

Objective:

Is it worth keeping UC Libraries open during the break?

1. Describe how you collected the data in enough detail that a group could replicate your data collection methodology in 2024 from your description. This includes the sample design as well as the actual process of data collection. It is recommended you keep a work diary as you go along for when you write your report. Some aspects you may cover.

The sample design was to survey three libraries, James Hight, EPS and Riccarton Library. The objective is to investigate if it is worth keeping the UC libraries open during break. We did the survey by recording the people going through the entrance and writing down their sex, race, group size and what they were doing at the library.

For my data collection, I collected my data in EPS. I would sit far away from the entrance and try not to look at them for a long time and write as quick as I could to not appear as a creep. Since EPS is a small library where not many people go and it was during term time, I can actually record them from the entrance and actually keep observing them what they were doing at the library e.g studying, chatting, eating and more. I would walk as close as to where they are to see what they were doing, pretend that I'm looking for books just so I can glance at what they are up to, or I would sit from away from them e.g. in the corner where I can see them. Since EPS is a small library, I had to be careful where to put myself as there is only so little spaces where I can actually keep my eye on them and not appearing as a creep at the same time.

2. Did you divide the survey population into subpopulations? If so why and how?

Yes we divided our survey population into subpopulation. We divided them into three populations: James Hight, EPS and Riccarton High School library (which is our control group). The reasoning behind this is to compare the population of people going to uni libraries and public libraries during term break, and use it to answer our objective.

3. How did you select your sample? Over space and time?

We all did 10 hours for one library. I did EPS and I selected my sample for every person walking through the entrance and keeping my eye on them during the time I was there

observing. I mostly did my survey for 1 hour each time and this was done over the first two weeks of the break (whenever I'm available).

4. How did you coordinate the sampling and measurement processes to get consistent results from all data collectors?

We formed a group chat to discuss potential demographics for our survey until we've all agreed to what will be included for the survey. We booked a discussion room, made a group spreadsheet for data collection during the break and discuss important topics like how to do the survey and what time and days we will do it.

5. Are there variations in the quality of the data?

Yes. I would say my data was more detailed than other data collectors since EPS is a small library and less popular. There were less people coming in so I can actually keep my eye on them and observed what those people were actually doing at the library like differentiating whether they are watching Netflix or Echo recordings.

For other data collectors, since James Hight and Riccarton Library are both popular libraries, more people will be coming in so there will be not enough time for collectors to actually follow and observe what they were doing at the library hence less detailed data.

6. Any consistency issues across data? – Consistency issues across data collectors? – Nonresponse issues?

Inconsistency issues across the data involved difficulty distinguishing ethnicities/races and inconsistency with the group size. I put only one observation and count how many is in the group rather than putting every person in that group as one observation. This in turn made me lose information on their sex and race, but I can still retain their purpose in the library. Also group size was difficult to count as there could be two groups arriving in different times forming one big group. Other data collectors have difficulty analysing what people were doing at the library as they could have multiple purposes at the library but not able to keep their eyes on them due to large volume of people coming in.

Inconsistency issues across data collectors will be our availability as we have different commitments (work, vacation and assessments during break) so it was hard to operate on similar times.

We had no nonresponse issues with each other. Everyone was replying and discussing when they are free. We also had no non response issues with the people we are collecting data as we are not talking or approaching them.

7. Effects on some data but not all data (e.g. it was pouring with rain one of the days)?

EPS only opens 1pm - 5 pm every weekends. So this cuts down my option as to what time when I could do my survey and cuts down the accuracy of the data as well because more people are more likely to go to the library in the afternoon as opposed to in the morning creating a little bias.

Furthermore, April 4 has plenty of observations since it was during the graduation ceremony so more people will be using the library. In addition there were more data a few days after the Easter holidays since students will be continuing their study and catching up in the library.

8. Explain any possible issues with selection probabilities (e.g. some types of people more likely be observed than others)

Since I did EPS, there weren't a lot of people coming in so everybody has an equal probability of being selected. However, I can think of possible issues with selection probabilities for other data collectors.

More likely to observe those arriving in group size as they have a larger presence and pretty easy to spot. More likely to observe those who dress up as they have an attracting presence. More likely to observe those who walk slow as collectors have more time to write down their observations. In contrast, less likely to observe those who walk fast. Imagine there is one person and there is a group both coming in at the same time, that one person will less likely to be observed as the priority would be the group.

9. How you structured your data file so if I gave it to someone else they could easily reuse it.

I used Excel spreadsheet so it can be imported and exported easily. I also put all of my data in one excel spreadsheet file so it can be accessed easily as opposed to having different files for different times. I also made sure to put the time and day so data is separated. All the functions and graphs I've made are also in that one file. I also made sure to write down some notes along the way e.g. what I did that day, what was unusual and how I managed to observe people without looking like a creep i.e. the tips and tricks.

10. I am not interested in anything more than very simple analyses. For example, simple comparison of means for various subgroups, sample totals (i.e. descriptive statistics) and graphs will suffice. Assume SRS for this (i.e. equal probability sampling). I will be impressed if you do a linear regression, but it won't contribute to your mark.. Do not attempt to calculate sample errors.

We have collected a total of 1750 observations in a span of three weeks in different days and times during the break.

- Name	Total Observations	Timespan	Sample Mean (people per hour)
- James Hight	- 740	- 9 am to 10 pm	- 56
- EPS	- 184	- 10 am to 5 pm	- 26
- Riccarton Library	- 826	- 10 am to 4 pm	- 138

James Hight has 740 from 9 am to 10 pm, EPS has 184 from 10 am to 5 pm and Riccarton Library has 826 from 10 am to 4 pm. The sample mean for each libraries are: 26 for EPS, 56 for James Hight and 138 for Riccarton Library. This means that Riccarton Library has the highest estimated number of people (138) using the library every hour whereas it's the opposite for EPS being the lowest (26). Due to EPS having a very low sample mean, this data suggests that EPS can be closed during the break as keeping it running for a few people would cost UC some money. It might be better to invest that money in student services/ facilities like more study pods and study areas for students which I will personally appreciate.

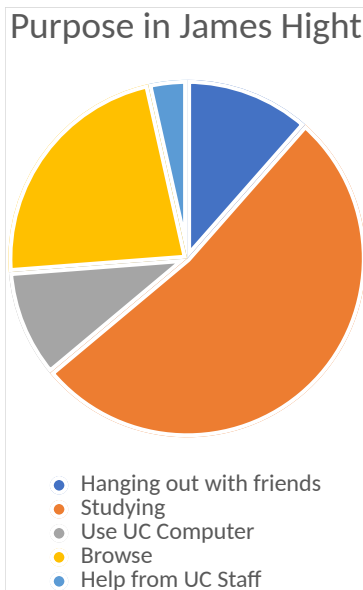
The data also shows that James Hight has estimated 56 people coming in every hour which is still a significant number. I don't think it would be wise for UC to close James Hight during the break because these estimated 56 people would then have to find new study places or libraries to use and most of the time that's not an option. Some people go to the library to study, to socialise, or to escape home and work and spend time with themselves. Since James Hight is the main library with 11 floors, the estimated 26 people from EPS can relocate to James Hight and they would have a space for themselves to do their study for example. Overall, James Hight can be kept open during the break.

These findings are coming from the data collected, but there are several factors involve in this. The data for Riccarton Library was collected during school time, so there were more high schoolers using the library. Also the data collected only spans from 10 am to 4 pm meaning the sample mean would be expected to be higher whereas James Hight's data spans from 9 am to 10 pm so it would have a lower mean. I think that if the time spans for

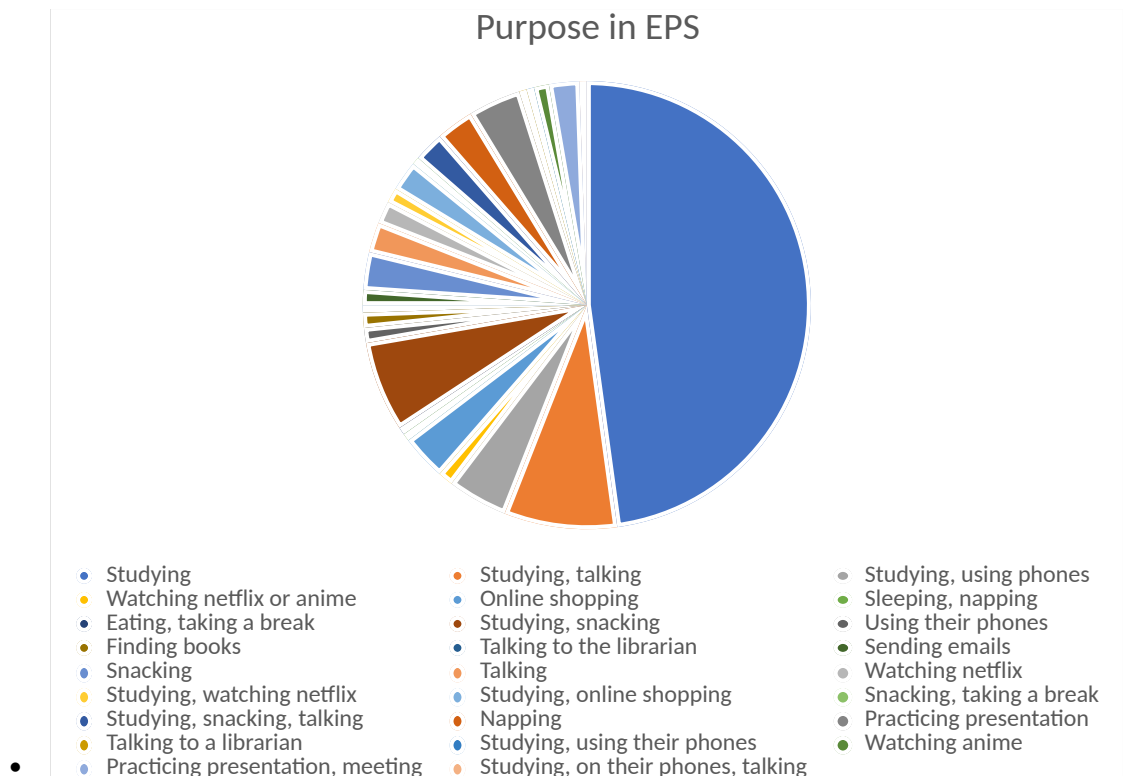
both libraries are the same, then we would have a different outcome. We may need more investigation on James Hight and EPS.

However, according to the data collected alone, James Hight can be kept open during the break as the estimated 56 people is still a significant number and it serves a lot of purpose for people.

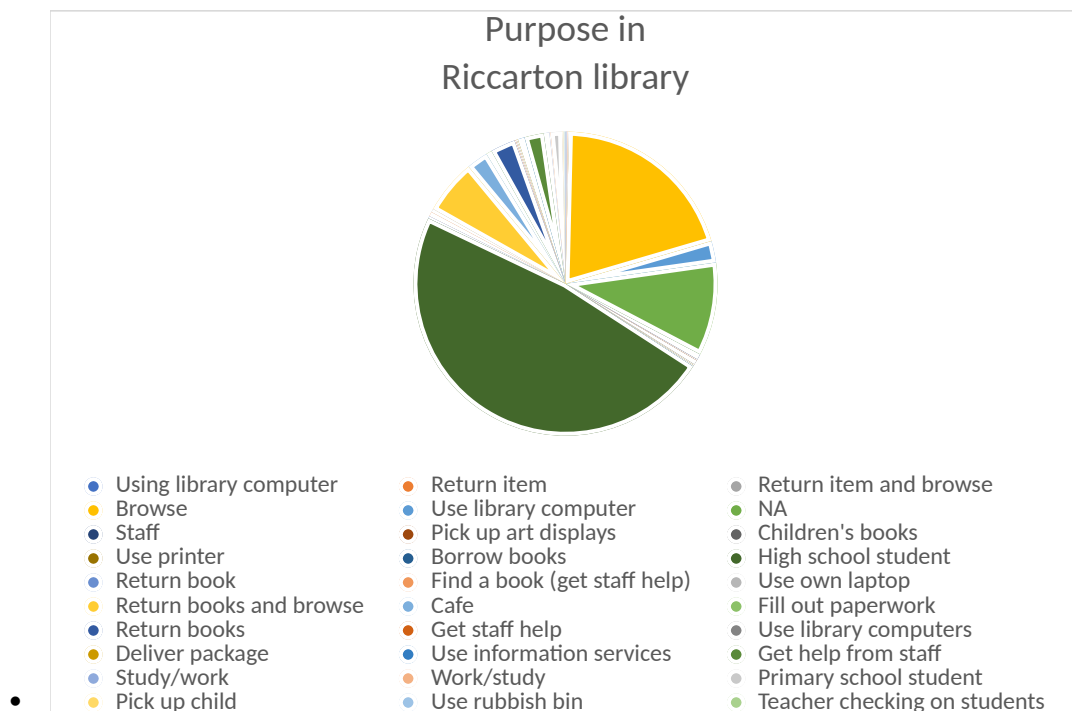
Talking about purposes, here's a breakdown of people's purposes for visiting the library.



James Hight was mostly used for study, browsing and computers during the break.



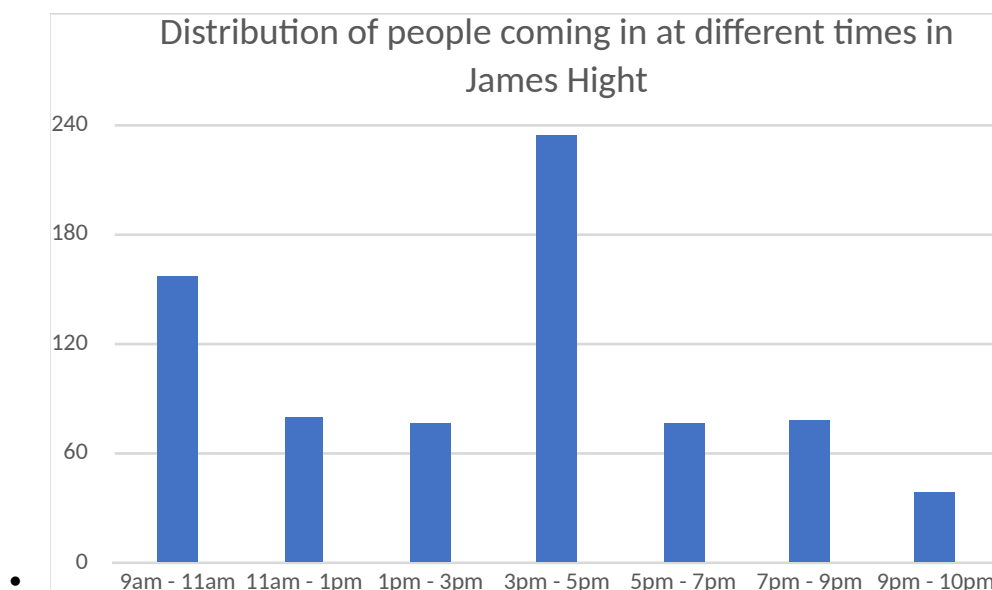
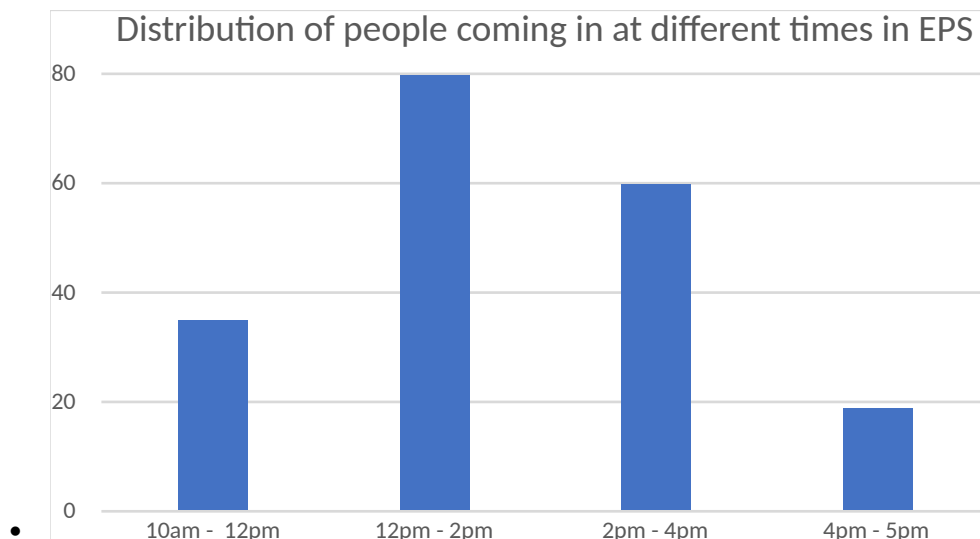
EPS was mostly used for study, study while talking and study while snacking.



Riccarton Library was mostly used by High school students. The data collector didn't specify what the students were doing exactly (as they were the only one to collect when there is a large volume of people coming in) but we can assume that most of the students were there to study as plenty of classes will be held in one of the big discussion rooms. I would know as I used to be a Riccarton High student.

All in all, this suggests that if Riccarton library can be kept open for study purposes, then EPS and James Hight can also be kept open especially for uni students trying to study.

However, EPS was only getting an estimate of 26 people per hour and it would be a loss of money to operate all the time during the break. We can look at the distribution of people coming in at different times for EPS and James Hight.



For EPS, the majority of people come in from 12pm – 4pm and it would make sense for EPS to only open at those times during the break. This way, UC can save electricity and not have to pay more workers to maintain and operate the library. In return, UC can save money and invest that into better/more student facilities/services. For James Hight, it looks about right to me. It can keep running its usual times.

Overall, James Hight can operate normally whereas EPS can be closed during the break or operate only from 12pm - 4pm to save money.

Some interesting demographics from our data. There were more females than males going into the library. More Caucasian in the data which is to be expected as most students at UC are Caucasian, followed by Asian. These findings also aligned with UC's official data. Photo below.

Ethnicity Priority	EFTS						Headcount					
	2018	2019	2020	2021	2022	2023	2018	2019	2020	2021	2022	2023
Asian	17.6%	18.1%	17.0%	14.7%	14.5%	15.7%	17.9%	18.6%	17.7%	14.9%	13.9%	15.5%
Maori	8.0%	8.2%	8.4%	9.4%	9.6%	9.5%	7.7%	8.1%	8.6%	9.8%	9.8%	9.5%
Middle East, Latin America & African	2.4%	2.4%	2.3%	2.2%	2.1%	2.3%	2.6%	2.6%	2.5%	2.3%	2.2%	2.3%
Not Stated	5.6%	5.3%	4.8%	3.7%	3.0%	2.8%	5.7%	5.4%	5.0%	3.9%	3.5%	3.1%
NZ European/Pakeha & Other European	61.8%	61.1%	62.9%	65.6%	66.5%	65.3%	61.0%	60.0%	61.4%	64.5%	66.1%	64.8%
Other	2.1%	2.2%	2.1%	1.7%	1.7%	1.8%	2.4%	2.6%	2.3%	1.8%	1.9%	2.1%
Pasifika	2.5%	2.6%	2.5%	2.6%	2.5%	2.7%	2.6%	2.7%	2.5%	2.7%	2.6%	2.8%

Ethnicity	EFTS						Headcount					
	2018	2019	2020	2021	2022	2023	2018	2019	2020	2021	2022	2023
Pasifika	2.9%	3.0%	2.8%	3.0%	2.9%	3.2%	3.0%	3.1%	2.8%	3.1%	2.9%	3.3%

Gender	EFTS						Headcount					
	2018	2019	2020	2021	2022	2023	2018	2019	2020	2021	2022	2023
Another	0.2%	0.3%	0.4%	0.9%	1.2%	1.2%	0.2%	0.3%	0.5%	0.8%	1.1%	1.2%
Female	49.8%	50.4%	51.2%	51.6%	52.1%	51.6%	50.7%	51.0%	51.5%	54.3%	57.7%	56.3%
Male	49.9%	49.3%	48.3%	47.5%	46.7%	47.1%	49.1%	48.7%	48.0%	44.8%	41.2%	42.5%

Most popular group size in the data is 1. Seems like people like to go to the library alone especially student for when they are studying. This is followed by group size 2 implying that some people also use the library to study while socialising with friends.

11. I would like you to discuss what is good about the data you collected and what are its deficiencies in terms of answering your question (e.g. how representative is it of what population? How good is the data at measuring your objective?). This is similar to what data collection agencies should do in advising possible users of their data.

The collected data is good because it compares two uni libraries with a public library (control group). It is a good representative of the population we were aiming for because we did the data collection during the break. The data is good at measuring our objective because it shows how many people use the library (for different purposes) and we can then see whether uni libraries are worth to keep open during the break (or only be opened at certain times)

In terms of deficiencies, the data was only collected during the break, we do not have data during the term so we can't compare to see if there's a notable difference.

12. Ideas for how the data collection process could have been done better based on what you learnt (and it can be done better next time!). You can include what you would do if you had more money or time or people, though ideas for improvements with what you had is desirable (i.e. if you knew then what you know now, ...).

If we have more people, then we could have two people doing the data collection at the same time so the large volume of people coming won't be an issue. One would be at the door and the other one would follow them to see what they were doing at the library to increase data accuracy. We would also collect before, during and after the break for better comparison and to answer the objective with a bit more evidence. We would also sample other libraries at uni (e.g. Macmillan Brown) for a better representation of the target population.

If we have more money, then we could take time off work to collect data for every day of the week and cover random, different times. We would also be able to rotate with each other and collect data from each libraries.

13. Any other information that you think would be useful or interesting. This is the section that is likely to lead to a very good grade as it is what I haven't explicitly asked for that I expect top students to supply

Riccarton library was only open during the first week of the break and was closed for the remaining weeks of the break because it had to undergo construction. This means that all data was only collected during the first week. And that it was during school week so more students will be using the library. This is useful to know because this creates bias within the data. We could've prevented this by collecting data when it is during high school break to keep it consistent with uni break, but we won't be able to due to time constraints and commitment to other things.

Another interesting thing was bias of the data collectors. As students, we have part time jobs during the break so we won't be able to just randomly survey whenever we can. We had to work around our schedule to make the data collection times as random as we can.

A treasurer or the person in charge of finance at UC would find this data useful. This data shows that not a lot of people go to EPS hence it would be a good idea to close it or open at certain times during the break to save money and to keep James Hight running normally. . This data would also be good for librarians as they can increase their staff to during busy hours.

