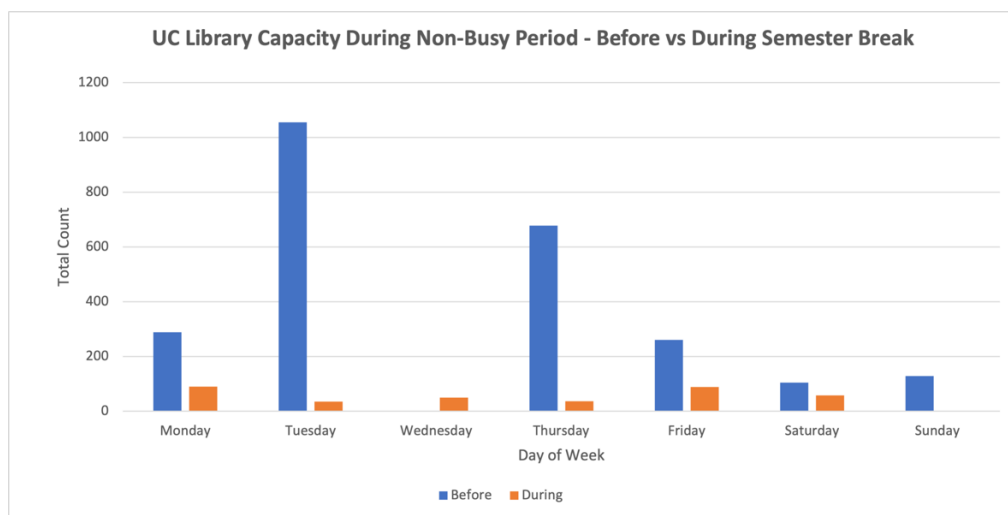
4.1 UC Library Capacity during Busy Period – P1 vs P2



Graph 1

From Graph 1 above, we observed that some data were not included in the analysis. Specifically, during P1, the observation count was not taken on Friday and Saturday due to the Easter public holiday. Another Friday and Saturday were counted again in the following week, within the semester break period. Given the randomness of the time frames when the data was collected, we cannot compare Monday and Tuesday for P1 and P2 in details. However, we can compare Monday and Tuesday roughly, and we identified a slight decrease in the number of occupied spaces in the UC library. These findings provide valuable insights into library usage patterns for P1 and P2.

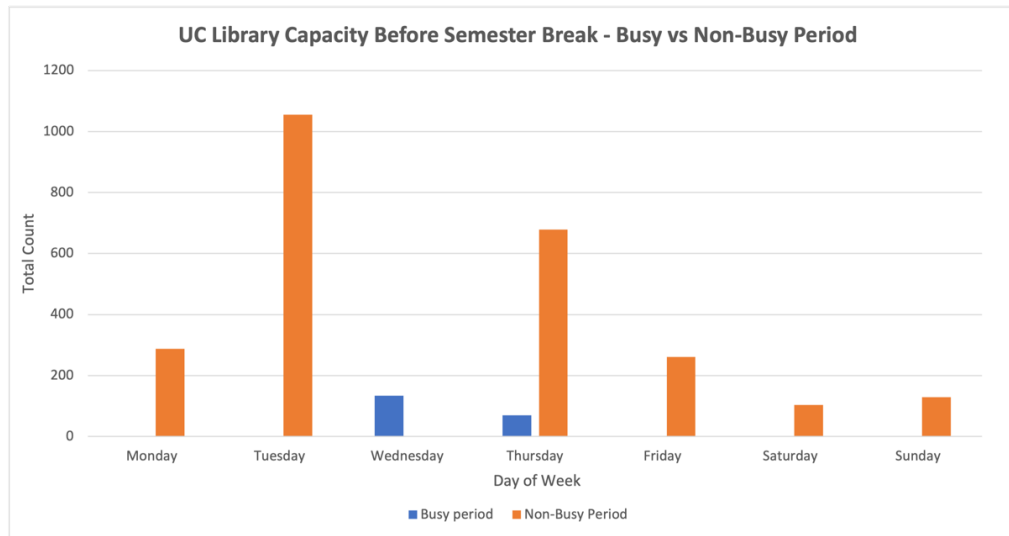
4.2 UC Library Capacity during Non-Busy Period – P1 vs P2



Graph 2

Upon analyzing Graph 2, we observed a more intriguing pattern compared to the previous graph. Given that the majority of observation counts were collected during non-busy periods, we can conduct a more thorough comparison with this data. Notably, we identified a declining trend from weekdays to weekends, except for a slight growth on Sundays for P1. As for P2, we found a slight increase from Tuesday until Friday, followed by a slight drop. These findings provide valuable insights into library usage patterns.

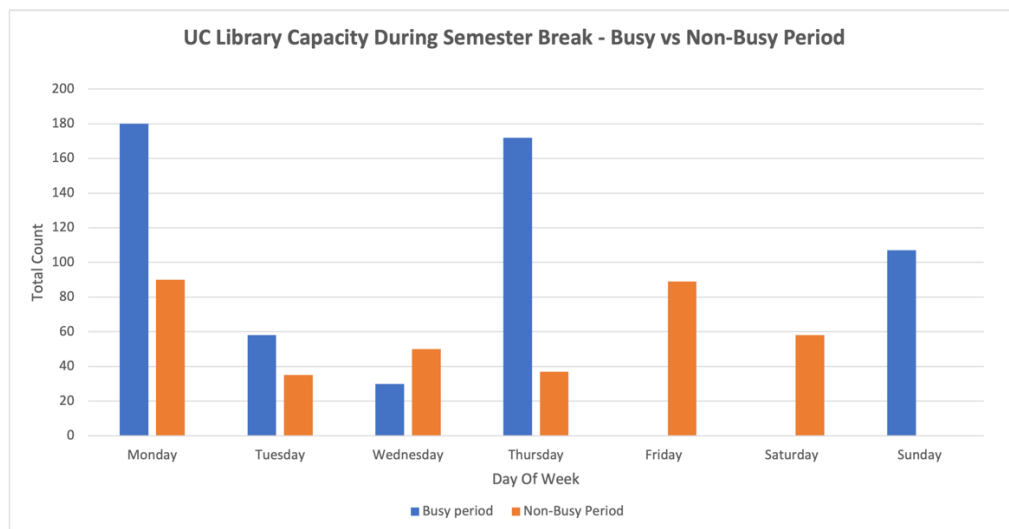
4.3 UC Library Capacity during P1 – Busy vs Non-Busy Period



Graph 3

Graph 3 above displays the number of observations counts during the busy and non-busy periods at the UC library. It is worth noting that only two days of data were collected during the busy period, which were on Wednesday and Thursday. Despite this, we observed a slight downward trend in the number of observations counts during this period. Conversely, it was easier to analyze the trend during non-busy periods. Specifically, we observed a sharp increase from Monday to Tuesday, followed by a gradual decline until the weekend. These insights are useful in understanding library usage patterns.

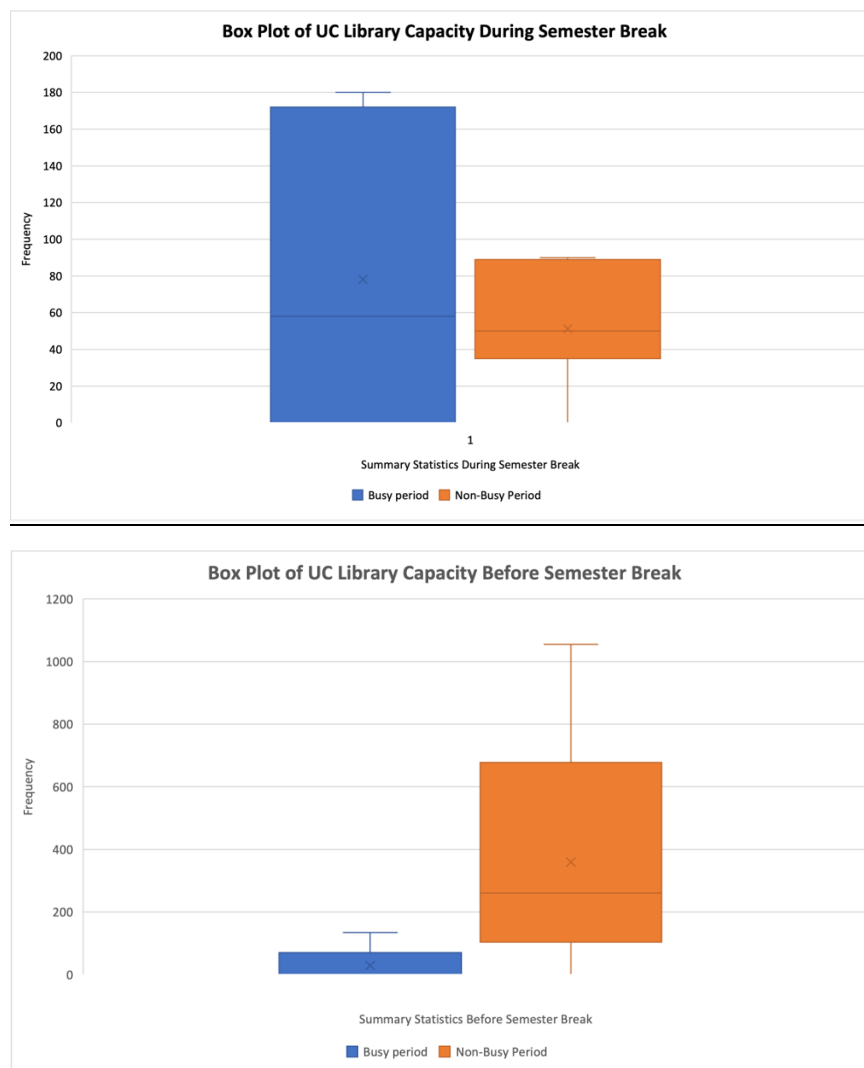
4.4 UC Library Capacity during P2 – Busy vs Non-Busy Period

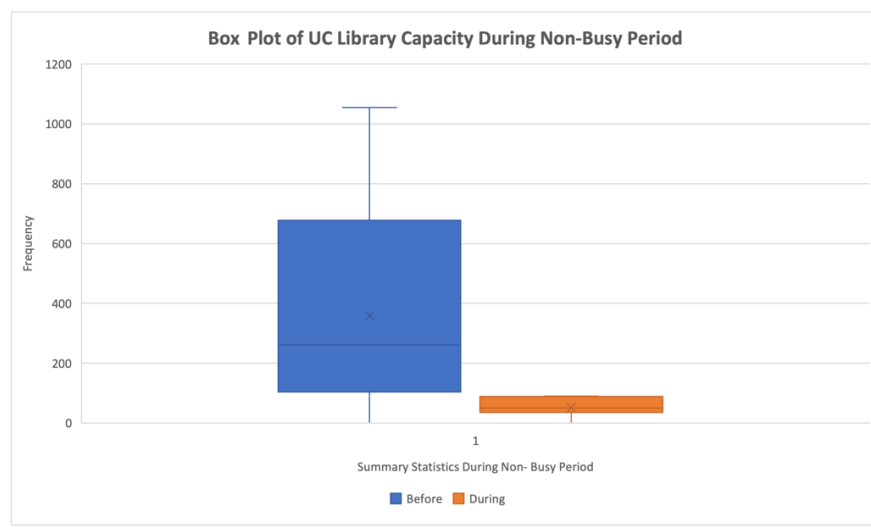
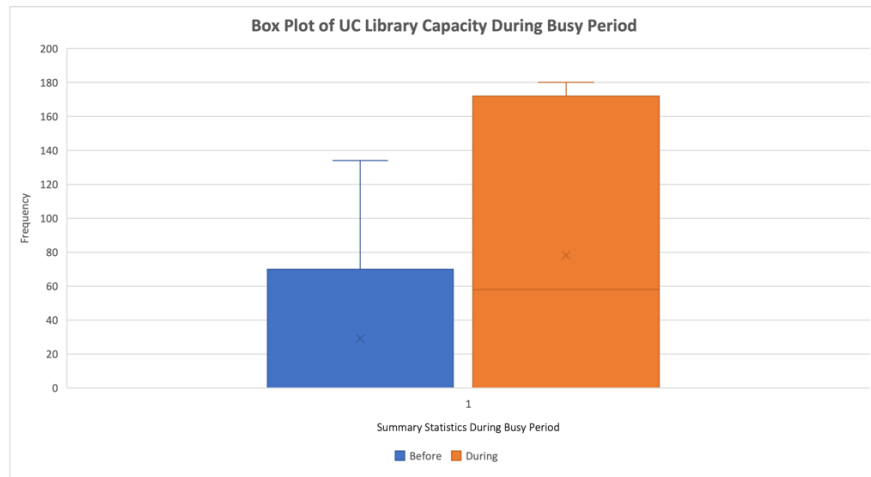


Graph 4

Graph 4 presents interesting findings, particularly with regards to the trend during the busy period. We observed an unexpected total count for Monday and Thursday, which was significantly different from that of Tuesday and Wednesday. Upon further analysis, we found that this was due to the random floors that were taken between these four days. It is possible that the data collected for Monday and Thursday were from the busier floors compared to those of Tuesday and Wednesday. These findings highlight the importance of considering various factors that may influence data collection and analysis and ensuring that the data used is representative and accurate.

4.5 Boxplot of UC Library Capacity during P1 versus P2 and UC Library Capacity during Busy versus Non-Busy Period





5.0 Discussion

This section aims to explain the challenges encountered during the observation process. One of the main difficulties was the randomness in the time frames chosen by each group member to collect data, as well as the different floor levels observed. This posed challenges in comparing and analyzing the data, as it was too random to draw clear conclusion. As evident from the graphs above, these challenges resulted in unusual patterns in the total counts, making it difficult to derive meaningful insights. To address such challenges in the future studies, it is important to standardize the observation process by establishing clear protocols for data collection, ensuring that the data is representative and selecting appropriate analytical tools for drawing reliable conclusions.

During the data collection process, I personally encountered issues with students leaving and occupying spaces in the UC library, which posed challenges in accurately collecting data. For instance, during the busy period, the 4th floor was usually the busiest, which made it time-consuming to collect the total count from start to finish, taking an average of 25 minutes. Additionally, while collecting data on occupied spaces from table to table, some students occupying the space left a few seconds after the count was taken, making it challenging to determine whether to include them in the data. To address this issue, I consulted with other group members, and we decided to include the data as it was collected, despite the potential for some inaccuracies. In future studies, it may be helpful to establish clear protocols for dealing with such challenges, such as a standardized approach to determine whether to include student who leave the space shortly after the count is taken.

In addition, we implemented a structured approach to assigning data collection hours to ensure that each member had specific hours and days to collect data. However, due to the low occupancy rate during P2, I found myself counting the entire population of the library during my assigned hours on Monday and Tuesday. Despite this, we made an effort to choose random floors during group data collection sessions to ensure the randomness and accuracy of the data, even though we were aware of the total population count during those specific time frames.

6.0 Conclusion

In conclusion, we can infer that during P1, the UC library had a higher capacity compared to P2, which was a week during the semester break. However, the randomness in the data collection process made it difficult to perform a detailed analysis of this trend. The unusual pattern in the data also hindered the use of comparative analysis methods. Despite the challenges faced during the data collection process, this research highlights the importance of collecting high-quality data to draw accurate conclusions.

To improve the reliability and validity of this research, it is essential to establish a structured protocol for data collection. This will help to address the limitations of the current approach, which suffered from deficiencies in terms of consistency and comparability. By creating a standardized method for data collection, we can ensure that all variables are accounted for,

and the data collected is more reliable and consistent. Additionally, this will help to minimized errors and inaccuracies in the data, allowing for a more robust analysis. Overall, a structured protocol is crucial for producing high-quality data, which is essential for meaningful analysis and insights.