Copyright Notice:

Commonwealth of Australia

Copyright Act 1968

Notice for paragraph 135ZXA (a) of the Copyright Act 1968

Warning

This material has been reproduced and communicated to you by or on behalf of Charles Sturt University under Part VB of the Copyright Act 1968 (the Act).

The material in this communication may be subject to copyright under the Act. Any further reproduction or communication of this material by you may be the subject of copyright protection under the Act.

Do not remove this notice.

Reading Description:


Reading Description Disclaimer:

(This reference information is provided as a guide only, and may not conform to the required referencing standards for your subject)
Collection Analysis: Powerful Ways to Collect, Analyze, and Present Your Data

By Amy Hart

As librarians, we continually perform collection analysis in our heads, evaluating our collections against the daily information needs of our students and faculty. This essential part of collection development is largely informal and undocumented.

It can be useful to quantify the strengths and weaknesses of a collection in a more formal analysis (see sidebars). The main obstacle to doing so, other than finding time, is determining a way to collect the data to analyze. If your library automation system includes a report module, you may be able to use existing reports. At Lexington Public Schools, we sent copies of our bibliographic data to Follett Library Resources, who produced a collection analysis for us. When we received their report, we realized we could use the data they provided to take the analysis further.

This article looks briefly at the kinds of data used in collection analysis and how to use library automation software to collect the data. It then describes the use of Microsoft Excel and its chart-making capabilities to enhance the presentation of the analysis.

Step One: Define Your Questions and Collect the Data

Counts and Percents of Titles Held in Each Dewey Class

The most basic question in a collection analysis is "Where is my collection weak, and where is it strong?" One way to determine this is simply by numbers. How many titles do we own in each of the major classifications of our collection?

Information on an item's Dewey class is available in the holdings database of your automation system. In a MARC-based system, it will be in the 852 field. Your automation system may have a report that sorts and then counts the number of titles in each Dewey class. Our system includes a report that sorts holdings by class. To get counts or percentages we would need to edit the report, or export the data to another application, or count by hand. However the data are collected, they need to be entered into a spreadsheet application as the next step.

Average Date of Publication

Another measure of collection strength is the relevance of the titles in each classification. A crude measure of relevance is date of publication. The average publication date for titles in each Dewey class gives you an idea of whether you need to update titles in specific classes.

The MARC bibliographic format contains year of publication data in both a fixed field and in the 260 $c publication date field. Reports may pull date information from either one of these fields. Your automation system may offer a report that will sort by Dewey class, sub-sort by publication date, and then calculate the average publication date. Our system offers a report that selects by Dewey range and sorts by publication date. To get the average dates, we would have to edit the report or massage the data in some other way. Again, the results need to be entered into a spreadsheet application. Some report modules allow report data to be exported in a delimited format so that they can be imported into other applications. This is the best way to move the information if the option is available.

In Lexington, data on the percentage and number of titles held in each Dewey class and the average date of publication for each were sorted, tallied, and averaged for our schools by Follett Library Resources.

Another option for collecting your data is to request a custom report from your library system vendor. Be aware that there may be a charge for this service.

Circulation Statistics

After we received Follett's collection analysis, we realized that we had more questions we wanted to ask. I entered all of Follett's data into Excel spreadsheets so that I could manipu-
late them myself. I also added another criterion to our collection analysis—circulation information. By comparing the composition of the collection (how many titles in each Dewey class) to circulation counts by Dewey class, we could learn where supply was meeting demand.

To collect data on circulation, we used several preformatted statistical reports from our automation system. The reports sorted circulation totals by Dewey hundreds and by collection codes (REF, B, FIC). After running the reports, I entered the data into an Excel worksheet.

Step Two: Analyze Your Data and Present Your Message

Once the data are in spreadsheets, it is time to begin the analysis. Go back to your original questions and determine how the data answer them. Then decide which answers or results you want to include in your analysis. Consider what kinds of charts will communicate these results most effectively. One chart type may obscure a point you are trying to make, while another may make it very clear.

In Lexington, our starting point for this process was the receipt of Follett's collection analysis report.

We looked at what they had done and thought about ways to expand on it. What follows is a review of some of the additional charts we made.

Follett used a pie chart to illustrate the percentage analysis by Dewey class and a bar graph to illustrate the age of the collection by Dewey class. We combined the two charts into one. Figure 1 shows a chart with two y-axes, combining a bar chart illustrating percentage breakdown by Dewey with a line graph showing average date of publication by Dewey. By presenting the information on one page, we eliminated the need to flip between charts. It became clear, for example, that even though the average date for the 000s was fairly recent, this had little affect on the overall age of the total collection, for the 000s made up less than one percent of the collection.

Next, our libraries coordinator requested a chart comparing the average age of each school's collection. I used a "line-column" Chart because, in addition to comparing average ages, I wanted to compare our average ages to the Massachusetts School Library and

Note on the Chart Examples

TRIVIA QUIZ

In what work of literature does each fictional school name appear?

1. Lowood
   Jane Eyre by Charlotte Bronte

2. Pencey
   The Catcher in the Rye
   by J. D. Salinger

3. Wayside
   Sideways Stories from Wayside School by Louis Sachar

4. Crunchem
   Matilda by Roald Dahl

5. Hogwarts
   Harry Potter and the Philosopher's Stone by J. K. Rowling

6. Prufrock
   The Austere Academy by Lemony Snicket

7. Gregory
   Harriet the Spy by Louise Fitzhugh

8. Harrison
   Nothing but the Truth by Avi

Chapter 3: The Library Media Specialist as Information Specialist 89
Elementary Schools—Comparative % of Collection by Dewey Hundreds

Figure 3

Total Circulation FY2002:
Total Volumes owned=135,000; Total Annual Circ = 118,515; Total Enrollment = 6,286; Per Capita Loans = 19

Lowood 30.00% 17998
Pencey 0.00% 7161
Wayside 4.20% 13329
Crunchem 15000
Hogwarts 6.00% 13471
Wyverley 6.10% 122988

Step Three: Add Circulation Statistics to the Mix

The first graph using circulation data was a pie chart showing total circulation for each school Figure 4. To add impact to the chart, total volumes owned by our libraries, district-wide total circulation, total enrollment, and per capita loans information was added. In FY 2002, our school libraries averaged 19 loans per student, and we loaned almost as many books as we owned.

To compare collection composition and circulation within each school, I decided to use the actual counts of titles in each Dewey class rather than a percentage. This allowed me to use a single y-axis in a “columns with depth” bar chart Figure 5. In the 500s, 800s and FIC collections, circulation appears to correlate with numbers of holdings. This finding suggests that our librarians are successfully meeting the reading needs of their students in these areas. However, the chart also identifies subject areas where the collection or circulation is weak. For example, the 600s and 700s have substantially more circulation than holdings. Should holdings in these subject areas be increased and

Media Association (MSLMA) draft 2002 standard. This standard recommends that, “Seventy percent (70 percent) of the entire print collection will have a copyright date no later than ten (10) years from the current year” (Massachusetts School Library Media Association State Standards, <www.mslma.org/whoweare/mslmastandards.html>). Based on this standard, our collections are fairly outdated. Adding a line to the graph visually illustrated the standard and added emphasis to the message Figure 2.

Libraries also wondered how their percentage breakdowns compared to each other. To answer this question, I created a bar graph that showed the compositions of the six elementary libraries compared to each other and to the district averages. Viewing seven bars next to each other, we could see that the elementary collections are remarkably similar to each other Figure 3. A comparison among the secondary schools showed more variation.
updated? Likewise, does the library need more books in the Easy classification? In contrast, the 300s, 900s and Biographies have lower circulation than holdings. Which curriculum areas should be using these books? Is there a way to encourage their use? Might a library unit be planned in these subject areas? Or is this a case where the titles owned, while adequate in number, need to be reviewed manually for relevance?

A second chart compared age of collection with circulation totals. Because age is expressed in years and circulation in integers, a double y-axis was necessary Figure 6. A “lines on two axes” chart was used to facilitate comparison of the two variables. If there were a correlation between age of collection and circulation, the two lines would either mimic each other or go up and down in inverse proportions. The lines go up and down independently of each other, implying that there is little relationship between the two.

We gained some interesting insights by comparing the circulation and composition charts between schools. Figure 7 illustrates one way two schools can be compared. Lines are used to compare circulation, and bars are used to compare holdings totals. Only one y-axis is necessary.

Comparing libraries can be helpful when they have made different collection development choices (e.g., one concentrates on fiction during difficult budgetary times while the other tries to buy a little in all areas).

Conclusion
A formal collection analysis can be quite useful. If your library automation system includes the right reports, data collection can be as simple as submitting them to run. If not, it is still possible to use your reports to collect data that can be massaged in other ways. Once your data are assembled, you can use spreadsheet applications and chart making abilities to define and fine tune your analysis. The same set of data can be used to convey several different points, depending on how the information is presented. Design charts and graphs with specific audiences and messages in mind and select chart types to enhance your message.

Amy Hart is the Technical Services Librarian at Lexington (Massachusetts) Public School.