

Patterns of Use and Usage Factors for Online Databases in Academic and Public Libraries

Carol Tenopir and Danielle Green
University of Tennessee
Knoxville, TN 37996

ABSTRACT

Database usage data from a random sample of academic and public libraries in the U.S. and Canada reveals patterns of use in selected types of libraries. Library users in both public and academic libraries tend to use commercial online databases most frequently early in the week, mid-day, and at times that correspond to the academic calendar (November in this six-month sample.) The mean number of simultaneous users is correlated with the size of the population served and the number of workstations available, but relatively low numbers of users are simultaneously logged on to research databases at all sizes of libraries. A questionnaire sent to these same libraries identified many other factors that might influence database use, including levels of instruction, availability of remote login, placement of a database on the library's homepage, although none of these factors was found to be statistically significant.

INTRODUCTION

Academic and public libraries today offer access to a complex mix of print, CD-ROM, online and web resources. To best allocate budgets and select the options that are best for library users, librarians must consider a variety of pricing options and licensing restrictions, and must predict likely levels of use for each database product. Predicting usage patterns and levels of use for commercial end user online resources is particularly important because prices are often determined by amount of use, total number of users, or by the number of simultaneous searchers allowed.

Predicting likely numbers of simultaneous users is especially difficult without prior usage history for a resource or type of resource. Seeing usage data from other libraries may, however, assist similar libraries to predict usage patterns. The project reported here examines database usage from nearly 200 academic and public libraries in the U.S. and Canada. The data reveal how many simultaneous users are logged on at any one time and when those users are online.

Usage patterns tell only part of the picture, however, because specific factors within a library may influence how much a given database is used. These factors may be obvious (including the number of workstations available, whether or not a library supports remote dial-in access, hours the library is open, etc.) or may be more subtle. More subtle contributing factors may include how each database is referred to in user instruction, how many steps on a library menu screen are required before reaching a database, or the library staff's attitude toward a database. To gather this type of information, a survey questionnaire was sent to the libraries for which I have usage data. The survey captures the opinions and attitudes of librarians, as well as environmental factors within each library that may influence the amount of online use by library users. Together, the usage data and survey data provide a picture of how and why library customers use online databases.

REVIEW OF THE LITERATURE

Several recent studies document the growth of digital resources in libraries and how libraries today embrace a variety of digital options (ALA, 1997; Tenopir & Ennis, 1998a; Tenopir & Ennis, 1998b). End user online searching of commercial databases is now an integral part of library reference services and users have come to expect online access through their libraries.

Collection development and reference librarians now spend much of their time evaluating and selecting online resources, a task more complex than, but not unlike, traditional selection and collection building. Added to the

traditional task, is the need to select the best possible pricing options from among many alternatives and to negotiate online licensing agreements (Abels, 1996; Tenopir, Banks, & Preston, 1998.) The best pricing option will vary from library to library and, within a library, may be different for heavily used resources and those used less often.

Many online resources are purchased with some type of user-based license. In simultaneous usage licenses, librarians must commit to how many online users will be permitted to be connected at any one time to a given database or to a family of databases. Providing for too many concurrent users will waste the library's money, allowing too few will cause frustration to users who must wait to log on to a desired database.

The challenge of anticipating users' needs and usage patterns is not new to libraries. Early usage studies for automated resources helped libraries determine how many public access terminals were optimal for an online catalog (see, for example, Tolle, et al., 1983). Predating automation, libraries tried to determine how many chairs to provide to accommodate people who came to the library or how many staff members to assign to the reference desk at a particular time. Turnstile counts were used to optimize reference department staffing or pickup schedules for shelving (Murfin, 1983; McGrath, 1996).

Although there is no definitive answer and all libraries have some unique situations of their own, learning from other similar libraries may be the best way to predict how many users will likely use a library or a database and when that usage will occur. The usage pattern data in this study was first reported in 1999 (Tenopir & Green; Tenopir, 1999a.). This is the first time that usage pattern data and survey data for both academic and public libraries is reported together. By looking at the two factors together, I hope to begin to go beyond the question of *how* (and how much) online databases are used in libraries, to *why* they are used. Librarians' opinions of why people use a database were solicited in this study, but are not reported here. For the opinion data see: Tenopir 1999b and 1999c.

METHODOLOGY: USAGE DATA

To answer the question of how databases are used requires large amounts of usage data recorded over time. For this study, a major database producer and aggregator provided data for all of its online databases. Although individual libraries often measure the levels of use for their online catalogs, to predict patterns and levels of use for many types of libraries requires this broader look across libraries and across many databases.

This database producer and aggregator provides access to approximately 100 database titles, including bibliographic, directory, and full-text journal article databases. Many of the databases that are both created and distributed by this company overlap in content. For example, the same journal titles may be available in a bibliographic-only (indexing) version or a bibliographic and full-text version; in a complete or an abridged version; or a current version or backfile version. Each library typically subscribes to only one version of a database, although it may subscribe to current and backfile versions of the same database. The most heavily used databases from this aggregator are general-interest titles, that provide access to articles found in major magazines and journals. Most of these titles are created by this company, as well as being distributed by them. Other titles that are distributed by this company are created by other database producers. Many of these are more specialized, including some databases that focus on medicine, psychology, business, and technology.

The library customers for which this company had usage data include over 1200 public, academic, school, and special libraries in several countries, but the majority are public and academic libraries in North America. A random sample of 200 public and academic libraries from the total customer list yielded 99 public and 96 academic libraries with usable data. Public libraries in the sample serve a variety of user populations and academic libraries come from different types of parent academic institutions. (See Table 1 and Table 2.)

Category	Carnegie Class	Number in Sample	Surveys Received
B1	Baccalaureate (Liberal Arts) Colleges I	7	3
B2	Baccalaureate Colleges II	9	5
M1	Master's (Comprehensive) Universities and Colleges I	32	21
M2	Master's (Comprehensive) Universities and Colleges II	3	1
D1	Doctoral Universities I	3	2
D2	Doctoral Universities II	6	5
R1	Research Universities I	15	6
R2	Research Universities II	6	5
AA	Associate of Arts Colleges	12	7
Bus	Schools of Business and Management	3	3
Total		96	58

Table 1. Carnegie Categories For Academic Institutions
(Source: www.carnegiefoundation.org)

Category	User Population Served	Number in Sample
0	Unknown	6
1	under 25,000	11
2	25,001-50,000	10
3	50,001-100,000	17
4	100,001-250,000	20
5	250,001-500,000	18
6	500,001-1,000,000	13
7	over 1,000,000	7
Total		99

Table 2. Categories of Public Libraries

Online usage for every library and each database is captured automatically by the database aggregator in five-minute intervals 24 hours a day. Users may be accessing the databases from within their library or by remote dial-in through their library's networks. At the time of the study the aggregator placed no restrictions on how many simultaneous users from each library could log in. For this study, usage data were sampled once per hour (on the half-hour) from the hours of 8a.m. to midnight from a six-month period (July-December 1997) [1]. Forty-seven different databases were accessed by the 195 libraries in the sample. In all, over one-half million datapoints were collected and analyzed in this study.

METHODOLOGY: SURVEYS

Usage data reveal how many simultaneous users are logged on to any one database at any one of the sampled times. Time stamps on the data provide information that shows average numbers of users by time of day, day of the week, and day of the month for each type of library and each database. What usage does not show is why specific databases are used or what factors within a particular library might influence usage. Information about each library was sought by sending a questionnaire to each of the libraries for which I had usage data.

Fifty-eight (of 96) academic libraries and 61 (of 99) public libraries responded to the questionnaire. The overall return rate is thus 61% (60% for academic libraries and 62% for public libraries.) As can be seen in the final column in Tables 1 and 2, the library respondents to the survey represent a range of library sizes and types.

Survey questions were grouped in 4 categories: 1) "Information About Your Library"; 2) "Information About the Databases You Provide"; 3) "Information About Databases From...[the aggregator for which I have usage data]"; 4)

"Other Factors" (mostly soliciting opinions from the respondents on what factors influence database use.) (See the Appendix.) Some questions varied in the survey of academic and public libraries (academic libraries were asked, for example, how many FTE students are enrolled, while public libraries were asked about the population they served), but most of the questions were the same for both types of libraries. Because the database aggregator does not wish to be identified, the survey questionnaire reproduced in the Appendix excludes the questions from Part 3 of the questionnaire.

Analyzed alone, the survey responses provide a picture of the libraries involved in this study and the opinions of those librarians. When analyzed in conjunction with the usage data, the survey data help identify what characteristics are present in libraries where online usage is higher than the norm.

ANALYSIS OF USAGE PATTERNS

Usage data were analyzed separately for academic and public libraries. Data presented for each aggregate all of the database titles used by the libraries. In academic libraries 38 different databases were used, while 36 databases were used by public libraries. (For more information on the usage pattern data for both academic and public libraries see Tenopir and Green.)

Not surprisingly, online usage in academic libraries follows the regular rhythms of the academic life. As can be seen in Figures 1, 2, and 3, academic library users at all types and sizes of institutions tend to use online databases at the time term papers are due (November in this six-month slice of data), early in the week (Mondays and Tuesdays are the peaks with another bump on Thursdays), and between 11a.m. and 4p.m. local time. Although these usage patterns correspond to the peak hours for library use (and any reference librarian or turnstile count likely would verify this), 72% of these libraries also offer remote online access. It appears that not many college students choose to be working online from a dormitory, office, campus lab, or from home at times other than the normal hours the physical library is open.

Online use in public libraries follows a similar pattern as that in academic libraries (see Figures 4, 5 and 6). Public library users are doing online research most often early in the week (Friday and Sunday are the low points) and in the middle of the day. November is also the peak of the six-month usage, but public library usage remains more constant from month-to-month than does academic library use. 68% of public libraries offer remote access in addition to database use from within the library.

The exact number of times more than one user is searching any single database varies from library-to-library (depending, of course, on the size of the clientele served by the library). Examined in an aggregate, the number of simultaneous users for all 36 databases used in public libraries ranges from zero users (the mode and median) to 39 users. (The mean is .30 users.) In academic libraries the range is from zero simultaneous users (again, the mode and median) to 66 users, with a mean of .28.

Averages of peak usage are of limited value, however, as libraries want to provide the capability for enough simultaneous users to inconvenience the fewest people, without overcommitting their resources. More useful is a cumulative usage table that will show what percent of the usage demands will be satisfied by a certain number of simultaneous usage ports. Table 3 shows that if a large public library (categories 6 and 7) provides the capability for ten simultaneous users, it would be satisfactory over 99% of the time (numbers may vary for extremely high demand or low demand titles.) Smaller libraries (categories 1-5) can provide satisfaction over 99% of the time with 2-5 simultaneous usage ports.

No. of Users	Cumulative Percent							
	Public Library Size							
	0	1	2	3	4	5	6	7
1	93.7	96.9	97.5	95.7	94.8	93.6	89.3	85.5
2	97.3	98.4	99.0	98.2	97.6	96.7	93.0	91.1
5	99.8	99.7	99.9	99.9	99.6	99.4	97.1	97.2
10	100.0	100.0	100.0	100.0	100.0	100.0	99.2	99.1
15	100.0	100.0	100.0	----	100.0	100.0	99.8	99.7
39	----	----	----	----	----	----	----	100.0

Table 3. Simultaneous Users, All Databases, Public Libraries By Size

Table 4 presents simultaneous usage data for academic libraries. All but doctoral and research universities (Carnegie class D2, R1, and R2) can provide over 99% satisfaction rates with 5 simultaneous users. Large research libraries need to provide ten usage ports to achieve the same satisfaction rates.

No. of Users	Cumulative Percent									
	Carnegie Class									
	B1	B2	M1	M2	D1	D2	R1	R2	AA	Bus.
1	97.8	96.8	94.5	91.5	96.4	95.2	91.8	89.9	96.1	90.5
2	99.2	98.7	96.9	96.1	99.0	97.8	95.0	93.3	98.2	94.9
5	99.9	99.9	99.1	99.7	99.8	99.7	98.3	97.8	99.8	99.3
10	100.0	100.0	99.7	100.0	100.0	100.0	99.4	99.7	100.0	100.0
15	100.0	100.0	99.8	----	100.0	100.0	99.6	99.9	100.0	100.0
25	----	----	99.9	----	----	----	99.8	100.0	----	----
66	----	----	----	----	----	----	100.0	----	----	----

Table 4. Simultaneous Users, All Databases, Academic Libraries By Carnegie Class

Not all institutions within the same Carnegie Class have the same number of students, however. Table 5 accounts for this difference by showing how many simultaneous users are required to provide satisfaction 99% of the time based on the number of FTE students at each academic library.

Number of Students Enrolled (FTE)	Average No. of Simultaneous Users Needed for Coverage (99 th percentile)	Number of Libraries
1-2499	2.33	6
2500-4999	3.2	5
5000-7499	3.71	7
7500-10999	5.2*	5
11000-14999	5.0	6
15000-19999	3.0**	3
20000-25999	5.67	3
26000-44999	5.67	3
45000-89999	(no data)	0
90000 or more	27 (26.5)	2

*Note that one library had 12 users, while others in the category had 3 or 4, making mean higher.

**One library was covered 99.9% of time with only 1 user. If outlier is removed, mean is

Table 5. Average Number of Simultaneous Users Required for Coverage (99th percentile), Academic Libraries by Student Enrollment

The database that was used the most in both academic and public libraries was analyzed separately to mitigate any effects on the data from seldom-used databases such as backfiles. In academic libraries this database is a general magazine/journal database specifically marketed to academic libraries. The database includes both bibliographic

records and full texts of some articles. Table 6 shows that a general database of wide appeal to students and faculty will attract more simultaneous users, but the law of diminishing returns still applies when selecting a maximum number of simultaneous use capability.

No. of Users	Cumulative Percent									
	Carnegie Class									
	B1	B2	M1	M2	D1	D2	R1	R2	AA	Bus.
1	95.7	92.1	84.2	85.7	91.3	86.9	78.9	68.9	86.7	85.3
2	98.5	96.4	89.4	92.8	98.2	92.4	84.0	76.3	92.6	92.2
5	100.0	99.5	96.4	99.4	100.0	98.4	91.9	90.6	99.1	99.1
15	----	100.0	99.2	----	----	100.0	97.7	99.7	99.9	100.0
25	----	----	99.7	----	----	----	98.9	----	100.0	----
66	----	----	----	----	----	----	100.0	----	----	----

Table 6. Simultaneous Users, Most Used Database (General Magazine/Journal), Academic Libraries

The most used database in public libraries is a combination bibliographic-full text database of general interest medical information. Table 7 shows that for this database, in small public libraries (categories 1-4) capability for two simultaneous users satisfies 99% or more of the usage, while larger libraries require five simultaneous users.

No. of Users	Cumulative Percent						
	Public Library Size						
	1	2	3	4	5	6	7
1	98.5	98.3	97.7	96.7	94.8	87.0	83.4
2	99.8	99.6	99.5	99.0	98.4	93.5	92.0
5	100.0	100.0	100.0	100.0	100.0	99.3	99.5
8	----	----	----	100.0	100.0	99.9	100.0
12	----	----	----	----	----	100.0	----

Table 7. Simultaneous Users, Most Used Database (Health-related), Public Libraries By Size

ANALYSIS OF SURVEY

The 58 academic and 62 public libraries that responded to the survey questionnaire offer a variety of electronic media for end user searching, including CD-ROM, commercial online, and World Wide Web access. (See Table 8.) Most purchase online access from more than one commercial aggregator and database searching is an important part of everyday reference activities.

Media Types Used By Respondents	Percentage	
	Academic Libraries	Public Libraries
CD-ROM	97	92
Locally loaded	38	46
Loaded on another library's computer	40	31
Commercial online from a vendor	95	82
Commercial online from an OPAC company	24	38
World Wide Web	91	90

Table 8. Types of Electronic Media Provided for End Users

Librarians believe that the databases analyzed in this study are some of the most popular among their users. In academic libraries nearly two-thirds ranked the databases from this aggregator as the most or second most popular with their end users. In public libraries these databases are rated as most popular with over 86% of the libraries.

Many factors might influence how much a database is used in a library. Active promotion of databases through classes, signs, handouts, etc. are more common in academic libraries than in public, but are offered in both. Table 9 shows what these libraries do to promote database searching.

Methods of Promotion Used	Percent	
	Academic Libraries	Public Libraries
Remote access	74	68
Workstations dedicated to these databases	7	21
Databases noted on library system main menu	64	70
Signs	21	42
Handouts	59	59
Databases mentioned specifically in training	81	61
Databases specifically mentioned in academic classes	66	N/A
Specific class assignments require use of these databases	57	N/A
These databases used by students for class assignments	N/A	93

Table 9. Ways Libraries Promote the Specific Databases Analyzed Here, Academic and Public Libraries

FACTORS THAT MAKE A DIFFERENCE

Although database use is undoubtedly influenced somehow by library policies, few of these factors proved to be statistically correlated to amount of use. For those academic libraries for which we received survey responses, size of student population was used to test for correlations between amount of use and various factors in each library. (If differences were examined only between Carnegie classes, any differences might be due to differences in size of enrollment.

Libraries were grouped into high, moderate and low usage groups and compared using Pearson's Chi Square and Fisher's Exact tests. Variables tested were remote login, the number of workstations provided, if the aggregator's databases were noted on the main menu, the number of steps required to get to the databases, if signs were posted, if handouts were available, and the percentage of workstations that allowed the aggregator's databases to be searched. The same tests were run using the same variables but examining usage as a continuous variable. One-way ANOVA and Mann-Whitney tests were also performed for number of workstations and for training, and Spearman's correlation was run for use with number of workstations.

For academic libraries, there were only two instances in which any statistically significant difference was found. One was for the number of workstations provided. In this case, a one-way ANOVA, run on usage per 10,000 students and number of workstations, provided showed statistically significant differences between those that had 20 or fewer and those with 21-50 (significance=.014), but no difference between the latter and those libraries with more than 50 workstations. The second instance of statistical significance occurred when Spearman's rho (correlation test value) displayed a value of .037 when comparing the number of users per 10,000 students and the number of workstations provided. This showed that as the number of workstations provided increases, the usage increases as well. For public libraries, no statistical differences were found among libraries serving different sized populations.

CONCLUSION

Online use in both public and academic libraries follows identifiable patterns that correspond to the workday and to the academic year. Peaks and valleys are clearly identifiable in all types and sizes of libraries. Even for general interest databases, there are many times when no one is searching. Peak usage can be quite high in large libraries, but average use of any one database or group of databases is low.

A combination of factors may influence database use. Librarians hope to influence use of a particular product through instruction, signs, and handouts, but few of these were found to be statistically significant.

The fact that so few statistical differences were found for usage among libraries serving schools and populations of such diverse sizes could lie with a number of possible factors. Differences may not be easily discerned due to the small sample size and, for most of the questions, the fact that the majority of libraries participate in the activity to which the variable referred (e.g., signs, training, remote access, etc.). Factors such as quality, depth and duration of training, type of training, level of student motivation, librarian-student interaction, and quality and utility of signs and interfaces are all qualitative aspects that could affect database usage but were not tested here.

NOTES

1. The hourly data is converted into the local time zone for each library, but, since the database aggregator's computer was set to eastern time, data from libraries in western time zones begin and end earlier (7am-11pm central time; 6am-10pm mountain time; 5am-9pm pacific time; 4am-8pm Alaska time; and 3am-7pm Hawaii time). Times are in the correct clock time (standard or daylight time) for those areas that change to daylight savings time.

REFERENCES

- Abels, E. (1996). Pricing of electronic resources: Interviews with three vendors. *Journal of the American Society for Information Science* 47 (3, 1996): 235-246.
- American Library Association. (1997). *Electronic services in academic libraries*, Chicago: ALA, 1997. [Executive summary available at www.ala.org/alaorg/ors/elecsvcs.html.]
- McGrath, W. E. (1996). Periodicity in academic library circulation: A spectral analysis. *Journal of the American Society for Information Science* 47 (February 1996): 136-145.
- Murfin, M. E. (1983). National reference measurement: What can it tell us about staffing. *College and Research Libraries* 44 (September 1983): 321-333.
- Tenopir, C. (1999a). Factors that influence database use. *Proceedings of the Association of College and Research Libraries (ACRL) Meeting, April 1999*: in press.
- Tenopir, C. (1999b). Influencing database use in academic libraries. *Library Journal* 124 (May 1, 1999): in press.
- Tenopir, C. (1999c). Influencing database use in public libraries. *Library Journal* 124 (June 1, 1999): in press.
- Tenopir, C., Banks, G., and Preston, L. (1998). Pricing options for end user products in libraries. *Proceedings of the National Online Meeting, New York, May 1998*, 419-432.
- Tenopir, C. and Ennis, L. (1998). The digital reference world of academic librarians. *Online* 19 (July/August 1998): 22-28.

Tenopir, C. and Green, D. M. (1999). Simultaneous usage of online databases in academic and public libraries. *Proceedings of the National Online Meeting, New York, May 1999*: in press.

Tolle, J. E. et al. (1983). Determining the required number of online catalog terminals. *Information Technology & Libraries 2* (September 1983): 261-265.

Figure 2. Average number of online catalog terminals required by library type and size.



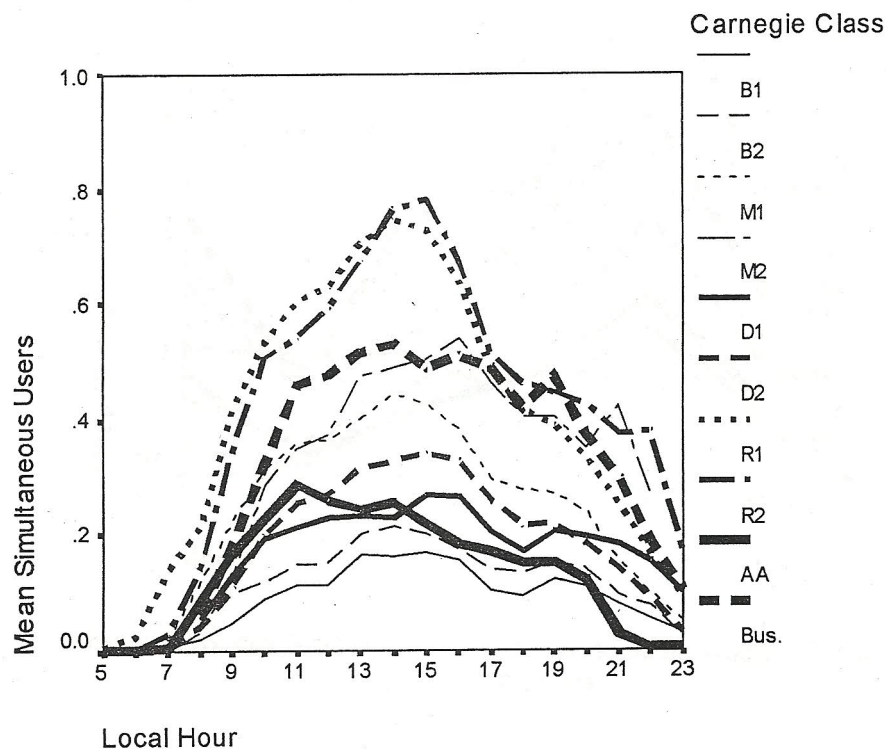


Figure 1. Average Simultaneous Use Patterns by Time of Day Academic Libraries by Carnegie Class

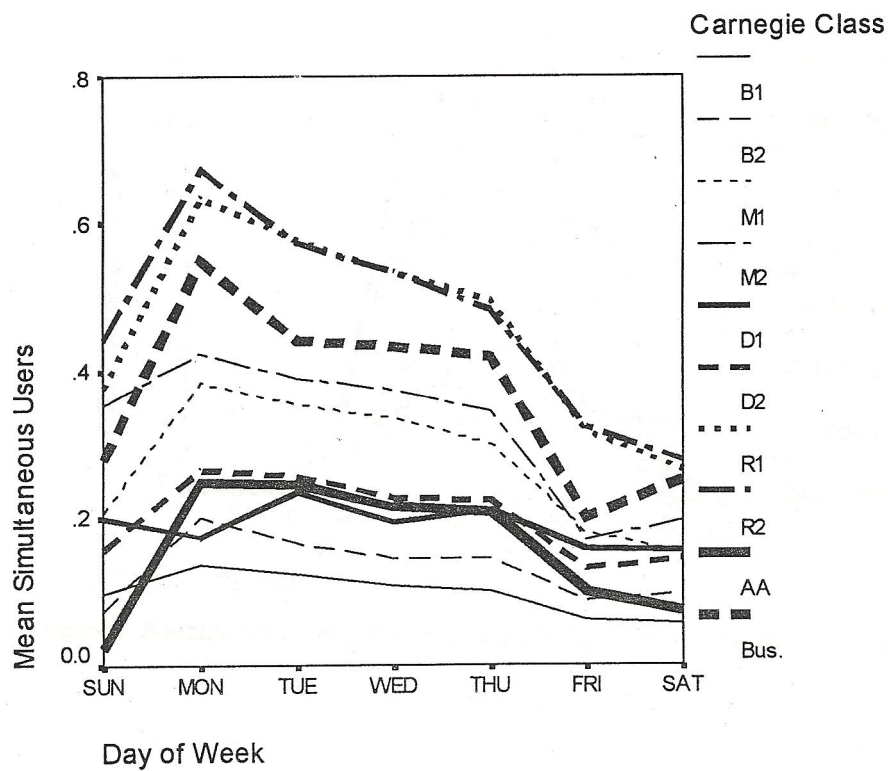


Figure 2. Average Simultaneous Use Patterns by Day of Week Academic Libraries by Carnegie Class

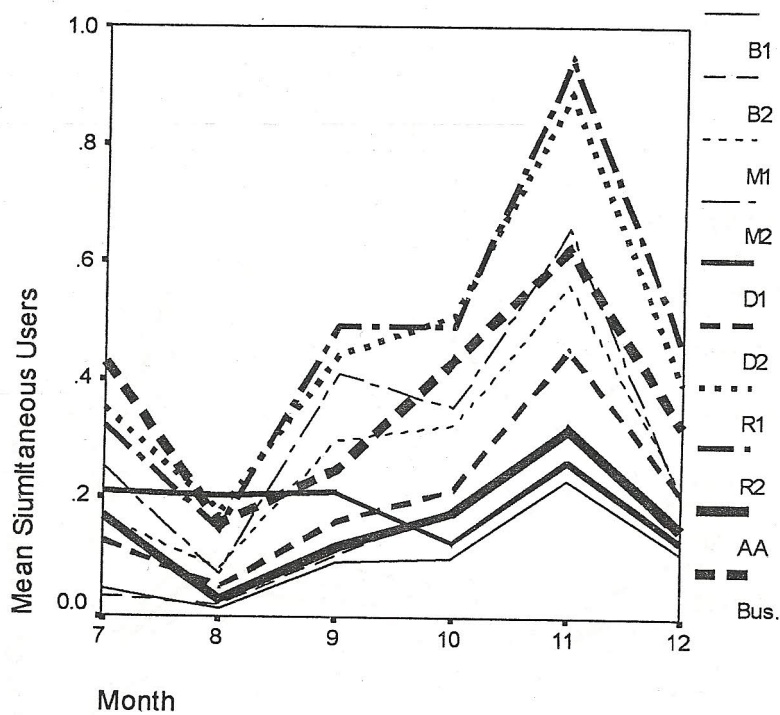


Figure 3. Average Simultaneous Use Patterns by Month, Academic Libraries by Carnegie Class

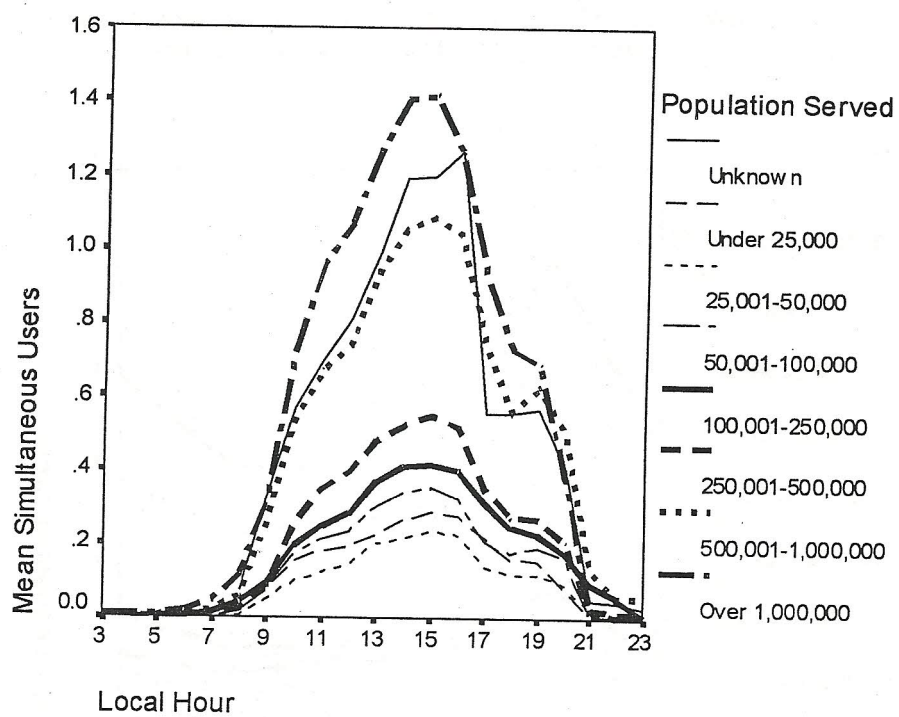


Figure 4. Average Simultaneous Use Patterns By Time of Day, Public Libraries By Size

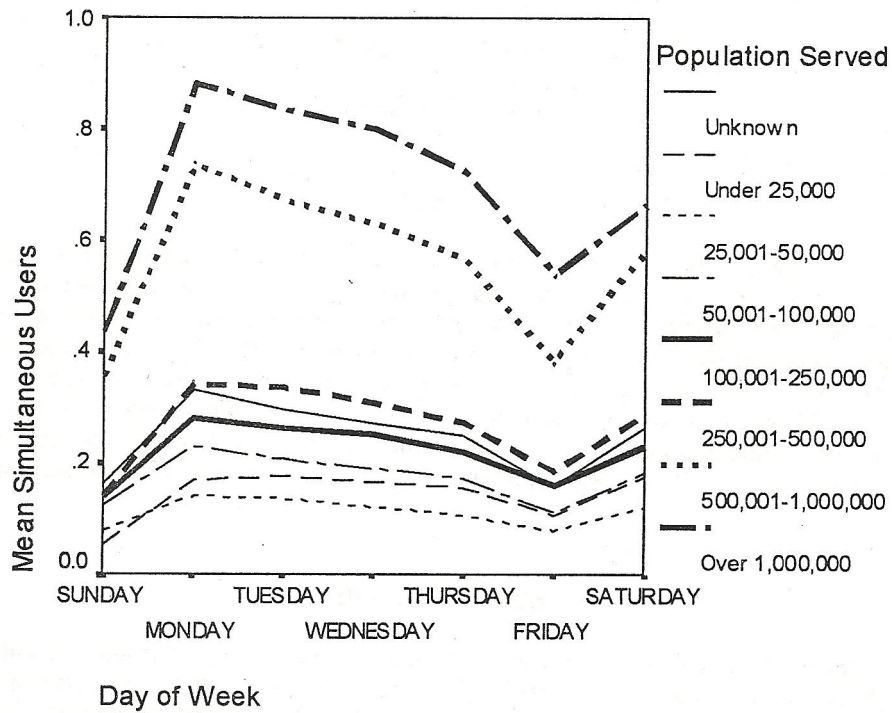


Figure 5. Average Simultaneous Use Patterns By Day of the Week, Public Libraries By Size of Library Community

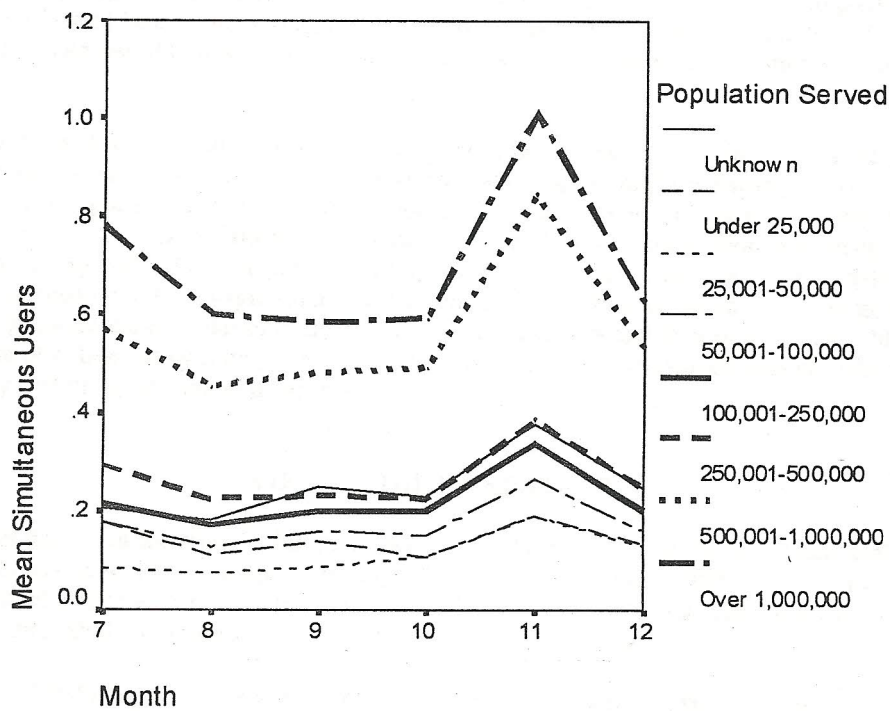


Figure 6. Average Simultaneous Use Patterns By Month (July-December) Public Libraries By Size of Library Community