**Library report**

 **By Lucy Zhang**

There were 790 books in the SZ library in total. And there were 15 different kinds of books. In the original data, some of the books had no classification. So, I supplemented the serial number of some books without a classification number, and also classified the remaining books of unknown classification as Z comprehensive books [1]. As the result the percentages of each kinds of books in the library was graphed as following *(Figure 1)*:

*Figure 1 Percentages of each kinds of books.*

B philosophy, religion; C General Introduction to Social Sciences; D Politics, law F economy; G Culture, science, education, sports; H language, text; I Literature; K History, Geography; O Mathematical Sciences and Chemistry; P Astronomy, Earth Science; Q Biological Sciences; Z Comprehensive Books

This chart is used in order to show more clearly of each percentage of tiny parts of the book kind.

The biggest part of book kind is literature, followed by the Comprehensive Books, then Language & Text and History & Geography were the third large kind. The rest of them were the tiny parts that only took the percentage that less than 5%. The whole percentage of this small part was 11% of all.

The chart of the prices of books was similar to unimodal and right-skewed *(Figure 2)*.



*Figure 2 Book prices.*

The left part of the chart was closer to each other than the right. The data was most concentrated in ¥18~¥36. There were 17 books that is more expensive than ¥130. In the beginning I did not do anything to the max of the x-axis, and soon found out the chart had a long tail contains 1 or 2 books but cost much. In the end, I set overflow number to 130 and cut the long tail. In order to make the chart shorter and easy to read.

These books are the top 10 popular books in the library. All of them were borrowed more than 20 times *(Figure 3)*.

*Figure 3 Books of top 10 borrowing times.*

The library can display these books for longer times and get more books of these kinds.

Some books, such as Harry Potter series, its writer was well-known in people. But in the data, it shows 0. I changed the writer of Harry Potter from 0 to J.K.Rowling. I listed the top 10 popular writers of the books in the library. Among the books written by the top ten authors, the percentages of books written by each author were listed as following *(Figure 4)*:

*Figure 4 The top 10 popular writers.*

Jin Yong is the most popular author, Keigo Higashino is second, San Mao and Haohui Sun together are the third, Haruki Murakami is the fourth, Lu Yao and Lu Xun together are the sixth, Cixin Liu and DangNianMingYue is the seventh, and at last is Jiang Nan.

The library can get more books of these author if possible. It is also necessary to collect more than one book for the same title.

The data showed that female borrowed books more than male *(Figure 5)*.

*Figure 5 Gender of readers.*

There were 71% of the readers were female, which was 618 people. 29% of readers were male. Which was 254 people. Within this, The borrowing times of all the female was 2886 times, and male had 1627 times *(Figure 6)*.

*Figure 6 Borrowing times.*

But as we can see in the chart, the female readers were nearly 3 times of male readers, while female readers’ times of borrowing books were only 2 times of male readers’ and even less. As a matter of fact, the average of male’s borrowing time was actually more than female *(Figure 7).*

*Figure 7 Average of times of borrowing books.*

(average) Each female borrows books for 4.67 times while each male borrows books for 6.40 times.

There were 5 different kinds of people among the readers: Graduation, Senior three/secondary school three, Senior two/secondary school two, Senior one/secondary school one and Other. This is the contingency table and the bar chart of the population of them *(Figure 8 & Figure 9)*.



*Figure 8 Contingency table of population in different degrees.*

*Figure 9 Readers of different degrees and different gender.*

It is clear to see that female was much more than male, especially in Senior three/secondary school three. The readers were concentrated in Senior three/secondary school three.

Among the Senior one/secondary school one students (no matter female or male), they all borrow books of Literature *(Figure 10*).

*Figure 10 Percentage of book kinds in Senior one/secondary one students with data.*

Among the Senior two/secondary school two students (no matter female or male), they usually borrow books of Literature *(Figure 11)*.

*Figure 11 Percentage of book kinds in first 11 Senior two/secondary two students with* *data.*

This data comes from the original data base. I chose the first 50 Senior three/secondary school three students with data, and this is the percentages of the book they borrowed. Literature takes most of the places.

Among the Senior three/secondary school three students (no matter female or male), they usually borrow books of Literature *(Figure 12)*.

*Figure 12 Percentage of book kinds in first 50 Senior three/secondary three students with data.*

This data comes from the original data base. I chose the first 50 Senior three/secondary school three students with data, and this is the percentages of the book they borrowed. Literature is in the first place, then History & Geography is the second. The rest of them are Philosophy & Religion, Politics & Law, Astronomy & Earth Science and Biological Sciences.

Among the graduated students (no matter female or male), they usually borrow books of Literature *(Figure 13)*.

*Figure 13 Percentage of book kinds in first 10 graduated students with data.*

This data comes from the original data base. I chose the first 10 graduated students with data, and this is the percentages of the book they borrowed. And the biggest part is Literature.

Because the Other part only has 8 people and they are not students, the chart didn’t include them.

In conclusion, Literature kind of books are the most popular and necessary. The library should get more Literature books.

Different libraries have different amount of borrowing times of books *(Figure 14 & Figure 15)*.

*Figure 14 Borrowing times of different libraries.*

*Figure 15 Readers’ average borrowing times*

The total book borrowing times and the average from more to less are: LS – SN – JH – SZ – ZF – YL.

The libraries with more borrowing times should obtain more books, and the libraries with less borrowing times should obtain fewer books.

These are data of the amount of lending books from each library.

*(Figure 16~21)*

*Figure 16~21 Data of the amount of lending books from the six libraries.*

From every library between April 2019 to September 2020, they are multimodal and mostly concentrated in September 2019 to December 2020. There are gaps in July 2019 to August 2019, January 2019 to March 2020, and July 2020 to August 2020, is probably influenced by the summer vacation and the winter vacation (or COVID-19).

This data was found in the original data base. It is the first 30 lending time data in April 2019. It shows that most of the students borrow books from the SZ library in 15:00 to 22:00. Other libraries can refer this data to decide when to open the library in the busy time *(Figure 22)*.

*Figure 22 First 30 lending time data in April 2019 of SZ.*

In these gap time, the libraries can have a rest or open for few hours just in case. But in the busy time, the libraries need to open up for a longer time, for example, be opened from 15:00 to 22:00.

According to the boxplot, all five libraries has many outliers, and there were many concentrated outliers. The SN library has the greatest number of concentrated outliers

*(Figure 23~24).*





*Figure 23~24 Numbers of days of borrowing books.*

When cutting of the outliers, the boxplot of the number of days of borrowing books and is as following *(Figure 25*).



*Figure 25 the boxplot of the number of days of borrowing books*

The outliers of SZ is those days are greater than 17, SN and ZF and JH and YL’s are greater than 14, and LS’s are greater than 13. The widths of outliers of SN and JH are the widest, while LS has the fewest. It tells that the readers of SN and ZF libraries probably lost their books very often or were not good at timing.

After deleting the outliers, we can see that SZ has a wider spread of borrowing days and LS has a little shorter borrowing day.

***Action Plan:***

1. The library should obtain more books about Literature in order to fit the needs.
2. The library can display the top 10 popular books for longer times and get more books of these kinds.
3. The library can get more books of the top 10 popular author if possible. It is also necessary to collect more than one book for the same title.
4. The Literature kind of books are the most popular and necessary. The library should get more Literature books.
5. The libraries with more borrowing times should obtain more books, and the libraries with fewer borrowing times should obtain fewer books.
6. In these gap time, the libraries can have a rest or open for few hours just in case. But in the busy time, the libraries need to open up for a longer time, for example, be opened from 15:00 to 22:00.
7. It is better for all the libraries to give their readers more time to read book if possible, especially for the SN library. Otherwise, readers might not be able to finish the whole book in one time.

[1] https://baike.baidu.com/item/中国图书馆图书分类法/1919634?fr=aladdin

**Reflection**

I deeply understand how hard it is to analyze data. In the beginning, the library data is messy and hardly to tell anything. The process is hard. First to do is to find out what data we actually need. Then we have to classify the data, to see if it is quantitative or categorical. And making charts. But after analyzing, after drawing those graphs, things we want to know shown up. It's like bubbles are emerging.

During the COVID-19, to analyze the data of the virus must be much harder than analyzing this library data. But those statistics did it. Statistics is very important and significant in our lives. After facing these boring number, it displays the patterns, the logic, even predict the future. All these things are trying to help people live in a better way.

After analyzing the first five charts, I nearly want to give up even though I know I do have something that haven’t finished. It is a battle between comfortable and high score (even high score may not come true after analyzing the original data). As I mentioned above, it is a tough process to look for data in this bunch of charts. But in the end, GPA won. I still opened the original data base and try to analyze a little bit of them. 😐