Fall 2022

The Freedom Collection 2017–2021: Part 1, The Composition of the Freedom Collection and UNCL's Downloads by Member and Subject

David C. Tyler
University of Nebraska-Lincoln, dtyler2@unl.edu

The Freedom Collection 2017–2021: Part 1,
The Composition of the Freedom Collection and
UNCL's Downloads by Member and Subject

Author:
David C. Tyler

Data:
David Macaulay
David C. Tyler

Submitted:
Fall 2022

Table of Contents

Sect. #	Section Subjects	Page
	Introduction	3
	Summary Findings	4
1	Composition of the Freedom Collection	6
2	UNCL Downloading by Member and Subject	12
	References	27

Introduction: The Freedom Collection (Elsevier) and UNCL

For a prior data review, David Macaulay was kind enough to assemble a sizable spreadsheet of Elsevier journal download data via Alma Analytics for the University of Nebraska Consortium of Libraries (UNCL) and its individual members. The download data covered the period from 2017 to 2021 for UNCL as a whole, the University of Nebraska at Kearney (UNK), the University of Nebraska at Omaha (UNO), the University of Nebraska Medical Center (UNMC), and the University of Nebraska – Lincoln (UNL).*

The original data draw was for a full-ish Elsevier title list, excluding zero-use titles not subscribed to nor in Freedom Collection for 2022, and Macaulay's spreadsheets included some UNCL price information for a portion of the title list. The journal subscription prices were taken from Elsevier's 2022 invoice to UNCL and may not necessarily have had a clear relationship to Elsevier's list prices for individual subscriptions, although the relationship for the subscribed-to journals in the aggregate did seem for 2021 and 2022 to have a consistent relationship to the journals' sum subscription prices (i.e., ~150% of list for the group).

After having taken a look at the data for the full title list (3,735 journals), the lead author thought that it might be interesting to take a closer look at the portion of the title list tagged as being part of the Freedom Collection, a discounted collection of journals made available to ScienceDirect Complete academic customers. Readers unfamiliar with the Freedom Collection can read up on it under the "Journal collections" subhead on Elsevier's "Journal title lists" page (see: https://www.elsevier.com/solutions/sciencedirect/journals/journal-title-lists).

When Macauley provided UNCL's Elsevier download data to the lead author, UNCL had not yet, to the lead author's knowledge, formulated and settled upon a set of research questions to put to Elsevier's products. To the best of my knowledge, this situation has not changed, so the lead author has elected to give this subset of the full dataset just a bit of a probing in order to get a better sense of UNCL's utilization of this Elsevier collection. As was the case with the prior report, the reader should keep in mind that the lead author is not particularly knowledgeable where Elsevier journals are concerned and has not been involved at all in any recent Elsevier projects, reviews, deliberations, or negotiations. So, when reading through this report, the reader should understand that this report is not part of any existing deliberations and does not argue towards any particular conclusions. The reader should also understand that the questions posed herein were merely general questions that occurred to the lead author. If the reader has additional questions or would like to plumb any of the questions posed below to a greater depth, the lead author would invite the reader to contact a member of one of the UNCL groups involved in assessing resources and/or in contract negotiations. Furthermore, the lead author is ignorant

---

* Note: the Marvin and Virginia Schmid Law Library IP range is included within the UNL range, so usage from Law is part of UNL's COUNTER reports.

Concerning what limitations there may be on sharing Elsevier data, so the reader should inquire with UNCL whether this or another dataset could be made available for further analysis. Please do feel free contact the lead author (dtyler2@unl.edu) if there are any questions about this particular report.

Regarding the structure of the report: In the report below, table and figure numbers will correspond to the sections in which they appear (e.g., all tables in Section 1 will be Table 1 plus a sequential letter to differentiate them). Tables and figures with the same number-letter combinations, if any, should reference the same data in whole or in part. Hopefully, this approach will keep things fairly straightforward.

Also, the lead author will be employing terms like "downloads" and "usage" interchangeably. Of course, "usage" encompasses a great deal more than does "downloads," much of it unmeasured in this dataset. For ease of discussion, however, the lead author will use both terms freely. To save space and to avoid giving a false impression of precision, the lead author will be rounding most calculated variables to the nearest $1/10^{th}$ or $1/100^{th}$ as appropriate.

Finally, it should be understood that the lead author's summaries and analyses have been derived solely from the data provided and should be understood to be very much tentative and preliminary. David Macaulay contributed substantially to the factual content and to the clarity of this report, but most of the analysis was performed by the lead author, David C. Tyler, so it should also be understood that any errors present likely belong to me alone. Any errors attributable to David Macaulay herein likely resulted from my not knowing enough about Elsevier and UNCL to ask the right questions.

SUMMARY FINDINGS:

This report employs 2017-2021 Elsevier Freedom Collection journal download data for UNCL and its members. For 2022, this collection was comprised of 2,243 journals, 60% of the full title list compiled by David Macaulay.[*] Of the 2,243 journals, 613 (27.3%) were tagged as "subscribed," and David Macaulay was able to enter UNCL 2022 subscription price information for these journals into the spreadsheet. Unfortunately, Macaulay did not have subscription pricing information for each of the remaining 1,630 journals.

In addition to the pricing information that David Macaulay was able to pull together, the lead author, via consulting the WorldCat database, was able to gather Library of Congress (LC) classification/subclassification information for 2,228 of the journals. For the remaining 15 journals, the lead author manually assigned LC classifications/subclassifications based upon

* A separate 2022 Freedom Collection journal title list compiled by Macaulay had 2,258 journals listed, sans download data. The lead author does not know what this slight discrepancy indicates. Perhaps there have been few journal additions or title changes for 2022?

classifications given to journals with similar subject headings. A small number of the journals had multiple LC call numbers assigned to them. Some were within the same subclassification (e.g., *Journal of the World Federation of Orthodontists* was assigned both RK1 and RK520), and these discrepancies were ignored: the journal is in classification R and subclassification RK, regardless of the particular assigned call numbers. Some had call numbers with differing classifications/ subclassifications (e.g., *Computers in Human Behavior* was assigned both BF39.5 and QA76.9.I58). In such instances, both call numbers were recorded. As a result, for this analysis, the Freedom Collection ended up with 2,335 LC call numbers in total, with 92 journals tagged as belonging to either more than one LC classification or subclassification. In cases of conflict, in the report below the first listed WorldCat call number will be treated as though it is the preferred number (e.g., *Computers in Human Behavior* will be treated as though its call number is BF39.5, so its class will be B and its subclass will be BF).

As was noted above, the full dataset was provided without a set of research questions, so the lead author posed some general questions to the dataset on his own. Unfortunately, the data suggested a sizeable number of questions, so the report has been divided into multiple parts to keep the report(s) manageable. The several parts to this report will be distributed as they are completed.

The questions posed in this first part of the Freedom Collection report were as follows:

PART 1:

1. What was the composition of the Freedom Collection?
2. What was the extent of UNCL's and its members' downloading?
3. Did downloading concentrate by member?
4. Were some subjects favored by UNCL?

This first portion of the report addressed question 1-4, with findings are summarized here:

1. The Freedom Collection (FC) would appear to be composed of 2,243 journals. By Library of Congress (LC) Classification, the provides access mainly to journals in the LC classes *Q – Science, R – Medicine, T – Technology, H – Social Sciences,* and *S – Agriculture*, in descending order. Access to modest numbers of journals (i.e., less than 50/LC class) in twelve additional LC Classes is also provided
2. Over the five-year interval, UNCL downloading was substantial (e.g., total = 3,560,291), as was the downloading of two UNCL member (UNL = 2,053,054 and UNMC = 1,068,906)
3. Downloading was substantially concentrated by member (UNL = 57.7% and UNMC = 30.0%)
4. Among the major LC classes, UNCL favored Q, R, and T. The performance of LC Class H, while clearly productive, lagged a bit behind these four. Among the smaller LC classes, UNCL showed strong interest in LC classes B, S, and L (note: FC LC class B was comprised almost entirely of psychology journals)

SECTION 1: Composition of the Freedom Collection

The Excel workbook provided by David Macaulay contained several tabs and a quite a bit of data. A summary of its pertinent spreadsheets appears in Table 1a:

Table 1a: Freedom Collection: Composition of the Dataset

Category*	Journals	Subscriptions	Total Price**	Total Downloads
Freedom Collection	**2,243**	**613**	**$2,517,839.10**	**3,560,291**
B – Philosophy. Psychology. Religion	35	16	$36,027.20	69,893
C – Auxiliary Sciences of History	5	2	$3,792.02	3,752
D – World History . . .	1	0	$0.00	372
G – Geography. Anthro. Recreation	49	7	$29,865.07	56,628
H – Social Sciences	210	61	$148,736.84	257,240
J – Political Science	2	1	$1,872.76	2,256
K – Law	6	1	$826.17	2,799
L – Education	15	8	$11,259.41	54,527
N – Fine Arts	2	0	$0.00	2,092
P – Language and Literature	16	3	$4,203.72	6,887
Q – Science	694	245	$1,174,251.55	1,092,172
R – Medicine	760	154	$491,331.77	1,111,260
S – Agriculture	67	30	$121,294.70	205,919
T – Technology	371	82	$489,799.79	684,635
U – Military Science	1	1	$2,492.19	629
V – Naval Science	1	0	$0.00	86
Z – Bibliography. Library Science. …	8	2	$2,085.91	9,144

* Readers wanting to consult the LC system of classifications can find a guide online at: https://www.loc.gov/catdir/cpso/lcco/
**According to David Macaulay, this amount is the total paid by UNCL for the Elsevier core subscriptions in 2022 (UNCL apparently also paid an additional $165,906 to get access to the remainder of the Freedom Collection). These journal subscription prices were taken from Elsevier's invoice to UNCL and may have no direct relationship with the publisher's posted list prices for individual subscriptions.

As one can see from the table, the Freedom Collection would appear to be comprised mainly of, in descending order, *Q – Science*, *R – Medicine*, *T – Technology*, *H – Social Sciences*, and *S – Agriculture* journals, subscriptions, subscription dollars, and downloads. It might be worthwhile to disaggregate these categories into their LC subclasses to get a better sense of their composition. It probably would also be worthwhile to note that one of the multi-subject categories which may seem an odd fit for Elsevier, *B – Philosophy. Psychology. Religion*, is actually comprised almost entirely of psychology journals (there is just one philosophy journal devoted to applied logic and no religious studies journals).

The composition of the Freedom Collection journals in LC classification Q (Sciences) is as follows:

Table 1b: Freedom Collection: Composition of the Q (Science) Journals

Category*	Journals	Subscriptions	Total Price**	Total Downloads
Freedom Collection	**2,243**	**613**	**$2,517,839.10**	**3,560,291**
Q – Science (General)	17	3	$7,739.39	12,441
QA – Mathematics	104	31	$94,593.17	47,558
QB – Astronomy	10	0	0	1,525
QC – Physics	77	22	$122,481.64	74,291
QD – Chemistry	96	32	$244,343.11	146,992
QE – Geology	48	20	$88,950.20	47,610
QH – Natural history - Biology	105	43	$199,249.20	195,743
QK – Botany	28	8	$34,728.23	56,937
QL – Zoology	27	10	$33,281.75	34,663
QM – Human anatomy	1	0	0	446
QP – Physiology	131	55	$258,026.35	359,540
QR – Microbiology	50	21	$90,858.52	114,426
Q TOTAL	**694**	**245**	**$1,174,251.55**	**1,092,172**

* Readers wanting to consult the LC system of classifications can find a guide online at: https://www.loc.gov/catdir/cpso/lcco/
**According to David Macaulay, this amount is the total paid by UNCL for the Elsevier core subscriptions in 2022 (UNCL apparently also paid an additional $165,906 to get access to the remainder of the Freedom Collection). These journal subscription prices were taken from Elsevier's invoice to UNCL and may have no direct relationship with the publisher's posted list prices for individual subscriptions.

As one can see, the Q LC class is mainly comprised of physiology, biology, and mathematics journals (the latter includes some computer and information science journals, as well). Each subclassification comprises around 15% of the Science category. The next largest LC subclasses, chemistry and physics, each comprise a bit more than 10% of the Q – Science LC classification. The rest of the subclasses each contribute less than 10% of the classification's journal list. All twelve of the LC subclassifications in class Q contributed journals to the Freedom Collection.

The composition of the Freedom Collection journals in LC classification R (Medicine) is as follows:

Table 1c: Freedom Collection: Composition of the R (Medicine) Journals

Category*	Journals	Subscriptions	Total Price**	Total Downloads
Freedom Collection	**2,243**	**613**	**$2,517,839.10**	**3,560,291**
R – Medicine (General)	74	12	$43,034.53	132,375
RA – Public aspects of medicine	63	17	$76,061.98	139,007
RB – Pathology	27	8	$18,030.87	33,852
RC – Internal medicine	278	57	$194,544.31	392,726
RD – Surgery	96	17	$28,218.24	92,995
RE – Ophthalmology	13	2	$2,126.00	4,164
RF – Otorhinolaryngology	9	1	$7,452.83	6,355
RG – Gynecology and obstetrics	33	2	$5,881.56	26,937
RJ – Pediatrics	30	5	$5,478.05	45,369
RK – Dentistry	18	8	$15,258.10	14,710
RL – Dermatology	8	2	$2,813.63	12,192
RM – Therapeutics. Pharmacology	57	8	$38,532.07	94,400
RS – Pharmacy and materia medica	24	5	$44,880.92	77,403
RT – Nursing	26	10	$9,018.68	38,476
RZ – Other systems of medicine	4	0	0	299
R TOTAL	**760**	**154**	**$491,331.77**	**1,111,260**

* Readers wanting to consult the LC system of classifications can find a guide online at: https://www.loc.gov/catdir/cpso/lcco/
**According to David Macaulay, this amount is the total paid by UNCL for the Elsevier core subscriptions in 2022 (UNCL apparently also paid an additional $165,906 to get access to the remainder of the Freedom Collection). These journal subscription prices were taken from Elsevier's invoice to UNCL and may have no direct relationship with the publisher's posted list prices for individual subscriptions.

As one can see, the R LC class is mainly comprised of internal medicine (which includes a number of psychiatry journals). The only other LC subclassification with more than 10% of the category's journals is surgery. Of the seventeen LC subclassifications that comprise the R – Medicine category, two contributed no journals to the Freedom Collection: *Subclass RV Botanic, Thomsonian, and eclectic medicine* and *Subclass RX Homeopathy*.

The composition of the Freedom Collection journals in LC classification T (Technology) is as follows:

Table 1d: Freedom Collection: Composition of the T (Technology) Journals

Category*	Journals	Subscriptions	Total Price**	Total Downloads
Freedom Collection	**2,243**	**613**	**$2,517,839.10**	**3,560,291**
T – Technology (General)	21	7	$37,384.39	28,152
TA – Engineering (General). Civil engineering	104	29	$170,183.20	163,491
TC – Hydraulic engineering. Ocean engineering	7	1	$3,860.30	3,846
TD – Environmental technology. Sanitary engineering	21	4	$34,800.64	72,302
TE – Highway engineering. Roads and pavements	1	0	0	487
TF – Railroad engineering and operation	2	0	0	92
TH – Building construction	5	0	0	10,871
TJ – Mechanical engineering and machinery	28	7	$38,779.81	52,841
TK – Electrical engineering. Electronics. Nuclear engineering	41	6	$42,559.08	38,426
TL – Motor vehicles. Aeronautics. Astronautics	12	1	$1,705.17	3,355
TN – Mining engineering. Metallurgy	21	3	$22,625.36	26,623
TP – Chemical technology	77	15	$101,391.03	189,098
TS – Manufactures	19	4	$9,965.03	40,832
TX – Home economics	12	5	$26,545.78	54,219
T TOTAL	**371**	**82**	**$489,799.79**	**684,635**

* Readers wanting to consult the LC system of classifications can find a guide online at: https://www.loc.gov/catdir/cpso/lcco/

**According to David Macaulay, this amount is the total paid by UNCL for the Elsevier core subscriptions in 2022 (UNCL apparently also paid an additional $165,906 to get access to the remainder of the Freedom Collection). These journal subscription prices were taken from Elsevier's invoice to UNCL and may have no direct relationship with the publisher's posted list prices for individual subscriptions.

As one can see, the T LC class is mainly comprised of general engineering and civil engineering. The only other LC subclassifications with more than 10% of the category's journals were chemical technology and electrical engineering, electronics, and nuclear engineering. Of the seventeen LC subclassifications that comprise T – Technology, three contributed no journals to the Freedom Collection: *Subclass TG Bridge engineering*, *Subclass TR Photography*, and *Subclass TT Handicrafts. Arts and crafts*.

The composition of the Freedom Collection journals in LC classification H (Social Sciences) is as follows:

Table 1e: Freedom Collection: Composition of the H (Social Sciences) Journals

Category*	Journals	Subscriptions	Total Price**	Total Downloads
Freedom Collection	**2,243**	**613**	**$2,517,839.10**	**3,560,291**
H – Social sciences (General)	11	2	$2,884.75	12,599
HA – Statistics	2	0	0	486
HB – Economic theory. Demography	25	11	$26,328.08	14,768
HC – Economic history and conditions	15	6	$16,953.32	22,879
HD – Industries. Land use. Labor	50	10	$20,014.80	51,916
HE – Transportation and communications	16	3	$8,727.15	9,935
HF – Commerce	34	12	$23,101.76	34,400
HG – Finance	19	5	$20,459.28	29,478
HJ – Public finance	1	1	$4,862.29	2,886
HM – Sociology (General)	5	2	$3,680.89	11,799
HN – Social history and conditions. Social problems. Social reform	1	0	0	31
HQ – The family. Marriage. Women	5	0	0	2,297
HT – Communities. Classes. Races	9	2	$4,238.34	9,371
HV – Social pathology. Social and public welfare. Criminology	17	7	$17,486.18	54,395
H TOTAL	**210**	**61**	**$148,736.84**	**257,240**

* Readers wanting to consult the LC system of classifications can find a guide online at: https://www.loc.gov/catdir/cpso/lcco/
**According to David Macaulay, this amount is the total paid by UNCL for the Elsevier core subscriptions in 2022 (UNCL apparently also paid an additional $165,906 to get access to the remainder of the Freedom Collection). These journal subscription prices were taken from Elsevier's invoice to UNCL and may have no direct relationship with the publisher's posted list prices for individual subscriptions.

As one can see, the H LC class is mainly comprised of journals devoted to business and to economics, with journals in LC subclasses HB-HJ making up just over 75% of the classification's journals. Elsevier's Freedom Collection would seem to have little interest in sociology and its related fields, with the exception of subclass *HV Social pathology. Social and public welfare. Criminology*, which made up 8.1% of the H class journals. Of the sixteen LC subclassifications that comprise *H – Social Sciences*, two contributed no journals to the Freedom Collection: *Subclass HS Societies: secret, benevolent, etc*. and *Subclass HX Socialism. Communism. Anarchism.*

The composition of the Freedom Collection journals in LC classification S (Agriculture) is as follows:

Table 1f: Freedom Collection: Composition of the S (Agriculture) Journals

Category*	Journals	Subscriptions	Total Price**	Total Downloads
Freedom Collection	**2,243**	**613**	**$2,517,839.10**	**3,560,291**
S Agriculture (General)	17	10	$51,093.09	100,208
SB Plant culture	13	8	$22,698.26	45,589
SD Forestry	2			5,435
SF Animal culture	31	12	$47,503.35	47,354
SH Aquaculture. Fisheries. Angling	4			7,333
S TOTAL	**67**	**30**	**$121,294.70**	**205,919**

* Readers wanting to consult the LC system of classifications can find a guide online at: https://www.loc.gov/catdir/cpso/lcco/

**According to David Macaulay, this amount is the total paid by UNCL for the Elsevier core subscriptions in 2022 (UNCL apparently also paid an additional $165,906 to get access to the remainder of the Freedom Collection). These journal subscription prices were taken from Elsevier's invoice to UNCL and may have no direct relationship with the publisher's posted list prices for individual subscriptions.

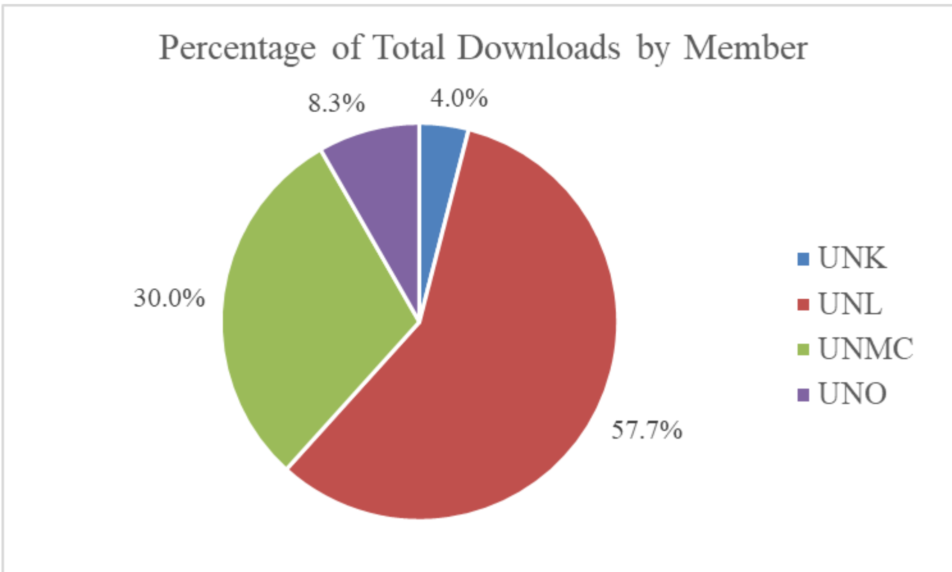
As one can see, the S LC class is largely (46.2%) comprised of journals devoted animal culture, with the bulk of the remainder given over to general agriculture and to plant culture. Elsevier's Freedom Collection would seem to give little attention to forestry or to aquaculture. Of the six LC subclassifications that comprise *S – Agriculture*, only one contributed no journals to the Freedom Collection: *SK – Hunting sports*.

SECTION 2: UNCL Downloading by Member and Subject

Two variables usually of interest in reports of this sort are usage and subject. A download summary for the 2017-2021 interval is presented in Table 2a below:

Table 2a: Freedom Collection Downloads: by Year and by Member

	Year					
	2017	2018	2019	2020	2021	Total
UNCL	667,079	768,052	765,406	661,968	697,786	3,560,291
UNK	24,317	25,758	31,309	25,761	35,276	142,421
UNL	400,679	445,867	417,093	356,040	433,375	2,053,054
UNMC	224,164	235,106	238,875	220,834	149,927	1,068,906
UNO	17,919	61,321	78,129	59,333	79,208	295,910



As can be seen from the table, members UNL and UNMC were responsible for almost all of the Freedom Collection's download activity over the interval. The author does not have the UNCL members' faculty and student numbers handy and so cannot calculate downloads-per-patron rates, but he suspects that UNL's and UNMC's level of activity may be high. This suspicion, of course, rests on the assumption that UNL is not, for example, roughly seven times as large as UNO.

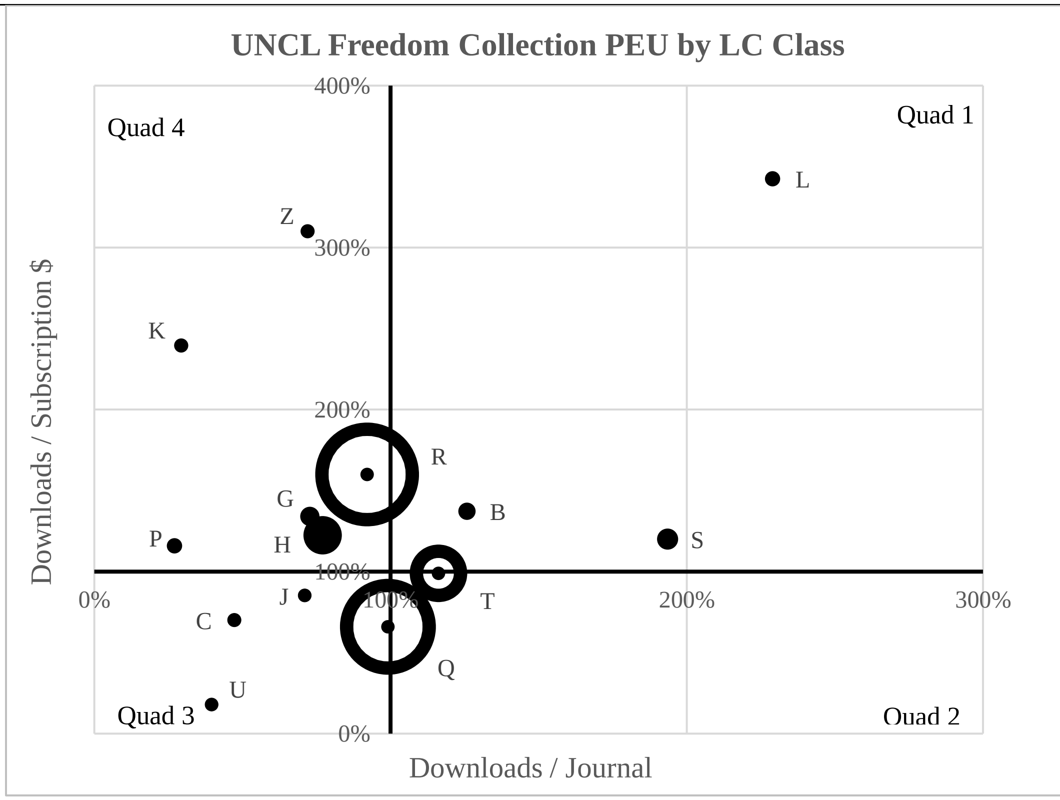
One of the questions that is often of the most interest in reports such as this is the question of whether resource usage and dollars are well allocated and aligned. One approach to this question that, as Harker (2022) has remarked, "has become ubiquitous and nearly universal for evaluating resources, especially renewable resources such as journals and databases" is the cost-per-use (CPU) metric (355). This metric has become popular no doubt in part because of its ease of calculation, as well as, as Harker has noted in her article on the history of CPU, because of the increased availability of usage data (Harker cites Brown & Stowers 2013, here) and because of ongoing pressures within the profession for libraries to demonstrate that they are cost effective and/or cost efficient (Harker cites Gilchrist 1971 and Walters 2016). Although there has been, as Harker has noted, some criticism of CPU (Harker cites Henderson 1992; in a report on UnSub, Tyler, Macaulay, and McClanahan [2021] also cautioned against a too uncritical employment of CPU), the author will employ CPU here due to its familiarity and ubiquity. Additionally, in order to demonstrate the extent of use in each LC class, the author will report a download-per-journal (DPJ) metric. Lastly, because many in the field seem to favor visual representations of data, the author will employ another handy approach: the calculation and plotting of relative use factors. Relative use factors offer a quick and easily interpretable assessment metric that, as Baker and Wallace (2002) have noted, allows one to "indicate a deviation from expected behavior" (208). Bonn (1974) may have been the first to introduce the use factor into the library literature as a "ratio of use to holdings in specific subject classes" (272). He suggested that libraries might, for example, calculate the "[p]roportionate circulation statistics by subject class compiled over a definite period … compared with proportionate holdings statistics by subject class" (272-273). Mills, in 1982, slightly modified Bonn's relative use factor and produced the Percentage Expected Use (PEU) metric, which presents the "ratio of the percentage of use of a subject to its percentage holdings" (5). The author, here, will employ Mills' PEU simply because experience has suggested that readers find it easier to intuitively interpret.

In Table 2b on the following page, the reader can see how "cost efficient" were the LC classes of the Freedom Collection via CPU, how active were the classes' journals (DPJ) as a group, and how proportional were the per-journal (X axis) and per-subscription-dollar (Y axis) downloads of the classes (PEU). Regarding PEU, the X axis should provide a relative indicator of how much the journals in each class were used per journal, and the Y axis should provide a relative indicator of how much the journals in each class were used per subscription dollar spent. One should read the PEU graph by starting in the upper right-hand corner and proceeding clockwise around the 100%-100% axis, which intersection indicates perfect proportionality. The farther a datapoint is from an axis, the greater the disproportionality of its activity. Thus, Quad 1 contains LC classes whose per-journal and per-subscription-dollar activity was higher than would be expected; Quad 2 contains LC classes whose per-journal activity was higher than expected but whose per-subscription-dollar activity was lower; Quad 3 contains LC classes whose per-journal and per-subscription-dollar activity was lower than would be expected; and Quad 4 contains LC classes whose per-journal activity was lower than expected but whose per-subscription-dollar activity was higher.

When reviewing Table 2b, the reader should be cautious and should keep in mind, where cost data were employed, that David Macaulay was able to produce UNCL subscription prices for just 27.3% of Freedom Collection journals, so the numbers reported likely are not entirely correct and positions along the Y axis are tentative. Also, the reader should keep in mind Gerd Gigerenzer's (2002) caution concerning relative measures: the impact of equal outputs can be greatly distorted by differences in category size, such that adding a single download to a tiny category's data could produce a sizable change in its outcome numbers while the effect of adding that same download to a big category would pass unnoticed. To ameliorate potential distortions, the author has made the datapoints vary in size in accord with the proportions of Freedom Collection journals contained in each LC class. Thus, larger datapoints should be taken as likely having more predictive value.

Table 2b: UNCL's Elsevier Freedom Collection Activity (2017-2021) by LC Class: Downloads-per-Journal (DPJ), Cost-per-Use (CPU), and Percentage Expected Use (PEU)

LC Class	DPJ	CPU*	LC Class	DPJ	CPU*	LC Class	DPJ	CPU*
B	1,996.9	$0.52	K	466.5	$0.30	S	3,073.4	$0.59
C	750.4	$1.01	L	3,635.1	$0.21	T	1,845.4	$0.72
D	372.0	n/a	N	1,046.0	n/a	U	629.0	$3.96
G	1,155.7	$0.53	P	430.4	$0.61	V	86.0	n/a
H	1,225.0	$0.58	Q	1,573.7	$1.08	Z	1,143.0	$0.23
J	1,128.0	$0.83	R	1,462.2	$0.44			



*Note: LC classes D, N, and V have no CPU values to report due to their having no journals with subscription cost data. Thus, they could not be included in the scatterplot.

As one can see from the table, among the minor LC classes of the Freedom Collection, the fifteen journals that comprise the LC class *L – Education* would seem to be an exceptional deal for UNCL, relative to the performances of the other LC classes. LC class *B – Philosophy. Psychology. Religion.* would also seem to be a good performer, although not to the extent that education was. The performances of the rest of the minor LC classes would suggest that a review might be warranted. Among the LC classes of middling size, LC Class *S – Agriculture* would seem to be an exceptional performer for UNCL in terms of downloads per journal, although its CPU was only slightly better than (i.e., lower than) the Freedom Collection's CPU of $0.71. LC class *H – Social Sciences* might warrant a bit of a review, unfortunately, for although its downloads per subscription dollar numbers were good, its downloads per journal numbers were relatively poor. One likely would find a number of journals among the social science titles whose download performances unimpressive. Lastly, among the Freedom Collection's major LC classes, *Q – Science* produced roughly as many downloads per journal as it should have, but its downloads per subscription dollar were somewhat poor; *R – Medicine* produced almost as many downloads per journal as it ought to have, and its downloads per subscription dollar were, surprisingly, quite good; and *T – Technology* produced slightly more downloads per journal than it ought to have, while producing roughly proportional downloads per subscription dollar.

Thus, it would appear that some of the LC classes and their Freedom Collection journals were of more interest to UNCL. If the consortium were to pursue alternative subscription arrangements, the journals from LC classes B, L, Q, R, S, and T would likely be of great interest to UNCL as a whole, and the journals from the rest of the LC classes would likely warrant a somewhat more critical and in-depth review.

To determine whether these apparent differences among the LC classes' performances might be real would require a bit more formal analysis and testing, which will follow. This would mark the point where Section 2 becomes rebarbative (i.e., unattractive and objectionable) for most readers, and those content with reviewing the more intuitive usage metrics and relative use factors reported on above should probably skip ahead to the next section, which will look into the UNCL members' subject usage profiles.

To get a sense of the LC classes and of UNCL's usage of their journals, the reader may review Table 2c on the next page, which presents some of the descriptive statistics commonly employed in informal analyses. As one can see, within the Freedom Collection dataset, there is considerable variability present in the numbers of journals in each LC class, in the numbers of downloads that each LC class produced, and in the size of the LC classes' Min.-Max. ranges. Most importantly for this report's purposes, there were also sizeable differences between LC classes' average downloads, which suggests that there may be statistically significant differences in performance. Unfortunately, there was also tremendous amounts of variability within the bounds of several of the LC classes. As a result, the standard deviation of the Freedom Collection itself and the standard deviations of ten of the fourteen LC classes for which standard deviations could be calculated were larger than their corresponding means, which suggests the download data, from the perspective of most familiar statistical tests, is unacceptably noisy and misshapen (i.e., the FC data may be a bit of a headache).

Table 2c: UNCL's Elsevier Freedom Collection Activity (2017-2021) by LC Class: Descriptive Statistics (Summed Responses)

LC Class	Journals	Downloads	Min.	Max.	Avg.	StdDev*
B	35	69,893	12	14,050	1,996.9	3,050.5
C	5	3,752	36	2,330	750.4	927.9
D	1	372	372	372	372.0	n/a
G	49	56,628	0	15,185	1,155.7	2,348.0
H	210	257,240	0	15,049	1,225.0	2,162.9
J	2	2,256	923	1,333	1,128.0	289.9
K	6	2,799	23	1,653	466.5	604.0
L	15	54,527	3	10,943	3,635.1	3,251.3
N	2	2,092	467	1,625	1,046.0	818.8
P	16	6,887	25	1,191	430.4	373.4
Q	694	1,092,172	0	19,795	1,573.7	2,483.7
R	760	1,111,260	0	44,146	1,462.2	3,229.9
S	67	205,919	0	17,886	3,073.4	3,818.9
T	371	684,635	0	24,664	1,845.4	3,179.9
U	1	629	629	629	629.0	n/a
V	1	86	86	86	86.0	n/a
Z	8	9,144	51	3,072	1,143.0	1,252.0
Totals	**2,243**	**3,560,291**	**0**	**44,146**	**1,587.3**	**2,907.5**

*LC Classes with a single journal have no calculable standard deviations

The dataset summarized in the table above is largely comprised of counts, and the output variable (i.e., Downloads) appears to be overdispersed. Given the large ranges and the positions of the Maximum values relative to their means, the output variable also appears to suffer rightward skewness. Unfortunately, as was noted above, this all suggests that the data may not be amenable to more familiar analytical approaches because of these statistical tests' restrictive assumptions. So, the author will first run a more familiar test, an Analysis of Variance (ANOVA), to compare the means of the LC classes, and then will follow up with a regression analysis that will compare the slopes of the lines that can be drawn through the LC classes' data in hopes that the regression analysis will agree with the impressions generated by the ANOVA. Negative Binomial regression will be employed because it has less restrictive assumptions and is a standard method used to model and analyze overdispersed count data via the Generalized Linear Model [GLZM] [Hilbe, 2007; Hoffmann, 2004; Orme & Combs-Orme, 2009]). That said, the author would here like to remind the reader that he is very much not a statistician, so there may well be superior approaches of which the author is unaware.*

---

* The author's process at arriving at NB regression was as follows: the unconditional mean of the outcome variable for the dataset as a whole would be much lower than its variance, if calculated, given that the mean is lower than its standard deviation. Similarly, regarding the conditional means and variances (i.e., for the LC classes), in several instances the variances within each category would be much larger than their means. This suggests that overdispersion is present and that a NB model might be appropriate. Just to be certain, the author attempted to employ Poisson regression, another frequently-used approach for discrete dependent variables. With the ancillary parameter set to 0, the log-likelihood goodness-of-fit statistic for Poisson regression was -3,347,996.991, which indicates a quite poor fit, and a simultaneously run Lagrange Multiplier test did indicate overdispersion. Ergo, NB regression.

In order for the analysis to work properly, there would need to be a minimum number of respondents in each group. Unfortunately, ten of the seventeen LC classes that comprise the Freedom Collection have very few journals. To remedy this, the author grouped all LC classes whose journal counts made up less than 1% of the Freedom Collection's total into a single Minor LC classes group. For this analysis, the LC classes that comprise this group are, as follows: C, D, J, K, L, N, P, U, V, and Z. Readers requiring a refresher concerning the subject matter covered by these LC classes are invited to review Table 1a above or to access the LC classification outline online: https://www.loc.gov/catdir/cpso/lcco/

With this combined grouping of minor LC classes in place, Table 2c above would appear as follows:

Table 2c-alternate: UNCL's Elsevier Freedom Collection Activity (2017-2021) by LC Class (Reduced): Descriptive Statistics (Summed Responses)

LC Class	Journals	Downloads	Min.	Max.	Avg.	StdDev*
B	35	69,893	12	14,050	1,996.9	3,050.5
G	49	56,628	0	15,185	1,155.7	2,348.0
H	210	257,240	0	15,049	1,225.0	2,162.9
Q	694	1,092,172	0	19,795	1,573.7	2,483.7
R	760	1,111,260	0	44,146	1,462.2	3,229.9
S	67	205,919	0	17,886	3,073.4	3,818.9
T	371	684,635	0	24,664	1,845.4	3,179.9
Minor LC	57	82,544	3	10,943	1,448.1	2,190.0
Totals	**2,243**	**3,560,291**	**0**	**44,146**	**1,587.3**	**2,907.5**

Analysis of Variance (ANOVA)

For the one-way Analysis of Variance (ANOVA), the first step in analyzing the potential effects on the dependent variable Downloads would be to test whether or not there are any statistically significant differences present in the dataset, with the tested null hypothesis being that there are no differences among the LC classes. To read Table 2d below, begin with the portion of the table devoted to the ANOVA, which tests for the presence of statistically significant differences between group means given the variability between and within groups. The F ratio and the p value (i.e., Sig.) indicate that we should reject the null hypothesis of there being no statistically significant differences in the dataset.[*]

The next step would be to move on to post-hoc head-to-head comparisons to identify where significant differences are present. Usually, this would involve employing the charmingly named Tukey's Honest Significant Difference test, but Tukey's has restrictive assumptions that this dataset violates (e.g., samples of equal size, normality, equal variances, etc). Just to test the water, the author ran a test of homogeneity of variance (Levene's Test for Equality of Variances), and the result (Levene [7, 2,235] = 4.669; $p < .0005$) indicated that normal post-hoc

* For this ANOVA, the critical F value should be around 2.01-ish, and the p value for significance was set to $p = 0.05$. Since the calculated F ratio is greater and the p value is smaller, we reject the null hypothesis.

procedures that assume equal sample sizes, normality and/or equal variances should not be employed. Instead, Dunnet's C Test (1980) was employed. Dunnet's C is recommended when sample sizes and degrees of freedom are large (Shingala & Rajyaguru, 2015) and variances are not necessarily equal. Also, it is reputedly robust to non-normality (Day & Quinn, 1989).

Table 2d: UNCL's Elsevier Freedom Collection Activity (2017-2021) by LC Class: Analysis of Variance (Summed Responses)

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.±
Between LC	228385353.7	7	32626479.10	3.894	.000
Within LC	1.872E+10	2235	8377669.317		
Total	1.895E+10	2242			

Post-Hoc Tests (Dunnett's C)

LC Class (I)	LC Class (J)	Mean Diff. (I-J)	S.E.	95% C.I. Lower	95% C.I. Upper
B	G	841.269	615.124	-1132.84	2815.38
	H	771.990	536.792	-953.21	2497.19
	Q	423.208	524.172	-1264.95	2111.37
	R	534.759	528.767	-1166.53	2236.04
	S	-1076.475	695.372	-3290.01	1137.06
	T	151.565	541.409	-1586.40	1889.53
	Minor	548.803	591.614	-1348.99	2446.59
G	H	-69.279	367.132	-1226.00	1087.45
	Q	-418.061	348.421	-1518.69	682.57
	R	-306.511	355.296	-1427.21	814.19
	S	**-1917.744***	574.616	-3723.87	-111.62
	T	-689.704	373.851	-1865.43	486.02
	Minor	-292.467	443.451	-1693.65	1108.72
H	Q	-348.782	176.538	-888.23	190.67
	R	-237.232	189.746	-816.59	342.12
	S	**-1848.466***	489.848	-3378.69	-318.24
	T	-620.425	222.560	-1300.22	59.37
	Minor	-223.188	326.216	-1244.31	797.93
Q	R	111.551	150.383	-345.57	568.67
	S	**-1499.683***	475.986	-2987.94	-11.43
	T	-271.642	190.116	-850.80	307.52
	Minor	125.595	305.005	-831.50	1082.69
R	**S**	**-1611.234***	481.041	-3114.40	-108.07
	T	-383.193	202.441	-999.70	233.31
	Minor	14.044	312.836	-966.07	994.16
S	T	1228.041	494.904	-316.61	2772.69
	Minor	1625.278	549.376	-97.15	3347.71
T	Minor	397.237	333.759	-645.37	1439.85

±Note: Sig. = .000 indicates $p < .0005$
*Indicates $p \leq .05$; significant differences listed only once in **bold**; positive Mean Diff. indicates the LC class on the left had superior performance, and negative Mean Diff. indicates the LC class on the right had superior performance.

From Table 2d above, it would seem that, although the main test indicated that the null hypothesis should be rejected, all that the post hoc test could comfortably affirm for UNCL's download totals is that over the five years UNCL used the *S – Agriculture* journals quite a lot. Table 2d may be an instance where the statistical analysis was overwhelmed. One-way ANOVA is normally quite robust to violations of its assumptions, hence its frequent use. But the Freedom Collection dataset – with its LC class counts of very different sizes and wild variability – may have so violated ANOVA's assumptions that the test was rendered somewhat useless, as evidenced by the enormous confidence intervals (C.I.) that the LC classes' journals' summed download totals produced.

Before moving on to approaching the data with a different statistical technique, the author thought that it might be worthwhile to look at the data again. But instead of testing for differences among the LC classes' journals' summed UNCL download totals, the author thought he would test for LC class differences using the journals' and UNCL members' individual download counts in each year. That is to say, instead of comparing the sums of the journals' downloads, the author would employ the journals' individual data cells. The results of the ANOVA would then be the product of an analysis of 44,860 individual counts (i.e., 2,243 journals x 4 UNCL members x 5 years), rather than of 2,243 summations. If the results of this analysis, presented in Table 2e below, match those presented in Table 2d above, then the reader should be more inclined to accept that *S – Agriculture* actually was the only standout LC class for UNCL as a whole. If not, then a closer look may be warranted.

As one can see from perusing Table 2e below, handling the data at the level of the individual response, rather than of the summed totals, provided a picture of the performances of the LC Classes that is closer to their actual behavior from year to year. With respect to the statistical

Table 2e: UNCL's Elsevier Freedom Collection Activity (2017-2021) by LC Class: Descriptive Statistics (Individual Responses)

LC Class	Journals	Responses	Downloads	Min.	Max.	Avg.	StdDev
B	35	700	69,893	0	1,786	99.8	220.6
C	5	100	3,752	0	576	37.5	99.6
D	1	20	372	0	90	18.6	27.8
G	49	980	56,628	0	3,463	57.8	229.3
H	210	4,200	257,240	0	3,203	61.2	184.1
J	2	40	2,256	0	309	56.4	74.6
K	6	120	2,799	0	178	23.3	37.9
L	15	300	54,527	0	3,047	181.8	325.0
N	2	40	2,092	0	355	52.3	87.3
P	16	320	6,887	0	334	21.5	40.8
Q	694	13,880	1,092,172	0	3,821	78.7	213.9
R	760	15,200	1,111,260	0	11,070	73.1	275.4
S	67	1,340	205,919	0	4,280	153.7	441.3
T	371	7,420	684,635	0	6,942	92.3	335.2
U	1	20	629	0	111	31.5	34.4
V	1	20	86	0	25	4.3	7.5
Z	8	160	9,144	0	664	57.2	107.9
Total	**2,243**	**44,860**	**3,560,291**	**0**	**11,070**	**79.4**	**265.5**

analysis, this approach should provide the post-hoc tests with considerably more responses and smaller standard deviations, which should allow us to fairly accurately analyze some of the LC Classes that had been absorbed into the Minor LC category above and which should produce tighter confidence intervals, which in turn should produce a more accurate picture of which LC Classes outperformed which during the interval.

Unfortunately, even with this approach, five of the seventeen LC Classes had fewer than one hundred responses, so these classes (D, J, N, U, and V) will still have to be combined into a Minor LC category. Thus, the adjusted Table 2e will look like the following:

Table 2e-alternate: UNCL's Elsevier Freedom Collection Activity (2017-2021) by LC Class (Reduced): Descriptive Statistics (Individual Responses)

LC Class	Journals	Responses	Downloads	Min.	Max.	Avg.	StdDev*
B	35	700	69,893	0	1,786	99.8	220.6
C	5	100	3,752	0	576	37.5	99.6
G	49	980	56,628	0	3,463	57.8	229.3
H	210	4,200	257,240	0	3,203	61.2	184.1
K	6	120	2,799	0	178	23.3	37.9
L	15	300	54,527	0	3,047	181.8	325.0
P	16	320	6,887	0	334	21.5	40.8
Q	694	13,880	1,092,172	0	3,821	78.7	213.9
R	760	15,200	1,111,260	0	11,070	73.1	275.4
S	67	1,340	205,919	0	4,280	153.7	441.3
T	371	7,420	684,635	0	6,942	92.3	335.2
Z	8	160	9,144	0	664	57.2	107.9
Minor	7	140	5,435	0	355	38.8	66.0
Total	**2,243**	**44,860**	**3,560,291**	**0**	**11,070**	**79.4**	**265.5**

As the table shows, the number of LC Classes and journals comprising the Minor LC category with the current framing of the data have been reduced. Hopefully, as was mentioned above, this alternate arrangement of the data and use of individual responses will provide a clearer picture of the potential by-LC Class performance differences present in the dataset.

As was the case above, for the one-way Analysis of Variance (ANOVA), the first step in looking for potential effects on the dependent variable Downloads is to test whether or not there are any statistically significant differences present in the dataset, with the tested null hypothesis being that there are no differences among the LC classes. To read Table 2f below, again begin with the portion of the table devoted to the ANOVA, which tests for the presence of statistically significant effects among the independent variables (i.e., improvements to the fit of the model) given the dataset's mean value and variability between and within groups. Once again, the F ratio and the p value (i.e., Sig.) indicate that we should reject the null hypothesis (i.e., H_0 = no statistically significant differences between the means of the LC classes and the dataset). The next step would be to move on to post-hoc head-to-head comparisons to identify where significant differences are present. Again, just to test the water, the author ran a test of homogeneity of variance, and the result (Levene [12, 44,847] = 52.341; $p < .0005$) indicated that normal post-hoc procedures should not be employed. So, Dunnett's C Test (1980) was again employed.

Table 2f: UNCL's Elsevier Freedom Collection Activity (2017-2021) by LC Class: Analysis of Variance (Individual Responses)

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.±
Between LC	16441005.52	12	1370083.794	19.533	.000
Within LC	3145594464	44847	70140.577		
Total	3162035469	44859			

Post-Hoc Tests (Dunnett's C)¥

LC Class (I)	LC Class (J)	Mean Diff. (I-J)*	S.E.	95% C.I. Lower	95% C.I. Upper
B	C	62.3271	12.9879	18.610	106.044
	G	42.0635	11.0995	5.181	78.946
	H	38.5995	8.8090	9.324	67.875
	K	76.5221	9.0274	46.437	106.607
	L	-81.9095	20.5336	-150.437	-13.382
	P	78.3253	8.6453	49.577	107.073
	S	-53.8238	14.6587	-102.499	-5.148
	Z	7.5783	9.2020	-22.995	38.151
	Minor	42.6971	11.9281	2.803	82.591
C	L	-144.2367	21.2427	-215.443	-73.031
	Q	-41.1667	10.1217	-75.505	-6.829
	R	-35.5892	10.2051	-70.197	-.981
	S	-116.1509	15.6365	-168.529	-63.773
	T	-54.7489	10.6911	-90.931	-18.566
G	K	34.4587	8.1011	7.466	61.425
	L	-123.9730	20.1436	-191.200	-56.746
	P	36.2618	7.6431	10.768	61.756
	S	-95.8872	14.1072	-142.714	-49.060
	T	-34.4852	8.2953	-62.021	-6.950
H	K	37.9226	4.4754	22.910	52.935
	L	-120.5090	18.9780	-183.883	-57.135
	P	39.7257	3.6437	27.615	51.837
	Q	-17.4391	3.3709	-28.611	-6.267
	S	-92.4233	12.3860	-133.527	-51.319
	T	-31.0212	4.8179	-46.988	-15.054
	Minor	22.4262	6.2593	1.397	43.455
K	L	-158.4317	19.0804	-222.184	-94.680
	Q	-55.3617	3.9060	-68.512	-42.212
	R	-49.7842	4.1172	-63.623	-35.945
	S	-130.3459	12.5422	-172.030	-88.661
	T	-68.9439	5.2064	-86.352	-51.536
	Z	-33.8250	9.2037	-64.808	-2.842

. . . table continued on the following page . . .

± Note: Sig. = .000 indicates $p < .0005$

¥ To save space, only significant differences are reported

* Significant at $p \le .05$; ONLY significant differences are listed and are listed once; nonsignificant differences not listed; positive Mean Diff. indicates the LC class on the left had superior performance, and negative Mean Diff. indicates the LC class on the right had superior performance.

Table 2f (continued): UNCL's Elsevier Freedom Collection Activity (2017-2021) by LC Class: Analysis of Variance (Individual Responses)

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.±
Between LC	16441005.52	12	1370083.794	19.533	.000
Within LC	3145594464	44847	70140.577		
Total	3162035469	44859			

Post Hoc Tests (Dunnett's C)¥

LC Class (I)	LC Class (J)	Mean Diff. (I-J)*	S.E.	95% C.I. Lower	95% C.I. Upper
L	P	160.2348	18.9026	97.103	223.367
	Q	103.0699	18.8519	40.112	166.028
	R	108.6475	18.8968	45.542	171.753
	T	89.4878	19.1636	25.504	153.417
	Z	124.6067	20.6118	55.681	193.532
	Minor	142.9352	19.5758	77.505	208.366
P	Q	-57.1649	2.9163	-66.872	-47.458
	R	-51.5873	3.1936	-62.209	-40.965
	S	-132.1490	12.2701	-172.879	-91.419
	T	-70.7470	4.5115	-85.726	-55.768
	Z	-35.6281	8.8293	-65.314	-5.943
Q	S	-74.9842	12.1918	-115.445	-34.524
	Minor	39.8653	5.8658	20.123	59.608
R	S	-80.5617	12.2611	-121.252	-39.872
	T	-19.1597	4.4872	-34.029	-4.291
	Minor	34.2878	6.0085	14.080	54.496
S	T	61.4020	12.6685	19.364	103.440
	Z	96.5209	14.7680	47.287	145.755
	Minor	114.8495	13.2837	70.641	159.058
T	Z	35.1189	9.3751	3.663	66.575
	Minor	53.4474	6.8012	30.646	76.248

± Note: Sig. = .000 indicates $p < .0005$

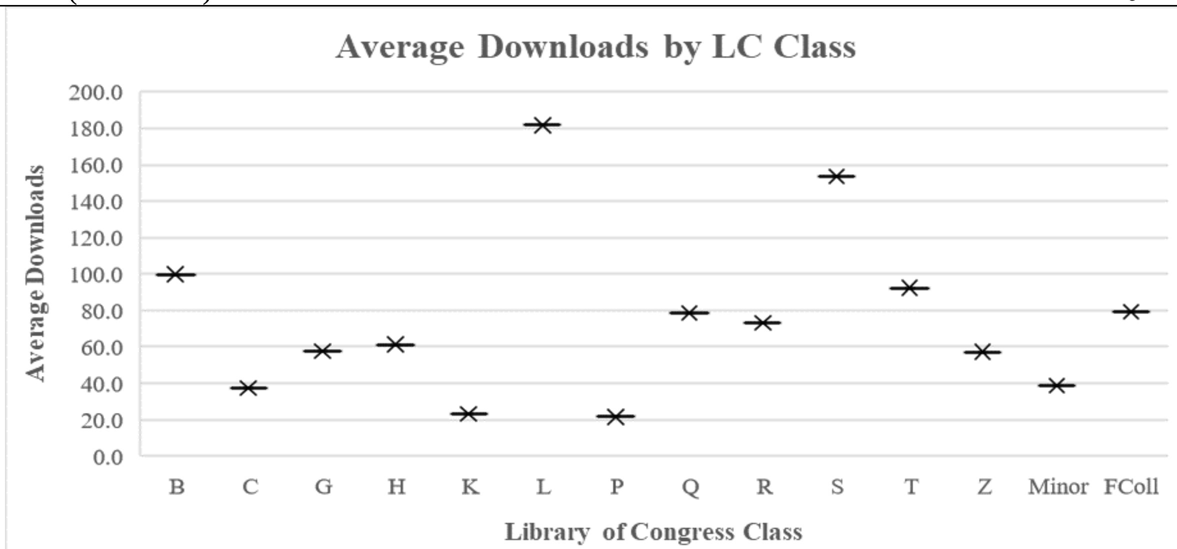
¥ To save space, only significant differences are reported

* Significant at $p \leq .05$; ONLY significant differences are listed and are listed once; nonsignificant differences not listed; positive Mean Diff. indicates the LC class on the left had superior performance, and negative Mean Diff. indicates the LC class on the right had superior performance.

As is unfortunately so often the case, the author's analyses have produced an avalanche of numbers and traced a bewildering array of relationships among the LC Classes. To provide a general sense of what the tests in the tables above indicate, the author has provided a Table 2f summary on the following page. Essentially, for UNCL, where average download performance is concerned, the LC Classes *B – Philosophy. Psychology. Religion*, *Q – Science*, *R – Medicine*, and *T – Technology* should be considered the workhorses of the Freedom Collection. The lesser lights among the LC Classes, but still productive performers, were *G – Geography. Anthro. Recreation*, *H – Social Sciences*, and *Z – Bibliography. Library Science.* The poor performers were *C – Auxiliary Sciences of History*, *K – Law*, *P – Language and Literature*, and the LC classes in the Minor LC category. The standout star performers among the smaller LC classes were *L – Education* and *S – Agriculture*.

Table 2f-Summary: UNCL's Elsevier Freedom Collection Activity (2017-2021) by LC Class (Individual Responses): Average Performance Comparisons

		Performance is . . .	
LC Class	Superior to	Equivalent to	Inferior to
B	C G H K P Z Minor	Q R T	L S
C		G H K P Z Minor	B L Q R S T
G	K P	C H Q R Z Minor	B L S T
H	K P Minor	C G R Z	B L Q S T
K		C P Minor	B G H L Q R S T Z
L	B C G H K P Q R T Z Minor	S	
P		C K Minor	B G H L Q R S T Z
Q	C H K P Minor	B G R T Z	L S
R	C K P Minor	B G H Q Z	L S T
S	B C G H K P Q R T Z Minor	L	
T	C G H K P R Z Minor	B Q	L S
Z	K P	C G H Q R Minor	B L S T
Minor (D J N U V)		C G K P Z	B H L Q R S T



Negative Binomial Regression (NB Regression)

Regression analysis is "a set of statistical procedures designed to examine relationships between one or more independent variables (IV) and one dependent (i.e., outcome) variable (DV)" (Randolph & Myers 2013, 109). One way to think about regression is to consider the mean, or average, of the outcome data and to think that regression calculates a sort of running mean through the data, with the model's estimation changing (i.e., going up or down) depending upon the conditions presented (e.g., the effects of the independent variable of interest) and the conditions controlled for (e.g., the effects of the other independent variables). So, in this case, if the ANOVA above correctly modelled UNCL's download data, one would assume some LC classes were average-ish and did not have large effects on downloading (e.g., B, Q, R, and T), some LC classes drove downloading counts substantially lower (e.g., C, K, P, and Minor LC), and some drove downloading counts substantially higher (e.g., L and S).

As was noted above, the author will be employing negative binomial (NB) regression, as opposed to more familiar linear regression approaches such as ordinary least squares (OLS), because it has less restrictive assumptions and is a standard method used to model overdispersed count data (Hilbe, 2007; Hoffmann, 2004; Orme & Combs-Orme, 2009). With NB regression, some of the reported parameters and the resulting regression equations may look a bit different from OLS and its more straightforward additive equations (e.g., $y = a + bx + \varepsilon$, where y = DV, a = y intercept, b = regression coefficient, x = IV, and ε = model error). The NB equation will generally take the form of $\lambda = e^{\eta}$ where the calculated rate/slope equals the exponentiated value of the linear predictor. Thus, the equation would be: $\lambda = e^{\eta}$ where e equals Euler's number and $\eta = \alpha + bx + \varepsilon$ (λ = DV or y, α = y intercept, b = regression coefficient, x = IV, and ε = model error).

As was the case with the ANOVA above, for the NB regression the first step in analyzing the possible effects of the several LC classes on the dependent variable Downloads would be to test whether or not there are any statistically significant effects present in the dataset. As Table 2g indicates, there was at least one statistically significant effect present (Note: Sig. = .000 denotes a p value < .0005):

Table 2g: Omnibus Test[a]

Likelihood Ratio Chi-Square	DF	Sig.[*]
421.428	12	.000

a: Compares the fitted model against the intercept-only model
*: Sig. of .000 indicates that p < .0005

The Omnibus Test having established that there was at least one statistically significant effect in the dataset, the next step would be to determine which of the independent variables (i.e., the LC classes) may have produced an effect. Normally, at this point, the author would create another table that presented the results of a test of model effects (Type III Test of Fixed Effects) to follow the Omnibus Test and show which subjects' performances were significantly different than the others' performances, but SPSS was unable to compute the necessary values.[*]

Instead, the author will move on to presenting the regression coefficients (β) for the several LC classes in the model. Often with regression analysis, an analyst will generate regression coefficients using the worst-performing or best-performing category as the comparison or baseline category, depending upon what he or she wishes to illustrate. In this instance, the

---

* SPSS was unable to compute the necessary Likelihood Ratio Chi-Square values because the estimable function had zero degrees of freedom.

author selected *Q – Science* as the baseline category as it was one of the larger categories, it had an average download performance and standard deviation that were close to those of the Freedom Collection as a whole, and its Percentage Expected Use value for downloads and journals was roughly proportional (see Tables 2e-Alternate and 2b above, respectively). This is all to say that the performance of *Q – Science* seemed to be pretty typical of the Freedom Collection as a whole.

Table 2h: Freedom Collection LC Class Parameter Estimates

			95% Wald C.I.		Hypothesis Test		
Parameter	β	S.E.	Lower	Upper	Wald χ^2	df	Sig.*
(Intercept α)	78.687	1.5539	75.641	81.732	2564.322	1	.000
B	21.160	8.9143	3.689	38.632	5.635	1	.018
C	-41.167	8.8774	-58.566	-23.767	21.504	1	.000
G	-20.903	4.5686	-29.857	-11.949	20.934	1	.000
H	-17.439	2.6930	-22.717	-12.161	41.935	1	.000
K	-55.362	5.2050	-65.563	-45.160	113.132	1	.000
L	103.070	24.4471	55.154	150.985	17.775	1	.000
P	-57.165	3.2091	-63.455	-50.875	317.321	1	.000
R	-5.578	2.0780	-9.650	-1.505	7.204	1	.007
S	74.984	9.8840	55.612	94.356	57.554	1	.000
T	13.582	2.9365	7.827	19.337	21.394	1	.000
Z	-21.537	10.6303	-42.372	-.702	4.105	1	.043
Minor	-39.865	7.7989	-55.151	-24.580	26.129	1	.000
Q	0[a]						
(Scale)	1[b]						
(N.B.)	5.400	.0349	5.332	5.469			

Dependent Variable: Downloads
Model: (Intercept) B C G H K L P R S T Z Minor Q
a. Set to zero because the parameter is redundant
b. Fixed at the displayed value
*: Sig. of .000 indicates that p < .0005

Despite the avalanche of numbers, interpreting the results in Table 2h is actually fairly straightforward. In a simple straight line equation, such as would have been produced via a linear regression method like Ordinary Least Squares, the intercept would be the value of Y when X equals zero (Keith, 2006), and the regression coefficients for each variable would indicate the amount that each Independent Variable ought to be multiplied by to produce a one unit change in Y as one moves through the data (Randolph & Myers, 2013). For a multiple regression analysis, the intercept would be the mean value of Y when all Independent Variables equal zero (Keith, 2006).

In this case, because all of the variables in the multivariate equation were categorical variables, they were coded as being equal to 1 or 0, to indicate presence or absence. Thus, to produce an estimate in the same manner as with linear regression, one would need only take the per data cell value of *Q – Science*, which is the Intercept because Q was set to be the reference case, and add to it the β value of the LC Class in question multiplied by the number of data cells in question. For example, the value of a single *S – Agriculture* data cell in the dataset is $Y = \alpha + \beta_S S$ or Y = 78.687 + 74.984(1). So, where *Q – Science* produces approximately 78.687 downloads per data cell, *S – Agriculture* produces approximately 153.671. The Confidence Intervals (C.I.) give the plus-or-minus bounds of the β estimates, and the Wald χ^2 and Significance values indicate whether or not the effect of each LC class was statistically significant (i.e., not explainable by chance). For this table, the author employed N.B. regression rather than OLS, and N.B. is usually a little more complicated to interpret, but in this instance the linear approach outlined above should be accurate enough.*

The takeaway from all of this number crunching, in a nutshell, would be that any LC classes with positive regression coefficients (β) in Table 2h should be understood to be somewhat above-average in their productivity, and those with negative coefficients should be seen as being below-average. With *Q - Science* serving as the average, big coefficients (β) indicate pretty big deviations from average productivity, and small coefficients indicate roughly average productivity. So, for example, *S – Agriculture* can be taken to be a big producer, K – Law (β_K = -55.362) can be taken to be a poor producer, and *R – Medicine*, even though its coefficient is negative, can be taken to be on a par with *Q – Science* because its coefficient (β_R = -5.578) is small.

The overall takeaway, of course, is that the results of Tables 2g and 2h largely replicate those of the ANOVA tables in the subsection above, so the sense of the relative productivity/importance of the several LC Classes provided by the ANOVA tables is likely correct.

---

* Because the equation for N.B. regression is $\lambda = e^{\eta}$, where the calculated rate/slope equals the exponentiated value of the linear predictor, the linear combination of predictors (η) have an additive effect at the log(y) scale (Orme & Combs-Orme, 2009). So, normally with the *S - Agriculture* example above, $\log(y) = \alpha + \beta_S S$, with $y = e^{(\alpha + \beta_S S)}$. However, since all of the independent variables in this model's equation were dummy variables (i.e., 1 or 0), this mathematical rigamorale should not be necessary. With only categorical independent variables in the model, the N.B. regression approach produced the LC Classes' relative per data cell values in roughly the same manner that linear regression would have. The only difference would be that the linear regression equation's intercept would have been α = 73.109 and *Q -Science*'s β value would have been 5.578, which would result in Q's data cell values being $y = \alpha + \beta_Q Q = 73.109 + 5.578(1) = 78.687$. So, effectively the only difference here between the OLS approach and the author's N.B. approach would be that OLS's intercept value would have been the average value of y when all LC classes' values were set to 0, and the N.B. intercept value equals that of *Q – Science*, the selected comparison category. Again, the author would like to emphasize that he is not a statistician, so anyone more mathematically proficient should feel free to correct any of the above should it prove to be incorrect. Honestly, I am teaching myself this stuff as I go.

References

Baker, S. L., & Wallace, K. L. (2002). *The Responsive Public Library: How to Develop and Market a Winning Collection*. Libraries Unlimited.

Bonn, G. S. (1974). Evaluation of the collection. *Library Trends, 22*(3): 265-304.

Brown, J. M., & Stowers, E. D. (2013). Use of data in collections work: An exploratory survey. *Collection Management*, *38*(2), 143-162.

Day, R. W., & Quinn, G. P. (1989). Comparisons of treatments after an analysis of variance in ecology. *Ecological Monographs*, *59*(4): 433-463.

Dunnett, C. W. (1980). Pair wise multiple comparisons in the unequal variance case. *Journal of the American Statistical Association*, *75*: 796-800.

Freud, S. (2002). *Civilization and Its Discontents*. Penguin.

Gigerenzer, G. (2002). *Calculated Risks: How to Know When Numbers Deceive You*. Simon & Schuster.

Gilchrist, A. (1971). Cost-effectiveness. *Aslib Proceedings*, *23*(9): 455-464.

Harker, K. R. (2022). The depths of cost-per-use: Historical context and applications. *Library Trends*, *70*(3), 355-386.

Hilbe, H.M. (2007). *Negative Binomial Regression*. Cambridge: Cambridge UP.

Hoffman, J.P. (2004). *Generalized Linear Models: An Applied Approach*. Boston: Pearson.

Keith, T. Z. (2006). *Multiple Regression and Beyond*. Boston: Allyn & Bacon.

Library of Congress. (n.d.). *Library of Congress Classification Outline*. Available at: https://www.loc.gov/catdir/cpso/lcco/

Mills, T. (1982). *The University of Illinois Film Center Collection Use Study*. Urbana, IL: University of Illinois, CAS Paper. (ERIC Document: ED227821).

Orme, J.G. & T. Combs-Orme. (2009). *Multiple Regression with Discrete Dependent Variables*. Oxford: Oxford UP.

Randolph, K. A., & Myers, L. L. (2013). *Basic statistics in multivariate analysis*. Oxford UP.

Shingala, M. C., & Rajyaguru, A. (2015). Comparison of post hoc tests for unequal variance. *International Journal of New Technologies in Science and Engineering 2*(5): 22-33.

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*, *120*, 103445

Tyler, D. C., Macaulay, D., and McClanahan, R. (2021) *UnSub: Sundry Thoughts, Employing Data from the Wiley e-Journal Package* (unpublished report). University Libraries, University of Nebraska-Lincoln.
Available at: https://digitalcommons.unl.edu/librarywhitepapers/17/

Walters, W. H. (2016). Evaluating online resources for college and university libraries: Assessing value and cost based on academic needs. *Serials Review, 42*(1): 10-17.