Geovisualizing Collections of Penn State University Libraries:
a geographical and statistical perspective of use, age, and relevancy

Sherry Steele Roth
Penn State University

Abstract

Academic libraries find themselves in a transformational environment from curating and marketing invaluable collections of intellectual and cultural achievement to facilitating collaborative teaching and learning in technology-rich spaces. Libraries must identify and market to administrators, patrons, and donors their relevancy as good stewards of library assets. Using geographic information systems (GIS) to visualize collection characteristics helps libraries find the right balance of book and patron occupancy, analyze collections and use trends, identify collection strengths and weaknesses, and allocate space effectively and economically.

Background

Academic libraries have been undergoing a transformation from being viewed as a place that houses an “almost incalculable and certainly irreplaceable” collection of “intellectual and cultural achievement (Coyle)” to being viewed as a place for “teaching and learning spaces and enhanced user seating (“Framework for Space Planning – MIT”, 2012). The challenge for academic libraries becomes how to balance the continuing "investment in assets for ‘just in case [information]’ at a time when the emerging paradigm for allocating resources for information is ‘just in time’ (Smith, 2004)." Even though accrediting bodies and professional organizations have used actual volume counts to evaluate and rank academic libraries, they are realizing that it “is no longer appropriate to treat most print resources as protected objects, or the college library as a museum for books (Tracy, 2011).” Consequently, university librarians need to perform a formidable balancing act: protecting tangible intellectual assets that cost overhead dollars against freeing the space occupied by these assets for alternative, but much needed, intellectual pursuits – teaching, learning, and collaboration. Some of the tangible assets can be seen Figure 1A.

Figure 1. (A) The stacks of central Pattee Library before moving 100,000 inches of infrequently-used books to off-site, high-density annex shelving in 2012. From Collection Maintenance Facebook (https://www.facebook.com/StacksAtPennState). **(B)** Infrequently used books moved off-site allow for alternative space use like the Knowledge Commons on the first floor of Pattee Library, opening in 2012 with computer workstations, group study rooms, a one-button studio, media-rich teaching spaces, and recently a 3-D printing lab. Photo from Knowledge Commons Facebook (https://www.facebook.com/psukc).

Libraries are now re-evaluating their roles and optimizing their spaces to find the right balance of housing and lending books and providing new services. They are moving away from past collection-based models to create user-centered buildings "where people connect with each other and with the services and resources provided and increasingly work collaboratively (Latimer, 2011)." Libraries in general and Penn State's University Libraries (UL) in particular are incorporating sustainability into collection acquisition and management and are re-assessing use of library spaces. Of its over seven-million copy collections, UL houses 1.5 million infrequently-used items in off-site shelving facilities within 4 miles of University Park campus. UL engaged in a CIC initiative that allowed Google to scan digitally 400,000 books with the potential to free even more shelf space. In addition, consortial agreements to store and share a single collection among partner institutions will continue to free shelf space. "During the next 5 - 10 years UL anticipates increased use of off-site facilities as more digital content is acquired and print use declines. As University Libraries responds to the demand for more collaborative study space, collections will be displaced from their on-campus locations (About the Annex, 2015)." Figure 1B above portrays one alternative use of library space.

Use of Visualization in Libraries

Geographic information systems (GIS) have been used to manage library facilities by measuring in-library book-use behavior related to height of the bookshelves (Xia, 2004b); observing wayfinding habits of users in order to strategically market services and collections along high-traffic areas (Mandel, 2009); record occupancy of study areas using floor plans and illustrate use of space graphically (Xia, 2005); and map user activities to better aid library policy and space planning decisions (Given & Archibald, 2015). Not only has GIS been used for assessing needs within the library but it has also been used to improve services through appropriate site selection and geographic accessibility (Park, 2012); improve virtual reference services (Mon., Bishop, McClure, McGilvray, et al., 2009); better understand local population characteristics (Adkins & Sturges, 2004); improve user experience (Brundin, 2007); and in locating books and consortial materials (Aguilar-Moreno & Granell-Canut, 2013) as well as access to collections and materials (Pfander & Carlock, 2004; Solar & Dolabar, 2005).

In many libraries, data visualization is currently being used to strengthen library assessment efforts and to market the value of libraries. Library collections content has been visualized and graphed in two- and three-dimensions based on classifications systems, allowing comparisons with other libraries and identifying collections strengths and weaknesses (Denton, 2012). A display of browse-able, virtual bookshelves helps patrons find nearby materials of related subject areas (Project: Stack Life, 2015). Starfields have been used to visualize in a two-dimensional grid book collections of a physical library's catalog, revealing collection strengths and weaknesses and allowing temporal filters. (Sanchez, Twidale, Nichols, & Silva (2005). Academic libraries have used data visualizations in Tableau to support library assessment by identifying and promoting library collections in concert with library promotions (Murphy, 2015) and to better understand user relationships (Goswami, Mukherjee, Kharbanda, Gupta, & Soni, 2010). Goswami *et al's* visualizations linked book use nodes to research-user communities to identify user needs and to suggest certain "uncommon" books to researchers in the same field, thus increasing library value to patrons. Data and statistics dashboards showing use and collection statistics are being used by both public libraries (2014 Statistics; North Carolina, 2015) and academic libraries (Activities at Harrell, 2014) to advertise use and value. However, while GIS and visualizations have been used separately in many forms within academic libraries, using GIS both to geographically locate books and to provide visual data analyses has not been attempted to date.

Managing and visualizing book collections at Penn State

Mapping the physical (geographic) location of each volume within libraries can open a myriad of visualization analyses that support decision making for retaining, moving off-site locally or collaboratively, and for efficient space allocations. Keeping one book on an open shelf has been estimated to cost $4.68 each year in 2015 dollars (Courant and Nielsen, 2010; CPI Tables, 2015). The cost of high-density (off-site) storage was estimated by Courant and Nielson to be $0.94 (2015 dollars), The average width of a Penn State book is 1.5 inches (V. Neff, personal communication, December 10, 2014). Therefore, we can estimate the maintenance dollar-amount saved and space freed by removal of these books, even if we cannot assess accurately the opportunity cost of continuing to use these spaces as book repositories.

One critical criterion often used by STEM (Science, Technology, Engineering and Math) librarians in making retention decisions is whether a book has been used in the past five years. Analyzing book age and use and visualizing this information in a relatable construct has potential cost savings for a library whose books number in the millions.

Thus the purpose of this study was to examine book usage and assess where cost-savings might be applicable for the Life Sciences Library at the Pennsylvania State University.

Methodology

Data Acquisition and Preparation

Penn State Libraries stores bibliographic, library location, and usage information in the Integrated Library System (ILS). Cumulative use data, starting with the first computerized system in 1978, have been retained for each book in the collections. Annual use data have been stored for each book since mid-2001 to current day.

For this study, book usage data for the Life Sciences Library were analyzed. This consists of approximately 305,000 copies of serials, journals or magazines issued at regularly occurring intervals, and monographs. The stacks area consumes 43% of Life Sciences floor space, with the other portion taken by staff offices and computers and seating for student use (Figure 2B).

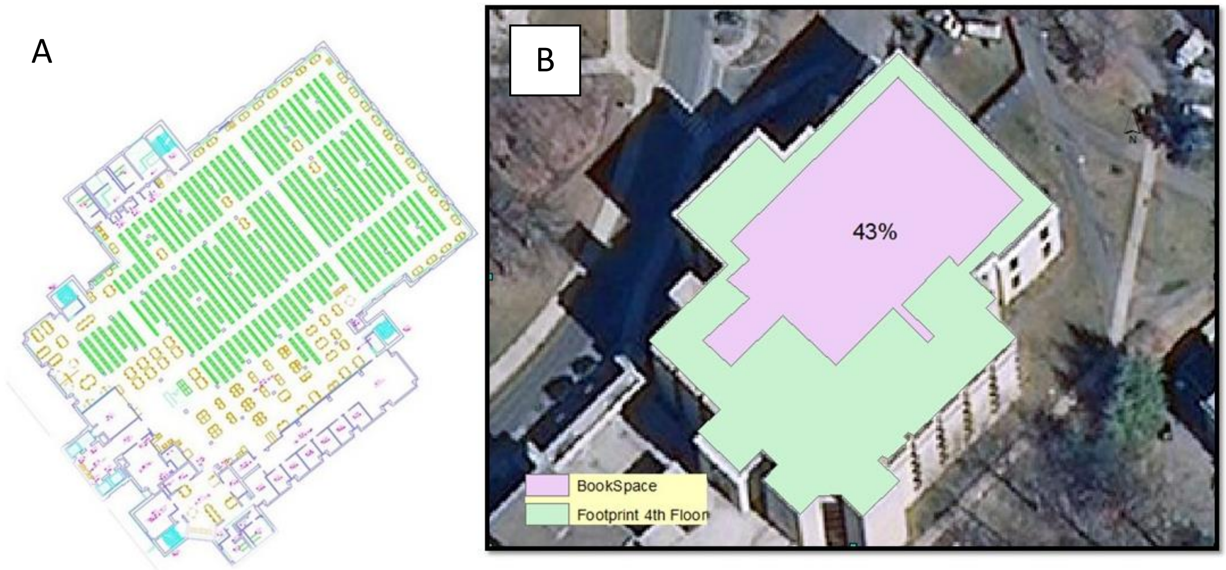

Figure 2. (A) A current CAD rendering of Life Sciences Library was used to draw the footprint of the stacks area as shown in (**B),** created using ArcMap 10.2.2.

In this study, Library of Congress (LC) call number classes and subclasses, in other words related subject areas, were evaluated in terms of number of checkouts, age (publish date), and most recent checkout data. Most research and academic libraries in the United State use the LC classification system that divides all knowledge into 21 classes, using a single letter. These classes are further divided into subclasses with two-letter, sometimes three-letter classifications. The Life Sciences Library houses Q, R, and S classes and the subclasses within them (see Appendix A for subject-area descriptions).

Two datasets were obtained as csv files from the library database. One dataset contained cumulative number of checkouts since 1978, the year during which the library became computerized. Attributes of the cumulative data included a unique ID (barcode), title, truncated call number (two alpha characters), year published, number of accumulated checkouts, the last checkout date, the year the book entered our system (which might be different from the publish date). This dataset will be referred to as “Monographs” for the remainder of this paper. The second dataset contained the number of checkouts recorded for each book in each calendar year, starting mid-2001 and continuing through the present year. Attributes of the annual data included the year, number of checkouts, title, truncated call number, year published, unique ID (barcode), and the year the book entered our system. This dataset will be referred to as “CheckoutsByYear” for the remainder of this paper.

Serials were excluded from these analyses for two reasons. Firstly, actual age of the volumes was inaccurate since each volume was assigned the first year of the first volume published rather than the year the volume represented, thus skewing the actual age of the volume. Secondly, the serials are moved off open shelves on a continuing basis as they are converted to digital formats and as consortial sharing takes place. Data for Monograph, on the other hand, were found to be accurate with only six records in need of editing to capture the actual date the monograph was published. eBooks were not included primarily because they are not part of the ILS database, because they are supplied by multiple corporate sources with each source checkout rate captured, if at all, in different data formats and types, and also because consortial sharing makes it difficult to attribute checkouts to any one partner institution. Consequently, only print monographs were analyzed.

Locating each book

Life Sciences Library holds over 1800 shelving units of seven shelves, each shelf filled with up to 35 inches of books. The floor footprint of the shelving units are 36 inches by 10 inches or 360 square inches. Sixty back-to-back shelving units span the width of the stacks with 36 inches between rows. To mimic the physical location of each book, a grid (known as a fishnet) was created to replicate the location of each book so that a total of 375 rectangles were placed in each of 368 rows. Monograph records were joined to the fishnet so that each rectangle had associated attributes of each of the monographs and represented the physical location of the book. Each book (rectangle) had a virtual dimension of approximately one-quarter inch. See Figure 3 below.

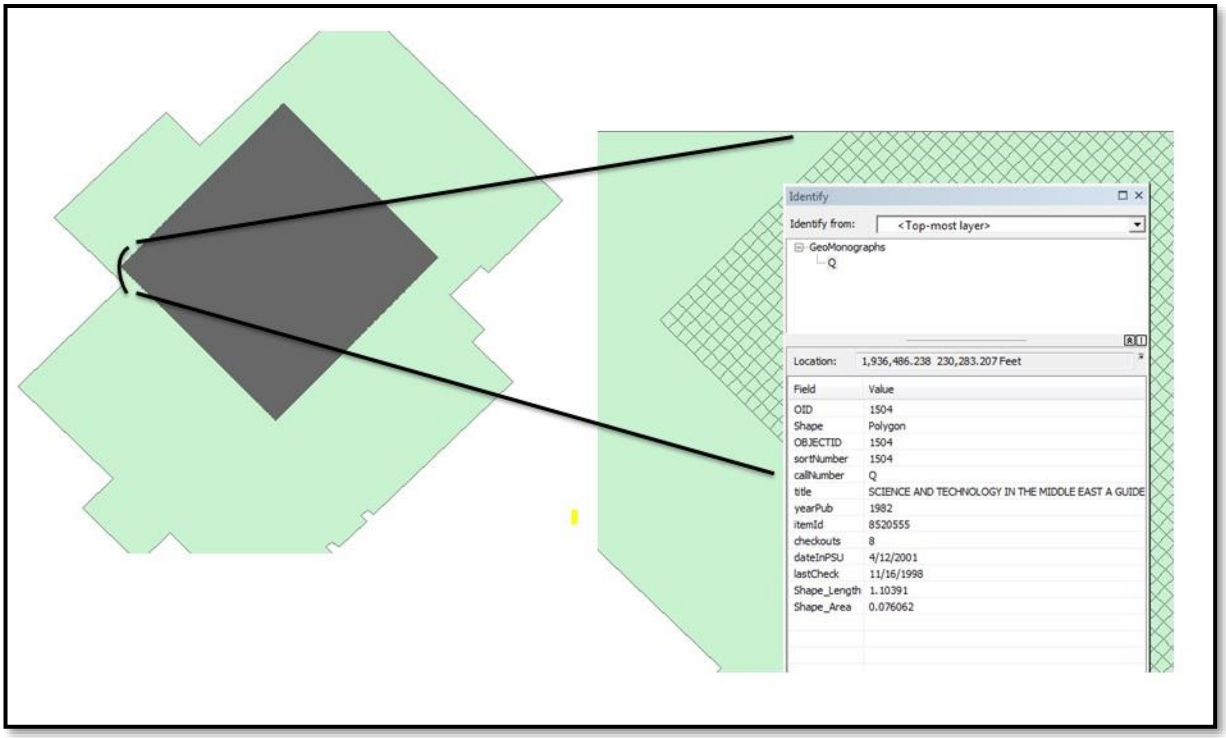

Figure 3. Individual monographs (books and attributes) loaded into a virtual representation of actual shelf locations, filling a portion of the actual footprint of the stacks area of the Life Sciences Library. Created using ArcMap 10.2.2.

One interesting aspect of loading monograph data into the geodatabase is that the data were loaded from the left (north) side of the fishnet to the right (south) side, returning to the left side to load the next row (Figure 4A). Libraries, however, fill the rows of shelves from left to right, turn the corner, and fill the next row from right to left (Figure 4B).

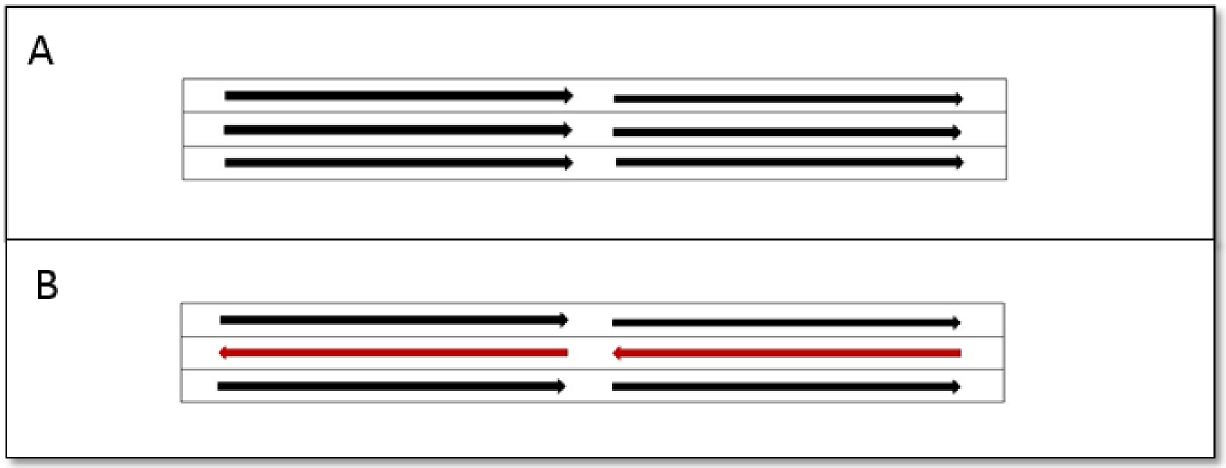

Figure 4. (A) Books load into a fishnet from left to right. **(B)** Books are shelved left to right, turn the corner, and are shelved right to left, alternating each row.

For this study, the csv file was modified to transpose alternating groups of 375 books (one row) so that the books were "shelved" as practiced. Thus, books that are closely related in subject matter remain in close proximity to books of the same subject matter, very comparable to Tobler's first law of geography, "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). As books progress to the rear of the library, they become less related in terms of subject areas to books in the front of the library.

Mapping and visualizing book attributes

Using the monograph-fishnet feature class, new feature classes for the LC class and subclass designations were created. Data were summarized by subject areas (subclasses) and loaded into a geodatabase table. Measures of central tendency were evaluated to identify average number of checkouts, average age and average last checkout date. Differences in number of checkouts, age, and use across the subject areas of the library (biology, health, and agricultural sciences) were then visualized in choropleth maps. Choropleth maps were used to evaluate the utility of visualizing indoor space for making collection decisions and to provide direction for criteria used in the analysis of potential book relocation. A three-dimensional (3D) analysis enhanced the relationships between age, use, and relevancy.

While the mapped analysis displayed characteristics of the collection that visualized its use and age in ways never shown before, these characteristics needed also to be visualized in forms that are typically used by librarians to make collection relocation decisions. Because often Excel tables and charts are used for these decisions (Greiner & Cooper, 2007), use, age, and relevancy were visualized in Excel charts.

Spatial analysis

A Kernel Density Estimation (KDE) was run on the monographs to look for spatial clustering of checkout rates. In theory, each book has the same opportunity of being checked out as every other book in the library, but the KDE could indicate a spatial variation in checkout rates. A radius of five feet was used so that checkout rates were smoothed over a radius of five feet. Since each virtual book was one-quarter inch, each radius in 360 degrees included over 180 books. To identify areas of statistically significant low and high checkout rates, a Getis-Ord Gi* analysis was used to assess the number of checkouts the collection experienced. The Getis-Ord Gi* was chosen because it identifies statistically significant spatial clusters of high values (hot spots) and low values (cold spots). In this case high checkout rates are considered hot spots and low checkout rates are cold spots (Getis-Ord Gi*, 2016). but the Getis-Ord GI* searches through neighboring books' checkout rates to see if there are higher or lower checkout rates than expected in this neighborhood. Inverse-distance spatial weighting was used with a Euclidean straight-line distance of five feet.

Cost and Space Evaluation

Books were identified that had not been checked out in the last five years, had less than subject-average checkouts (10), and were older than the subject-average age (28). These books were identified using a compound selection process in ArcMap and were mapped into a new layer for visual analysis.

Once identified these books were costed out to estimate yearly expense to University Libraries using Courant and Nielson's (2010) cost of keeping a book on an open shelf adjusted with Consumer Price Index information to 2015 dollars. Using the average of 1.5 inches for a Life Sciences Library book, shelf and ultimately floor space that could be freed by book relocation was calculated.

Results

Collection Profile

The 137,721 shelved monographs in Life Sciences in 2015 have been checked out over 1.3 million times. 20,916 books have never been checked out. Checkout and age frequency are displayed in Figure 5 below.

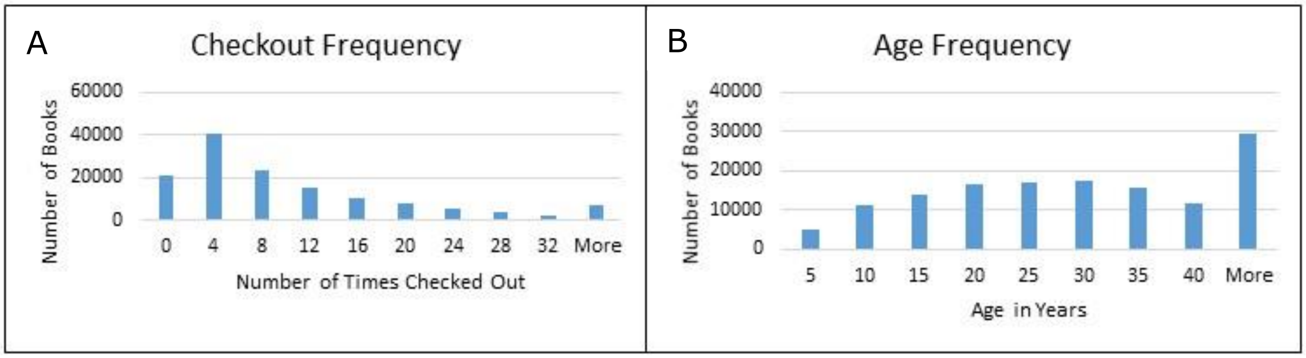

Figure 5 (A) Checkout frequency of books and **(B)** age frequency of books.

Figure 6 shows that the publication date of the collection varies widely from early 1800's through 2016, with an average age of around 28 (stdev 17) years. Checkout frequency shows there is a wide variation in checkouts of individual books. Average checkout rate per monograph in Life Sciences Library is 10 (stdev 13) checkouts. Created using Excel 2016.

Year Published		**Number of Checkouts**	
Mean	1987	Mean	10
Median	1989	Median	6
Mode	1990	Mode	0
Standard Deviation	17	Standard Deviation	13
Range	2016	Range	796
Minimum	1801	Minimum	0
Maximum	2016	Maximum	796
Count	137876	Count	137876

Figure 6. Statistical measures of age and checkout frequency of the collection. Created using Excel 2016.

A scatterplot of age and checkouts in Figure 7A below shows the heavier usage in recently published books while also showing declining use of print materials. Heavily used books reflect the more recent practice of buying text books for students to borrow for limited time periods. These materials are course reserve books. In Figure 7B visualizing the ratio of average checkouts to average age of the subject areas indicates that the QA (computers and programming) and the RX (homeopathy) sections are either heavily used and/or among the newer collections.

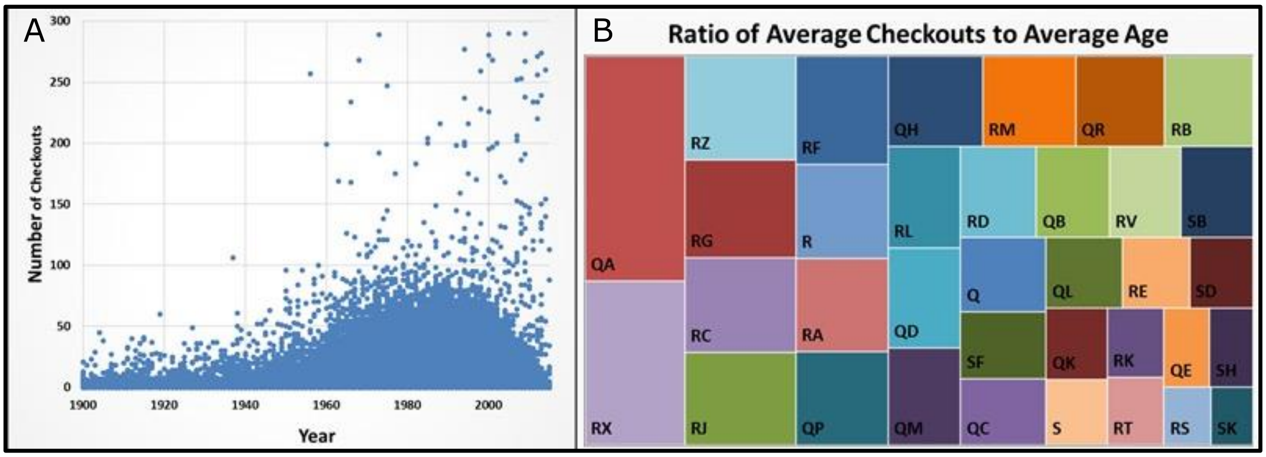

Figure 7(A) Scatterplot of number of checkouts to the year published shows heavier use in recently published books. **(B)** Ratio of average checkouts to average age shows that QA's and RX's might be heavily used and/or recently published. Created using Excel 2016.

Spatial Analysis

Visualizing monograph usage and age in choropleth maps revealed that QA's and RZ's (other systems of medicine that include chiropractic, osteopathy, mental healing, magneto therapy, among others) were checked out most often and checked out most recently (Figure 8B and 8D), indicating a currency and relevancy. Conversely RT's (nursing) and SK's (hunting sports) were checked out infrequently and not very recently. Generally, average checkouts were higher in the front of the library where computers and programming languages reside and lower in the rear of the library where agricultural sciences are located (Figure 8B).

In terms of subject-area age, QA's, and R to RB's (general medicine, public aspects of medicine, and pathology) are among the newest in the collection Figure 8C). Computers and programming languages are the most current and the most recently used, but the relationship of these attributes is less clear in other subject areas. For example, the RZ's which indicated recent, frequent use, do not indicate recent publication date. RX subject area, homeopathy, has experienced a high checkout rate but has an older averaged age.

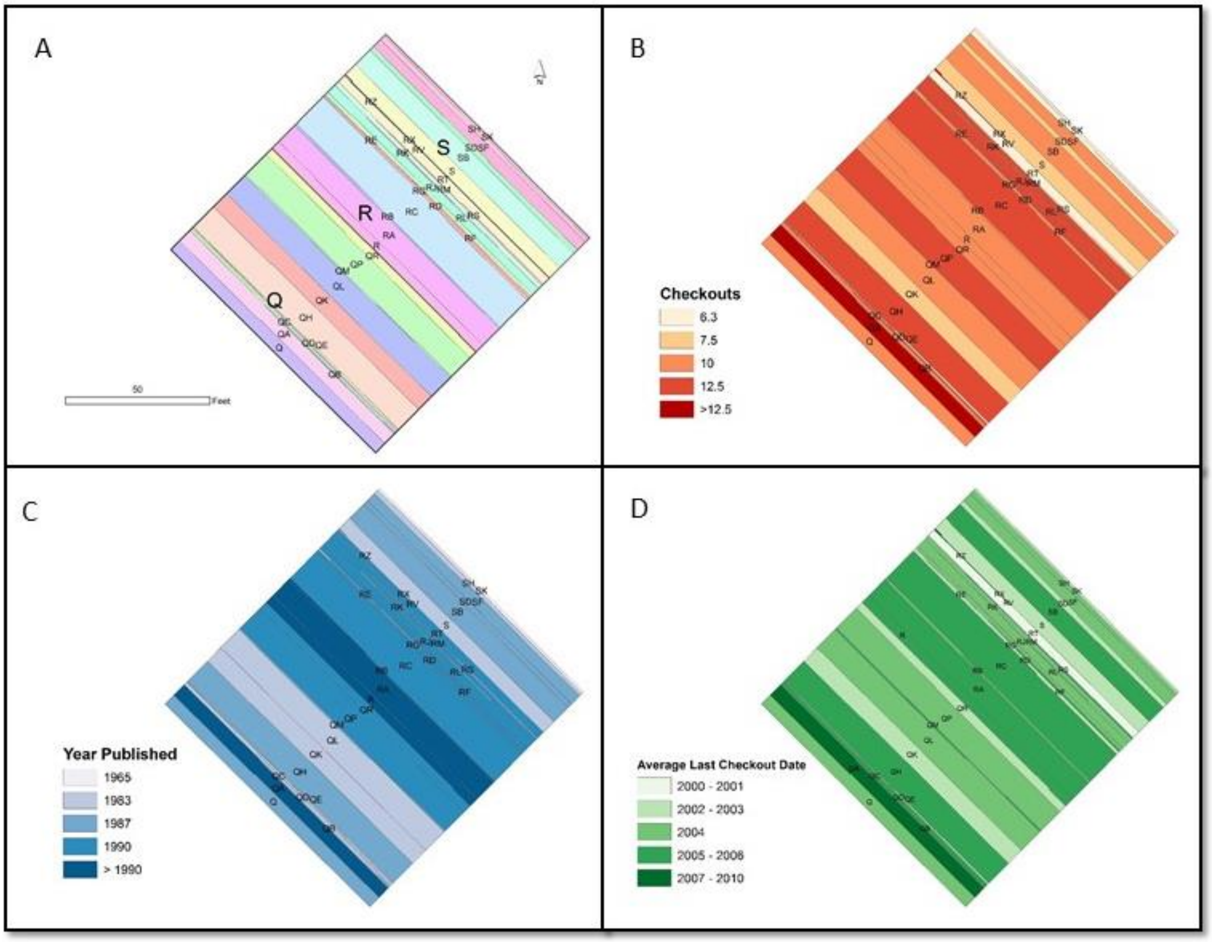

Figure 8. (A) Subject areas, LC subclasses, are delineated by different hues within the LC classes of Q, R, and S. Averaged checkout rates **(B**) reveal that QA's (computers and computer programming) are used most frequently, while SH (aquaculture, fisheries, and angling) and SK's (hunting sports) are used less frequently. Averaged year published **(C)** indicates that recently published books reside in subject areas QA (computers and programming), R (general medicine), RA (public aspects of medicine), and RB (pathology). **(D)** Most used books reside in the QA (computers and programming) section. Created using ArcMap 10.2.2.

Moving age, use, and relevancy graphs into a three-dimensional (3D) representation in Figure 9A makes attribute comparisons easier. QA's are published recently, used often and used recently. Adding a 3D layer of each monograph's checkout rate to the averaged checkout rate in Figure 9B facilitates identifying outliers. In Figure 9B, checkout rate was divided by 10 so that the visualization could be contained in a small map, but also to aid in identifying only those frequencies that truly represent an unusual checkout rate. These outliers represent books purchased as course reserve books and can be checked out every two hours by students who do not otherwise have access to these materials. These books are selected for a use other than to supplement and enhance class and research efforts, and they do not typify collection use.

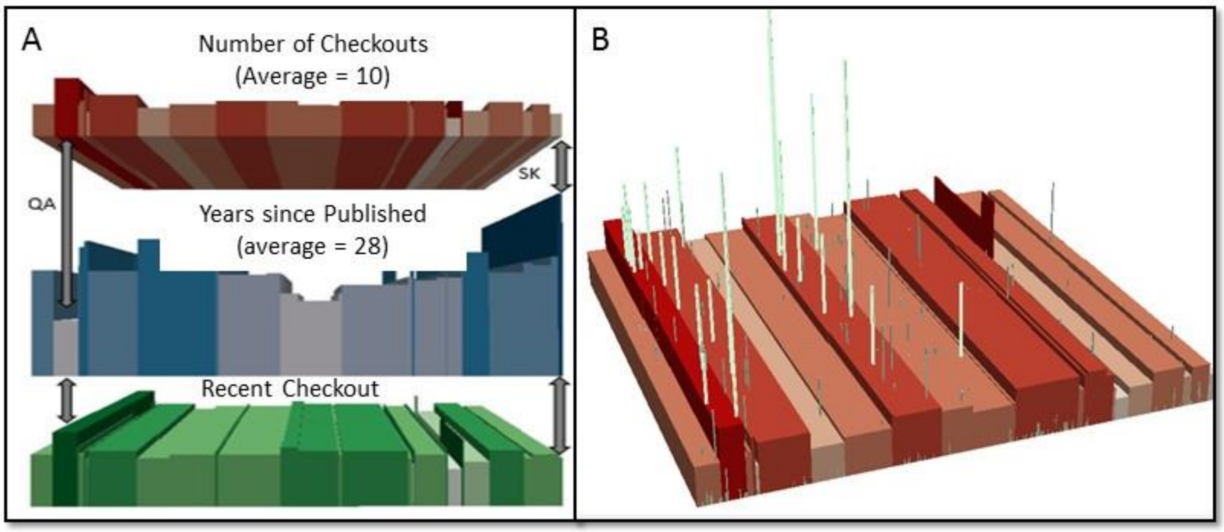

Figure 9. (A) A 3D representation of use, age, and relevancy indicates QA's are heavily used, published recently, and checked out recently. Conversely, the SK (hunting sport s) are not used heavily or recently and are an older collection. **(B)** Outliers, turquoise cylinders, are identified by their height above the average checkouts for each subject area. Created using ArcMap 10.2.2.

A density of book usage was illustrated in Figure 10A in which the KDE indicated that high checkout rates are concentrated in certain subject areas, giving a spatial relationship to checkout rates. The Getis-Ord Gi* analysis in Figure 10B indicates that there is a statistically significant higher checkout rate for the QA section from the overall collection checkout rates. Similarly, the agricultural section to the rear of the library was shown to be statistically significantly lower than the checkout rate for the overall collection. Looking at the entire collection in this manner for each of the three attributes of age, use, and relevancy can point the collections development librarian to areas of the stacks where intervention may be needed.

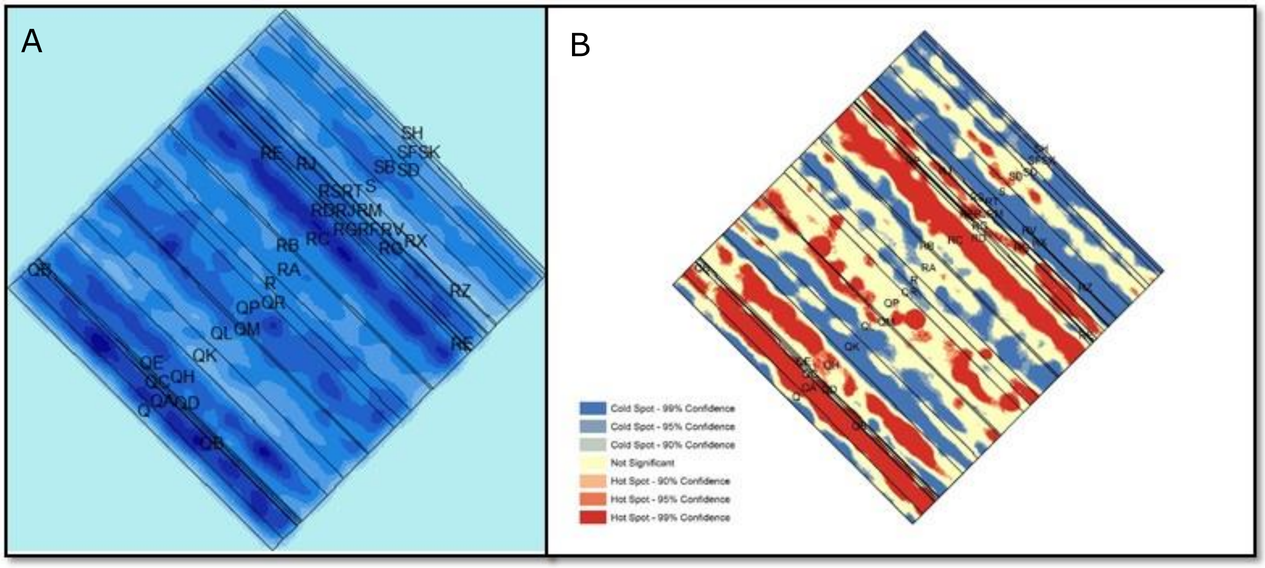

Figure 10. (A) Kernel Density Estimation using a five-foot radius shows spatial clustering of number of checkouts. **(B)** A Getis-Ord Gi* shows statistically significant cold and hot spots of number of checkouts using a five-foot radius across subject areas. Created using ArcMap 10.2.2.

Cost and Space Evaluation

Of the 137,876 monographs on Life Sciences floor, 32,163 or 23% meet criteria that indicate possible removal from the floor. These represent books that could be moved to off-site storage and, if needed, be brought to the lending desk for “just-in-time” service. Using the logic and thorough analysis found in “The Cost of Keeping a Book” by Courant and Nielsen, their presence in Penn State’s open stacks will accrue continuing annual cost and usurp space that could be used for alternative purposes.

In Figure 11A, the QA subject area (computer and programming) shows mostly white areas, indicating that few books qualify for removal. The rear of the library indicates a heavier population of books that can be removed. In Figure 11B, solid-hued areas within the S (agricultural) section meet the criteria for removal. These candidates for removal have not been used since 2010, have less than 10 total lifetime checkouts, and are more than 28 years old. According to Courant and Nielsen’s cost estimate, revised to 2015 dollars, these books cost University Libraries $150,522 each year that they remain in the open stacks. Not all of these books will be disposed of; most will ultimately be moved to off-site, high-density storage. Storing these books in off-site shelving would cost $30,271 annually. Using Courant and Nielson’s metrics, University Libraries spends $120,000 more each year the books remain in open-stacks storage and not in high-density closed-stacks storage. However, there is a cost to moving these books to off-site storage, a cost that includes wage and benefits of moving personnel and truck and equipment depreciation and gas costs, which would be taken out of the first year's savings. Also, the cost of keeping a book on an open shelf includes costs that would not disappear with their removal: heating, cooling, cleaning, lighting, and insurance for example. Consequently, assessing the value of space freed offers an alternative approach.

Using the average of 1.5 inches for a Life Sciences Library book, these books occupy 48,225 inches of linear shelf space, 608 shelves, 229 shelving units. These units occupy 82,440 square inches of floor space or 573 square feet. Because these shelving units would occupy at least three rows of shelving units with 36 inches between each row, an additional approximate 750 square feet would be freed, giving a total of 1,323 square feet of freed floor space. This freed space could be re-purposed so that the university would not have to rent or build new space. Consequently, the university could better identify out-of-pocket cost savings, which could be considerable. Yet another method of assessing the true cost for keeping these books is to identify the opportunity cost to students and faculty of not having a collaborative work or teaching space in which to work. Assessing lost opportunities is not as easily quantified.

Subject area details of candidates for removal can be found in Appendix B.

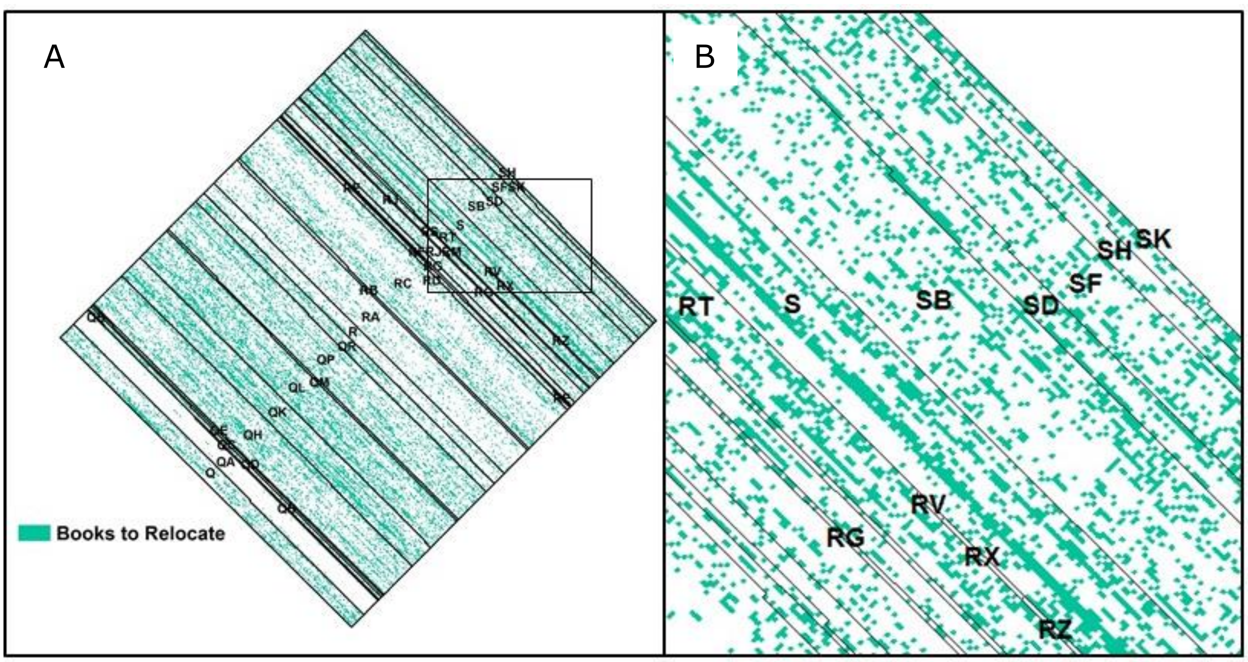

Figure 11. (A) Books older than subject collection average (28), used less than subject collection average (10) and not used in the last five years. **(B)** A closer view of books in the S class (agriculture) that could be relocated. Created using ArcMap 10.2.2.

Discussion and Conclusion

Academic libraries have been undergoing a transformation from being viewed as a place that houses invaluable collections of intellectual and cultural achievement to being viewed as a place that facilitates collaborative teaching and learning in technology-rich spaces. The purpose of this study was to assess the utilization of these spaces by analyzing book demand using spatial analysis methods in the Life Sciences Library at Penn State University. To do so a GIS was used to assess book usage across the broad subjects in the Life Sciences Library. Book usage statistics were analyzed using spatial analysis methods that provide visualizations of where books reside and using spatial statistical analysis of subject areas highlighting “hot topic areas” vs “cold topic areas” (that is popular vs. less popular subjects). A cost and space analysis was also performed to estimate cost-savings and space-freed through removal of unused books.

Geovisualizing book attributes presents an easily consumed and understood representation of use, age, and relevancy. Using GIS, libraries can benefit from visually analyzing collections and use trends, can identify collection strengths and weaknesses, can allocate space to its best use, and monitor statistics in a way that is impossible with current spreadsheets and without costly commercial assessment add-ons. This analysis has confirmed many of the benefits of GIS visualizations for collection analysis and trends, but also has gone one step further in quantifying cost savings from collection relocation or withdrawals and in quantifying floor space liberated for collaborative and creative effort.

In 2011, Andrew Coyle, writing in Library Hi Tech said, “GIS is going to be implemented in libraries sooner rather than later. The libraries that implement GIS early will have an intellectual advantage over those coming on-board late.” Coyle argues that librarians can benefit from visually analyzing the collection and its use trends and that analysis “will allow librarians to forecast demand for future allocations, uncover collection strengths and weaknesses, and monitor statistics in a way that is impossible with current spreadsheets.” This analysis has confirmed many of the benefits of GIS visualizations for collection analysis and trends, but also has gone one step further in identifying cost savings from collection re-allocations or withdrawals and in identifying floor space liberated for collaborative and instructional use. As academic libraries continue to migrate from a “just-in-case” model to a “just-in-time model” for print collections, enter into consortial sharing of print collections, add electronic books to their collections, and re-purpose space for collaborative and creative patron use, visually analyzing book usage and relevancy will play an important role in prioritizing collection withdrawal and relocation decisions. As universities continue to adopt sustainable fiscal and environmental operations, appropriate campus space allocation through effective use of GIS will become increasingly important.

References

2014 Statistics Dashboard. Canton Public Library. Retrieved from https://www.cantonpl.org/about-us/dash.

About the Annex (2015). Penn State University Libraries. Retrieved from http://www.libraries.psu.edu/psul/annex/abouttheannex.html.

ACRL Visual Literacy Competency Standards for Higher Education (2011). Association of College and Research Libraries. The American Library Association. Retrieved from http://www.ala.org/acrl/standards/visualliteracy.

Activities at Harrell HSL in 2014. George T. Harrell Health Sciences Library. Penn State Hershey. Retrieved from http://www1.libraries.psu.edu/hersheydashboard/.

Adkins, D., & Sturges, D. K. (2004, May/June). Library service planning using GIS and census data. Public Libraries, 43(3), 165-170. **Reprinted as a chapter in The PLA Reader for Public Library Directors and Managers, 2009.

Aguilar-Moreno, E., & Granell-Canut, C. (2013). Sistema de informacion geografica para unidades de informacion. Geografic para unidades de informacion. *El Profesional De La Información*, *22*(1), 80-86. doi:10.3145/epi.2013.ene.11

Brundin, M. R. (2008). Using a geographic information system to analyze public library performance measures: A spatial analysis of the 2004 Ontario public library statistics. Paper presented at the ASPRS 2007 Annual Conference Tampa, Florida May 7–11, 2007.

CPI Tables (2015). Consumer Price Index. Bureau of Labor Statistics. Retrieved from http://www.bls.gov/cpi/tables.htm.

Courant, P. N., & Nielsen, M. (2010). On the cost of keeping a book. *The idea of order: Transforming research collections for 21st century scholarship*, 81-105.

Coyle, A. (2011). Interior library GIS. Library Hi Tech, 29(3), 529–549. doi:http://dx.doi.org.ezaccess.libraries.psu.edu/10.1108/07378831111174468

Data Visualization: Tools, Techniques, and Practice. 1996-2015. Association of College and Research Libraries. The American Library Association. Retrieved from http://www.ala.org/acrl/datavisualizationpreconference.

Denton, W. (2012). On dentographs, A new method of visualizing library collections. *Code4Lib Journal,* (16).

ESRI 2014. ArcGIS Desktop: Release 10.2.2. Redlands, CA: Environmental Systems Research Institute.

Ferguson, C. (2007). Technology left behind - GIS and the library: Part 2. *Against the Grain, 19*(1), 80-82.

Given, L. M., & Archibald, H. (2015). Visual traffic sweeps (VTS): A research method for mapping user activities in the library space. *Library & Information Science Research*.

Goswami, S., Mukherjee, A., Kharbanda, M., Gupta, A., & Soni, P. (2010). Visualisation of relationships among library users based on library circulation data. *DESIDOC Journal of Library & Information Technology, 30*(2), 26.

Greiner, T., & Cooper, B. (2007). *Analyzing library collection use with excel*. Chicago: American Library Association.

Hot Spot Analysis (Getis-Ord Gi*). (2016). ArcGIS for Desktop. Environmental Research Systems Institute. Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/hot-spot-analysis.htm

Latimer, K. (2011). Collections to connections: Changing spaces and new challenges in academic library buildings. *Library Trends, 60*(1), 112-133. Retrieved from http://search.proquest.com.ezaccess.libraries.psu.edu/docview/903206239?accountid=13158.

Mandel, L. (2010). Toward an understanding of library patron wayfinding: Observing patrons' entry routes in a public library, *Library & Information Science Research*, 32(2), 116-130. http://dx.doi.org/10.1016.

Mon, L., Bishop, B., McClure, C., McGilvray, J., Most, L., Milas, T., & Snead, J. (2009). The geography of virtual questioning. *The Library Quarterly, 79*(4), 393-420. DOI: 10.1086/60538. Stable URL: http://www.jstor.org/stable/10.1086/605381

Murphy, S. A. (2015). *How data visualization supports academic library assessment: Three examples from the Ohio State University libraries using tableau*. Chicago: Association of College and Research Libraries./j.lisr.2009.12.004.

North Carolina Public Library Data. State Library of North Carolina. Retrieved from http://plstats.nclive.org/data.html.

Park, S. J. (2012) Measuring public library accessibility: A case study using GIS, *Library & Information Science Research*, 34:1, http://dx.doi.org/10.1016/j.lisr.2011.07.007.

Project: Stack Life. Library Innovation Lab. Harvard University. Retrieved from http://librarylab.law.harvard.edu/projects/stack-life.

Sanchez, J. A., Twidale, M. B., Nichols, D. M., & Silva, N. N. (2005). Experiences with starfield visualizations for analysis of library collections. Paper presented at the Proceedings of SPIE*, 5669*(1) 215-225. doi:10.1117/12.587990.

Tobler, W. R.. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46*, 234–240. http://doi.org.ezaccess.libraries.psu.edu/10.2307/143141

Ward, S. M. (2015). *Rightsizing the academic library collection*. Chicago: ALA Editions, an imprint of the American Library Association.

Xia, J. (2004a), Library space management: a GIS proposal, *Library Hi Tech*, 22(4), 375 – 382. http://dx.doi.org/10.1108/07378830410570476.

Xia, J. (2004b). Using GIS to measure in-library book-use behavior. *Information Technology and Libraries, 23*(4), 184-191.

Xia, J. (2005). Visualizing occupancy of library study space with GIS maps. *New Library World, 106*(5), 219-233. Retrieved from http://search.proquest.com.ezaccess.libraries.psu.edu/docview/229577836?accountid=13158.

Appendix A.

LIBRARY OF CONGRESS CLASSIFICATION OUTLINE

CLASS Q - SCIENCE

Q – Science (General)
QA – Mathematics, computers, computer software
QB – Astronomy
QC – Physics
QD – Chemistry
QE – Geology
QH – Natural history, biology
QK – Botany
QM – Human anatomy
QP – Physiology
QR – Microbiology

CLASS R - MEDICINE

R – Medicine (General)
RA – Public aspects of medicine
RB – Pathology
RC – Internal medicine
RD – Surgery
RE – Ophthalmology
RF – Otorhinolaryngology
RG – Gynecology and obstetrics
RJ – Pediatrics
RK – Dentistry
RL – Dermatology
RM – Therapeutics. Pharmacology
RS – Pharmacy and materia medica
RT – Nursing
RT – Specialties in nursing
RV – Botanic, Thomsonian, and eclectic medicine
RX – Homeopathy
RZ – Other systems of medicine

CLASS S - AGRICULTURE

S1 – Agriculture (General)
SB – Plant culture
SD – Forestry
SF – Animal culture
SH – Aquaculture. Fisheries. Angling
SK – Hunting sports

Appendix B.
Candidates for Removal

Subjects (Subclasses)		Averaged publication year	Averaged checkouts	Averaged last date checked out	Number of books identified for removal	as % of removal total
Q	Science (General)	1971	3	7/16/1997	1574	4.73%
QA	Mathematics (includes computers)	1972	4	7/20/1998	120	0.36%
QB	Astronomy	1965	4	11/22/2000	53	0.16%
QC	Physics	1966	3	3/6/1998	353	1.06%
QD	Chemistry	1967	4	8/5/1998	199	0.60%
QE	Geology	1969	3	5/14/1997	185	0.56%
QH	Natural history (General) and Biology (general)	1972	4	1/16/1997	3265	9.80%
QK	Botany	1966	2	10/7/1997	2770	8.32%
QL	Zoology, Animal behavior, Morphology, Anatomy, Embryology	1969	3	2/3/1998	3789	11.37%
QM	Human anatomy	1972	4	10/21/1998	192	0.58%
QP	Physiology	1975	4	8/16/1996	3015	9.05%
QR	Microbiology	1975	4	5/11/1995	1051	3.16%
R	Medicine (General)	1975	4	9/30/1998	561	1.68%
RA	Public aspects of medicine	1979	4	3/4/1998	1225	3.68%
RB	Pathology	1977	4	11/12/1995	134	0.40%
RC	Internal medicine	1976	4	12/17/1996	2834	8.51%
RD	Surgery	1975	4	7/22/1996	314	0.94%
RE	Ophthalmology	1971	4	12/24/1995	85	0.26%
RF	Otorhinolaryngology	1975	5	9/4/1998	97	0.29%
RG	Gynecology and obstetrics	1975	4	4/24/1997	254	0.76%
RJ	Pediatrics	1976	4	3/24/1996	708	2.13%
RK	Dentistry	1970	4	8/24/1998	46	0.14%
RL	Dermatology	1972	5	2/11/1999	35	0.11%
RM	Therapeutics. Pharmacology	1975	4	7/27/1995	448	1.34%
RS	Pharmacy and materia medica	1973	3	2/12/2000	127	0.38%

RT	Nursing	1978	3	7/9/1993	1383	4.15%
RV	Botanic, Thomsonian, and eclectic medicine	1982	6	12/14/2010	1	0.00%
RX	Homeopathy				0	0.00%
RZ	Other systems of medicine	1968	4	7/1/2000	16	0.05%
S	Agriculture (General)	1971	2	10/28/1997	2992	8.98%
SB	Plant culture	1973	4	8/18/1998	2224	6.68%
SD	Forestry	1971	2	7/15/1998	791	2.37%
SF	Animal culture	1971	4	9/22/1998	1354	4.06%
SH	Aquaculture. Fisheries. Angling	1967	3	10/4/1999	557	1.67%
SK	Hunting sports	1956	3	12/10/1998	559	1.68%