Fall 2022

The Freedom Collection 2017–2021: Part 2, UNCL Members' Subject Usage Preferences and Profiles

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

The Freedom Collection 2017–2021: Part 2,
UNCL Members' Subject Usage Preferences and Profiles

Author:

David C. Tyler

Data:

David Macaulay

David C. Tyler

Submitted:

Fall 2022

Table of Contents

INTRODUCTION

The Freedom Collection (Elsevier) and UNCL

This report is the second in a series of reports on Elsevier's Freedom Collection (FC) and the University of Nebraska Consortium of Libraries' (UNCL) usage of same. The first report concentrated on describing the FC, describing UNCL's and its members' usage of the FC, and exploring whether UNCL had any preferences among the subjects (i.e., Library of Congress classifications) that comprise the FC.

This follow-up report will look into the UNCL members' usage in more depth, with the implicit question of whether, were Elsevier to break up the FC into smaller subject-based collections, UNCL's members could easily agree on which FC subject collections to subscribe to. Thus, this report will dive into whether there has been a by-subject relationship between the FC's composition and UNCL's downloading, whether UNCL's by-subject downloading from the collection has been consistent, whether the UNCL members' have by-subject relative use preference, and whether UNCL members' by-subject usage profiles differ significantly.

The Data:

As was noted in the first part of this report, for a prior 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 – Lincoln (UNL), the University of Nebraska Medical Center (UNMC), and the University of Nebraska at Omaha (UNO).*

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. 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 closer looks 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 FC 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).

As was the case with previous reports, the reader should keep in mind that the author is not particularly knowledgeable where Elsevier journals are concerned and has not been involved 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

---

* 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.

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 author. If the reader has additional questions, the 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 author is ignorant 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 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 4 will be Table 4 plus a sequential letter or a content note to differentiate them). Tables and figures with the same number-letter combinations should reference the same data in whole or in part. Hopefully, this approach will keep things fairly straightforward. Section numbering will proceed from the prior portion of the FC report, so the reader may find Sections 1 and 2 in Part 1 of this report, which was distributed earlier, and begin reading here with Section 3.

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

Finally, it should be understood that the author's summaries and analyses have been derived solely from the data provided and, if employed in deliberations, 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 the analysis was performed by the listed author, David C. Tyler, so it should also be understood that any errors present likely belong to me alone. Any errors herein attributable to David Macaulay 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. Just prior to 2022, this collection was comprised of 2,243 journals, 60% of the full title list compiled by David Macaulay.* In addition to the information provided by David Macaulay, the 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

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

remaining 15 journals, the author manually assigned LC classifications/subclassifications based upon those 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 was treated as though it was 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).

The questions posed in this second part of the Freedom Collection report will be as follows:

5. Were the by-subject composition of the Freedom Collection journals and UNCL's downloading by subject similar?
6. Were UNCL's and the UNCL members' by-subject download counts consistent from year to year?
7. Did UNCL members have by-subject relative use preferences?
8. Were UNCL Members' by-subject download profiles similar?

This FC sub-report's findings are summarized here:

5. The subject profiles of the Freedom Collection by journal and of UNCL/UNCL members by download were extremely similar, which suggests that the composition of the Freedom Collection may have had a sizeable influence on its by-subject utilization by UNCL and by its members (Section 3);
6. By-subject download counts for UNCL and for its members were extremely consistent from year to year, which suggests that each institution had a predictable character across the interval (Section 4);
7. Calculation of relative use factors for the FC's subjects indicated that the UNCL members likely have local subject preferences (Section 5);
8. Further, more formal, analysis suggested that the UNCL members do indeed have differing by-subject usage profiles, and outside of *Q – Science,* where there was consortium-wide agreement, the members' subject preferences may differ substantially, (Section 6).

Freedom Collection Subjects and UNCL Downloads

SECTION 3: Journal and Download Count by-Subject Correlations

The author thought it would be informative first to look into possible relationships between the Freedom Collection's (FC) subject profile and the UNCL and UNCL member by-subject download profiles to see whether UNCL's activity may have been at least partially determined by the subject composition of the FC. To that end, for Table 3a the author calculated whether the numbers of journals in each FC Library of Congress (LC) classification were correlated with the numbers of downloads produced by UNCL and by its members in each LC classification in each year and in total.

Table 3a: Freedom Collection Journal Counts and UNCL Downloads by Library of Congress Classification: Pearson Product-Moment Correlations

	Year					
	2017	2018	2019	2020	2021	**2017-2021**
UNK	0.955	0.968	0.971	0.976	0.965	**0.972**
UNL	0.843	0.822	0.843	0.853	0.851	**0.842**
UNMC	0.891	0.883	0.886	0.876	0.895	**0.886**
UNO	0.632	0.931	0.956	0.931	0.946	**0.938**
UNCL	0.992	0.990	0.994	0.995	0.981	**0.993**

As one can see from the table, the number of journals in each subject would appear to be very strongly correlated with the numbers of downloads generated by each UNCL member in each year and in total and would appear to be correlated almost perfectly with download counts generated by UNCL as a whole. One could therefore conclude that, to a sizable degree, the by-subject distribution of UNCL's and its members' downloads within the Freedom Collection universe may have largely been determined by Elsevier's composition of the collection.*

*For readers less familiar with Pearson's product moment correlation coefficients, Cohen (1988, 1992), has offered some general rules of thumb for interpreting the correlation coefficients and their effect sizes: r = .10 (small effect); r = .30 (medium effect); and r = .50 (large effect). As Field (2013) and Ellis (2010) point out, the square of r functions as a proportion of variance explained index. Thus r = .10 results in r^2 = .01, which suggests that the effect explained 1% of the variance, and r = .90 results in r^2 = .81, which would mean that the effect explained 81% of the variance. So, those would be quite a small and quite large effects, respectively.

SECTION 4: UNCL By-Subject Download Consistency

The author thought it would also be informative to look into whether or not UNCL and its members behaved consistently within the interval under examination. If members behaved unpredictably from one year to the next, then it would be meaningless to proceed on to looking for UNCL subject-usage profiles. Their downloading by subject in one year would not be informative (i.e., have predictive value) for other years, so the usage profile for the entire interval would essentially be the product of a seemingly random process, which would result in the profile having deceptive descriptive value and spurious predictive value. Given the results in the Section 3, randomness would seem unlikely, but it is worth exploring.

Table 4a: UNCL Download Counts by Library of Congress Classifications: Pearson Product-Moment Correlations (r)

		Year				
	Year	2018	2019	2020	2021	**Avg r**
UNK	2017	0.989	0.995	0.992	0.971	**0.988**
	2018		0.997	0.996	0.979	
	2019			0.997	0.978	
	2020				0.988	
UNL	2017	0.997	0.998	0.997	0.994	**0.998**
	2018		0.999	0.998	0.997	
	2019			0.999	0.998	
	2020				0.999	
UNMC	2017	0.999	1.000	0.998	1.000	**0.999**
	2018		1.000	1.000	1.000	
	2019			0.999	1.000	
	2020				0.999	
UNO	2017	0.859	0.789	0.767	0.724	**0.903**
	2018		0.991	0.971	0.964	
	2019			0.986	0.988	
	2020				0.995	
UNCL	2017	0.998	0.998	0.994	0.990	**0.995**
	2018		0.999	0.995	0.996	
	2019			0.998	0.992	
	2020				0.986	

Note: There were journals in seventeen Library of Congress Classifications in the Freedom Collection. Readers wishing more information are invited to peruse Part 1 of this report.

To determine whether or not UNCL's members had consistent and predictable by-subject download profiles from year to year, the author calculated whether UNCL'S and each UNCL member's download counts by Library of Congress (LC) classification were correlated every year in the dataset. As can be seen in Table 4a, UNCL's and its members' by-subject download activity would seem to be preternaturally consistent from one year to the next. One could certainly conclude that each UNCL member and UNCL itself had a fairly fixed character throughout the interval.

The only member-year whose correlations were below r = .900 was UNO in 2017. UNO 2017's relationships with UNO's other years were still quite strong, but the discrepancy between UNO 2017's row in the table and the entirety of the rest of the table did catch the author's eye. A quick look at the data showed that UNO had 1,598 zero-use journals in 2017, while zero-use journals for UNO ranged from just 576 to 668 in 2018-2021, so it is possible that UNO may have experienced an expansion of access in 2018 rather than a small change in character. The reader interested in solving this minor mystery is invited to consult the UNCL Acquisitions Group's minutes or the collections staff at UNO.

SECTION 5: By-Subject Relative Use Preferences

Before moving to the final section and the analysis of the UNCL members' subject-usage profiles, the author thought that he would calculate and plot relative use factors for the member institutions and the FC subjects (Note: UNCL's relative use factors were calculated and presented in the previous section of this report). As the author noted in Part 1, relative use factors offer a quick and easily interpretable assessment metric that, as Baker and Wallace (2002) have said, allows one to "indicate a deviation from expected behavior" (208). As was the case in Part 1 of this report, the author will be using Mills' Percentage Expected Use, the "ratio of the percentage of use of a subject to its percentage holdings" (5), to place the LC classes into one of four quadrants in the scatterplots below.

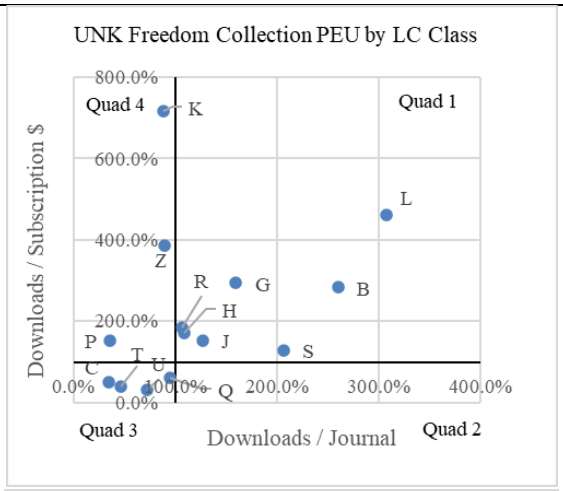
In the scatterplots, the X axis measures proportional downloads per journal and the Y axis measures proportional downloads per invoiced subscription dollar, with the X-Y axis indicating the point where both were perfectly proportional.* One should read each graph by starting in the upper right-hand corner of the plot and proceeding clockwise around the X-Y axis: Quad 1 contains LC classes whose per-journal and per-dollar activity was higher than expected; Quad 2 contains LC classes whose per-journal activity was higher than expected but whose per-dollar activity was lower than expected; Quad 3 contains LC classes whose activity was lower than expected on both points, i.e., the converse of Quad 1; and Quad 4 contains LC classes whose per-journal activity was lower than expected but whose per-dollar activity was higher. The farther any one datapoint is from an axis, the greater the disproportionality of its activity along that axis.

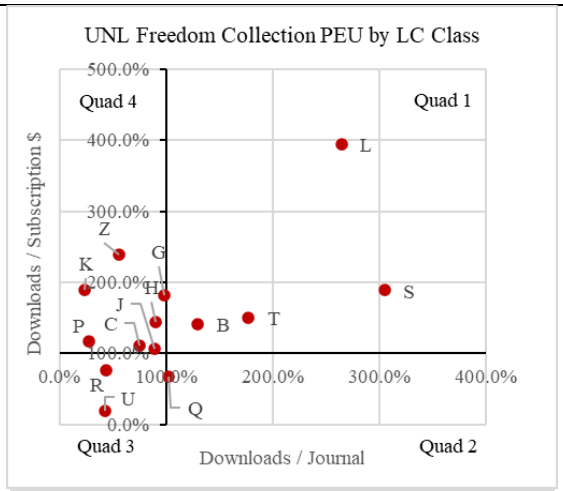
To approach the question of the LC classes' relative usage for the four UNCL members, the author has replicated the UNCL scatterplot from Part 1 in Table 5-a on the next page for each UNCL member. A question concerning library consortia that has for some time been of great interest to the author is whether or not the usage patterns visible at the level of a consortium are replicated at the level of the consortium's members. The author's inclination has been to assume that if consortium members' usage patterns are similar, then members should have an easier time coming to agreement where the allocation of resources is concerned, but where usage patterns differ greatly, one should expect greater conflict. As the table shows, the UNCL members' relative use factors would seem, within the bounds of Elsevier's Freedom Collection, to reveal somewhat different interest profiles.

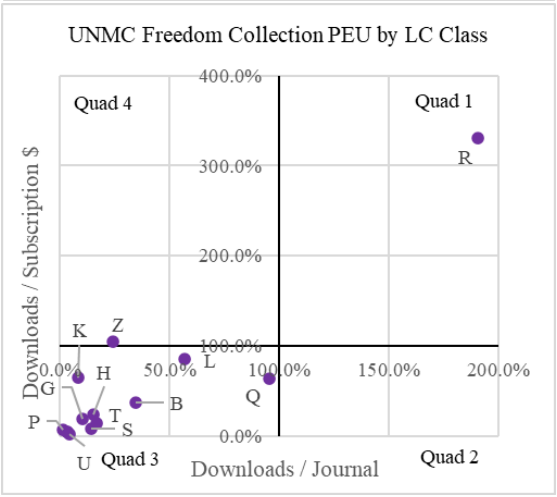
---

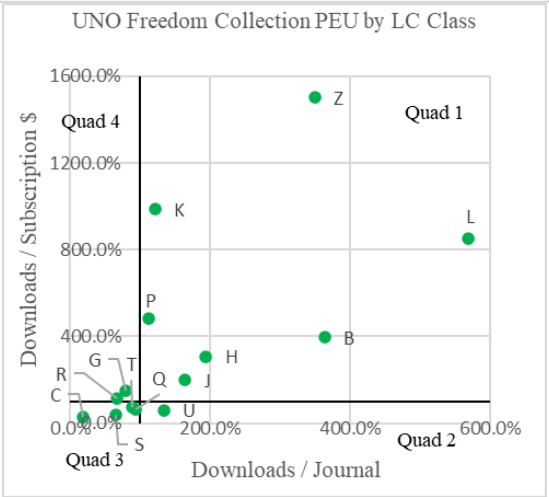
* When reviewing the PEU scatterplots, the reader should be cautious and keep in mind that David Macaulay was able to produce UNCL subscription prices for just 27.3% of Freedom Collection journals, so the positions shown along the Y axis are tentative and contingent. Should more UNCL pricing data become available or should UNCL switch to individual subscriptions, the plotted datapoints could change their Y positions substantially.

Table 5a-1: UNCL Members' Freedom Collection Download Activity (2017-2021): LC Class Percentage Expected Use (PEU) Quadrants









Note: The subscription dollar amounts employed were for UNCL as a whole, not for the individual members; LC classes D, N, and V had no journals with subscription cost data, so they had no PEU values to report on the Y axis and were not included in the scatterplots.

From the scatterplots, it would seem that all four members have a solid, proportional interest in *Q - Science* (proportional to the number of journals but not quite to the subscription dollars spent). Beyond this shared consistency, it would seem that UNK has a solid interest in LC classes B, G, H, J, L, R, and S, and has less interest in LC classes C, T, and U. UNL's interests seem to be most strongly focused on LC classes B, L, S, and T, and UNL seems to show more relative disregard for LC classes R and U. UNMC had an outsized and unsurprising interest in *R – Medicine*, which the author had expected to be the bulk of what the Freedom Collection has to offer UNMC. Finally, UNO seemed to be interested in LC classes B, H, J, K, L, P, and Z, and to slightly disregard classes C, G, R, S, and perhaps T.

Thus, if Elsevier were to break up the Freedom Collection in favor of subject subcollections, one would expect there to be some conflicts of interest between the UNCL members.

While some of the more obvious differences between the UNCL members' relative use of the LC classes may be easily spotted in Table 5a-1 above, the accompanying quadrant table, which is visually less busy, should probably also be reviewed to spot all potential commonalities and differences.

Table 5a-2: UNCL Members' Freedom Collection Download Activity (2017-2021): LC Class Percentage Expected Use (PEU) Quadrants

LC Class	UNK	UNL	UNMC	UNO
B	Q1	Q1	Q3	Q1
C	Q3	Q4	Q3	Q3
D	n/a	n/a	n/a	n/a
G	Q1	Q4	Q3	Q4
H	Q1	Q4	Q3	Q1
J	Q1	Q4	Q3	Q1
K	Q4	Q4	Q3	Q1
L	Q1	Q1	Q3	Q1
N	n/a	n/a	n/a	n/a
P	Q4	Q4	Q3	Q1
Q	Q3	Q2	Q3	Q3
R	Q1	Q3	Q1	Q4
S	Q1	Q1	Q3	Q3
T	Q3	Q1	Q3	Q3
U	Q3	Q3	Q3	Q2
V	n/a	n/a	n/a	n/a
Z	Q4	Q4	Q4	Q1

Key:
Q1 = % downloads > % journals and % subscription $
Q2 = % downloads > % journals but < % subscription $
Q3 = % downloads < % journals and % subscription $
Q4 = % downloads < % journals but > % subscription $

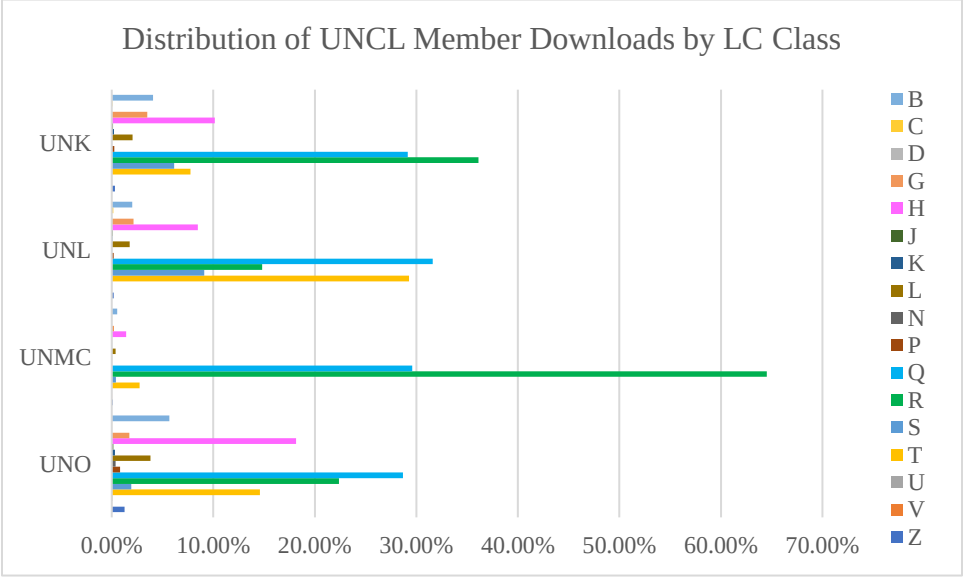
SECTION 6: UNCL Member Subject Usage Profiles

Having established that the distribution of downloads by subject may have largely been determined by Elsevier's composition of the Freedom Collection, having established that UNCL and its members have been very consistent in their by-subject downloading over the interval, and having explored whether UNCL's members seem to favor some subjects over others, we may now turn to the author's main question for Part 2 of the report: Are the UNCL members' by-subject download profiles substantially different?

As was noted in the previous section, the author's assumption where library consortia are concerned has been that, where members' usage patterns are similar, consortia members should have an easier time coming to agreement where the allocation of resources is concerned. Tables 5a-1 and -2 above suggested that UNCL's members may well have different interests. Thus, it would be worthwhile to explore this issue a bit further to determine whether or not these apparent differences in relative use preferences are real and substantial.

Perhaps the easiest and most intuitive approach to this question would be to calculate the LC Classes' percentage download performances for each UNCL member and present them as profiles for informal review. To that end, feel free to peruse the percentage download profiles in Table 6a (Bar Graph).

Table 6a (Bar Graph): UNCL Members' Freedom Collection Activity (2017-2021): LC Class Percentage Download Profiles

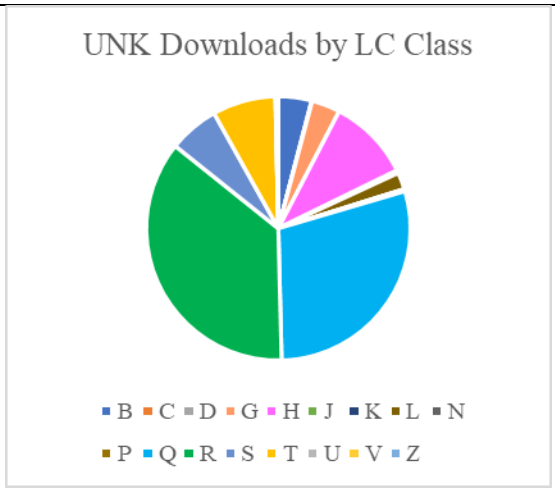


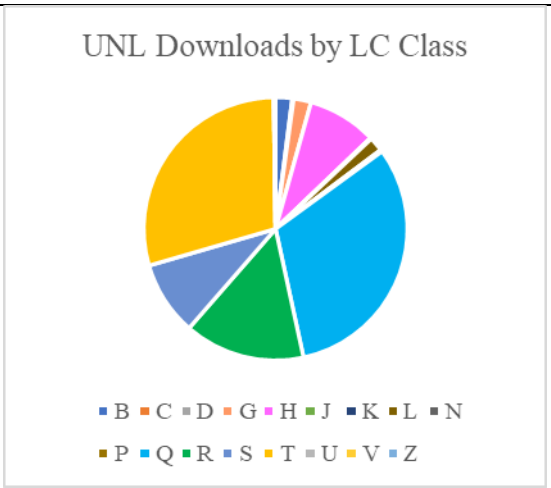
As was the case with the PEU tables in the previous section of the report, the most pronounced download footprint is UNMC's: about 65% of usage went to *R – Medicine* and nearly 30% went

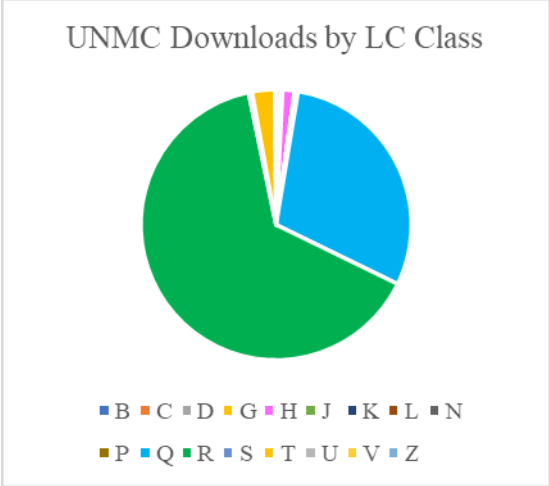
to *Q – Science*. The bulk of the scanty remainder went to *T – Engineering* and *H – Social Sciences*.

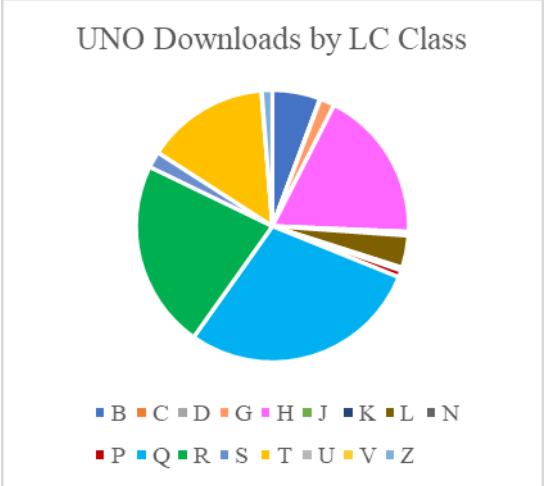
If the tiny bars in the charts above are too difficult to parse, the data above could also be perused via pie charts, such as in this alternate table, Table 6a (Pie Charts):

Table 6a (Pie Charts): UNCL Members' Freedom Collection Activity (2017-2021): LC Class Percentage Download Profiles









Interestingly, as was seemingly evident from the PEU graphs in Table 5a-1 in the prior section, all UNCL members' percentage downloads do appear to have favored *Q – Science* roughly equally. Thus, a science journal subject collection would seem to be equally attractive to all four campuses.

Looking beyond LC class Q, UNK and UNL seemed to have some similarities, although UNK employed *R – Medicine* and UNL employed *T – Engineering* quite a bit more, respectively, than did the other. Both, as was noted above, employed *S – Agriculture* more than did UNO. UNO and UNK appear to have employed the psychology journals in LC class B a little bit more avidly

than did UNL, and UNO appears to have employed the largely business-focused social sciences journals of LC class H more than did UNK or UNL. Thus, the bar charts/pie charts of Tables 6a would seem to support many of the intuitive inferences one could draw from the Percentage Expected Use tables of the previous section.

To determine whether or not these readily apparent differences between the UNCL members are likely to be real and to point to substantially distinct institutional characters, rather than being misleading invitations to indulge in what Freud (2002) called the narcissism of small differences, would require a more formal exploration and testing. As was the case with Section 2 of the previous issue of the FC report, this would be the point where the report becomes a less pleasant read, and those with an aversion to statistics and formal analysis are invited to skip the rest of this section of the report.

The author's first thought with regard to the UNCL members' subject profiles was that utilizing Analysis of Variance (ANOVA), as in Section 2 of the earlier report, would be an approach that might allow for a comparison of the Freedom Collection's proportional performance at all four campuses, but ANOVA is intended for interval/ratio data and for the comparison of the means of multiple samples or a single sample under multiple conditions (Field, 2013), not for the comparison of proportions or percentages (McDonald, 2014). Percentages violate ANOVA's assumption that values are free to vary around the mean without imposed limits (i.e., percentages, of course, can never be less than 0% or more than 100%, except under rare and unusual circumstances [Stuart, 1971]). Additionally, percentage/proportion measurements of a whole are not entirely independent of one another: One response category's percentage/proportion values constrain those of all other dependent variables, to the point where, in this case, knowing the values of sixteen of the seventeen LC classes would determine the seventeenth value.

Another approach that seemed interesting to the author was Latent Profile Analysis (LPA), which would allow one to look at the "choices" each member made in terms of downloads and identify which members were most similar in their interests. Unfortunately, the author has never run an LPA and is not familiar enough with the technique to give it a try. Also, the literature suggests that a sample size of at least 500 respondents would be preferable (Spurk et al., 2020), and there are only four UNCL members. The author considered that it might be possible to reverse the question and to analyze and group the journals on the basis of their download recipients, but, again, the author would need to familiarize himself with the technique or contact a statistician or social sciences researcher for assistance in order to attempt an LPA.

A technique recommended online for percentage/proportion data was beta regression with the R package (Mangiafico, 2016). Unfortunately, the author is unfamiliar with R and is unfamiliar with how to employ this technique in SPSS using the GUI, and the suggested SPSS syntax was well beyond the author's rather limited abilities.

Another recommendation was that the UNCL and LC proportion data be treated as polychotomous nominal variables and that the data be analyzed via multinomial logistic regression (McDonald, 2014; Orme & Combs-Orme, 2009). Not being a statistician, the author cannot say whether any of these techniques would be preferable to the others, but this last is one that resides within the author's rather limited bag of tricks.

Before proceeding, the author thought it might be worthwhile to express Table 6a with the UNCL members' download counts and percentages by LC class, just in case the reader wanted to have the actual counts and percentages, rather than just the visualizations above, to refer to.

Table 6a (Counts & Percentages): UNCL Members' Freedom Collection Activity (2017-2021): LC Class Percentage Download Profiles

	Download Counts				Download Percentages			
LC Class	UNK	UNL	UNMC	UNO	UNK	UNL	UNMC	UNO
B	5,774	41,542	5,778	16,799	4.05%	2.02%	0.54%	5.68%
C	109	3,436	83	124	0.08%	0.17%	0.01%	0.04%
D	0	254	2	116	0.00%	0.01%	0.00%	0.04%
G	4,965	44,098	2,405	5,160	3.49%	2.15%	0.22%	1.74%
H	14,441	173,937	15,133	53,729	10.14%	8.47%	1.42%	18.16%
J	161	1,628	33	434	0.11%	0.08%	0.00%	0.15%
K	335	1,275	230	959	0.24%	0.06%	0.02%	0.32%
L	2,934	36,277	4,059	11,257	2.06%	1.77%	0.38%	3.80%
N	43	859	44	1,146	0.03%	0.04%	0.00%	0.39%
P	364	4,028	117	2,378	0.26%	0.20%	0.01%	0.80%
Q	41,525	649,294	316,472	84,881	29.16%	31.63%	29.61%	28.68%
R	51,445	304,019	689,600	66,196	36.12%	14.81%	64.51%	22.37%
S	8,775	186,883	4,541	5,720	6.16%	9.10%	0.42%	1.93%
T	11,046	600,982	29,462	43,145	7.76%	29.27%	2.76%	14.58%
U	46	387	20	176	0.03%	0.02%	0.00%	0.06%
V	3	78	2	3	0.00%	0.00%	0.00%	0.00%
Z	455	4,077	925	3,687	0.32%	0.20%	0.09%	1.25%
TOTAL	**142,421**	**2,053,054**	**1,068,906**	**295,910**				

Table 6a (Counts & Percentages) would certainly seem to support the sense given by Tables 5a-1 and -2 in the section above and by the preceding iterations of Table 6a that there should be real and meaningful differences in how each UNCL member expressed its content preferences by subject over the interval. And, again, the numbers support the idea that LC class Q, at roughly 30% of downloads at each location, is the FC's workhorse for UNCL, regardless of campus.

Multinomial logistic regression is used to model and analyze nominal outcome variables in which the log odds of the outcomes are modeled as a linear combination of the predictor variables (Orme & Combes-Orme, 2009). For this analysis, the predictor variables will be the LC classes, with Table 6a's download percentages converted into counts out of one thousand

five hundred downloads per location. The UNCL members will serve as the polychotomous outcome variables. One way to think of what the author is doing here is to imagine that you have four ice cream-loving kids, you have several different flavors of ice cream, and you want to use a percentage measure of the kids' ice cream consumption of the flavors to suss out whether or not each child might have different favorites from the others.*

Before moving forward, the author would like to anticipate a few of the reader's other questions about the proceedings. Firstly, a quick glance at Table 6a (Counts & Percentages) above would suggest that a goodly number of the LC classes produced relatively little activity, so it would intuitively seem that little could be firmly concluded about these classes. The author would agree, so these LC classes will not be treated separately. As was the case with the prior issue of this report, the author will combine the small LC classes into a single category: MinorLC. The LC classes that will comprise MinorLC are C, D, J, K, N, P, V, and U. The remaining nine LC classes produced more activity, so it is to be hoped that something can be concluded about them vis-à-vis UNCL.

If the reader was to look far ahead to Table 6c (Parameter Estimates), the reader would see that the table has nothing to say about the performances of UNK or of LC class *Q – Science*. Readers less familiar with regression may be wondering why this is the case. Regression, as an analysis technique, requires reference groups against which to anchor and compare the behaviors of the other variables, groups, and/or outcomes in the model (Bell & Sanders, 2000; Field, 2013; Gelman, Hill, & Vehtari, 2021; Orme & Combs-Orme, 2009). That is to say, the reference groups provide the context or measure against which to evaluate the performance of the other variables. UNK was chosen to be the reference location simply via alphabetical order, and *Q – Science* was chosen because its performance was close to that of the Freedom Collection as a whole, which made it a good representative case, and because it behaved roughly the same at all campuses, which made it a consistent baseline across UNCL.

---

*As has been the case with almost all of my collections work, the approach here follows a particular branch of consumption theory, Samuelson's (1938, 1948) Revealed Preference Theory (RPT). In the early-to-mid-20th century, Samuelson argued that, in order to make useful and accurate analyses of and decisions/predictions about consumer behavior, one would need to observe and empirically analyze individuals' or institutions' actual behaviors and choices in a decision-making or choice-making environment to infer their actual preferences and purchase/utilization habits. This, of course, stands in contrast to the stated-preferences approach, in which one asks respondents to avow their preferences or to predict their own future choice-making actions, an approach much favored to this day by pollsters, surveyors, and many professionals. So, if as a reader you find yourself objecting to my approach and critiquing it by noting that the author has not asked the librarians nor the faculty for their opinions or thoughts, let me concede here that you are entirely correct. I am taking the RPT stance alone, and I am doing so simply because we, as a profession, tend to so favor the stated-preferences approach that our decision-making processes have frequently seemed to me to neglect consideration of actual usage. So, please do read my work here and elsewhere as an attempt to counterbalance, rather than to ignore or repudiate, the field's strongly favored and well-established approach to gathering input.

Lastly, the reader may be wondering why, with the entirety of the UNCL members' data available, would the author reduce the dataset's counts. This answer has two parts. First, the author wanted to compare the relative enthusiasm of the four campuses for the FC subjects, not the performances of the four campuses themselves. If the author had analyzed the data as provided, it would assuredly have revealed that UNL and UNMC were a lot more active than were UNK and UNO, but the author is uncertain as to whether it would have told us whether UNK's proportional downloading of *R – Medicine* was different than UNL's. Putting the campuses on an equal footing should allow for a focus on proportions.

Second, the author wanted to ensure that whatever statistical differences/effects might arise in the analysis would be real and substantial. As Kahneman, Sibony, and Sunstein have recently discussed (2021), part of the logic of statistical analysis and testing is that one is looking for real effects or outcomes in a dataset (i.e., signal), effects that are so noticeable that they likely cannot be explained away by chance and/or error (i.e., noise). One way for a signal be detectable is for it be large and/or consistent. Another way, because of how many test formulas are constructed, is to greatly increase the size of one's sample. Had the author used the data as provided or set the sample cap to a large number like one hundred thousand or one million, the number of statistically significant differences/effects would likely have increased, but some of these effects would have been trivial and unimportant in the practical sense. For example, the effects of LC class *Z – BIBLIOGRAPHY. LIBRARY SCIENCE. INFORMATION RESOURCES (GENERAL)* with the current cap might not be significant, but if the author greatly increased the download counts, he could push them into statistical significance. This strategy would, essentially, be like looking for differences among the UNCL members' downloads and moving from examining the downloads by eye to examining them with a microscope. At a million downloads, the performance of most every subject would become statistically significant. But, again, those differences would be spurious for practical decision-making. The meaningful points of distinction among the campuses' preferences would be obscured via the inflation of all differences, even the most minor. Hopefully, the employed download cap will be sufficiently large to reveal real, substantial effects among the Freedom Collection's main LC classes and to reveal some of the more outsized effects among the smaller LC classes, but not be so large as to inflate trivial effects into significance.*

---

*The author would be the first to admit that he knows little-to-nothing about statistics, but it seems that the logic of many statistical tests rewards effects for being large as compared to the total variation in a dataset, which is comprised of the hoped-for effect, of the variation unexplained by one's model, and of the error in one's measurements. This relationship between one's desired effects and the total variation is usually tempered by the size of one's sample. Thus, many statistical tests have this sort of logical structure:

$$\frac{\text{Desired Effect}}{\frac{\text{(Unexplained Effects + Measurement Error)}}{\text{Sample Size}}}$$

Sample size's being in the denominator of the denominator of an equation is equivalent to its serving as a multiplier for the numerator, so massively large samples will shrink confidence intervals and/or magnify the size of effects. Hence the author's adjustment of the UNCL Freedom Collection downloads' scale.

After the adjustments to the data, the distribution of downloads by UNCL members were as follows:

Table 6b: UNCL Members' Freedom Collection Downloads (Adjusted for Analysis)
N = 6,000

LC Class*	UNK	UNL	UNCL UNMC	UNO	**Total**
B	61	30	8	85	**184**
C	1	3	0	1	**5**
D	0	0	0	0**	**0**
G	52	32	3	26	**113**
H	152	127	21	272	**572**
J	2	1	0	2	**5**
K	4	1	0	5	**10**
L	32	27	6	57	**122**
N	0	1	0	6	**7**
P	4	3	0	12	**19**
Q	437	474	444	430	**1785**
R	542	222	968	336	**2068**
S	92	137	6	29	**264**
T	116	439	42**	219	**816**
U	0	0	0	1	**1**
V	0	0	0	0	**0**
Z	5	3	2**	19	**29**
Total	**1500**	**1500**	**1500**	**1500**	**6000**

*LC classes in italics will be grouped into the catch-all category MinorLC to facilitate analysis
**These counts were each adjusted by 1 to correct for rounding errors.

Even prior to running the analysis, one can glance through Table 6b and see that there likely are some differences between the UNCL locations' by-subject download activity. One can also see that several of the LC classes have produced very little activity, too little to be effectively analyzed. These LC classes were italicized in the table and will be placed in a catch-all category, MinorLC, to facilitate analysis. These MinorLC classes are: C, D, J, K, N, P, U, and V.

As was remarked on above, the author performed the multinomial logistic regression analysis (MLR) with UNK as the reference group among the UNCL members and with LC class *Q – Science* as the reference group among the predictor variables. As is usually the case, the analysis produced an avalanche of numbers, the first of which involve the model fit (see: Table 6c [Model Fitting]).

In the table, the likelihood ratio chi-square of 1,733.8 with a p-value under 0.0005 tells us that our model as a whole fits significantly better (i.e., reduces estimation error) than does the empty, null-hypothesis model (i.e., the Intercept Only model with no predictors). Thus, we should suspect that there will be at least one significant effect among the predictor variables at the several UNCL locations.

Table 6c (Model Fitting): Multinomial Logistic Regression				
	Model Fitting Criteria	Likelihood Ratio Test		
Model	-2 Log Likelihood	Chi-square	df	Sig
Intercept Only	1886.329			
Final	152.506	1733.823	27	.000
Note: Sig (p) of .000 indicates that p < .0005				

Having established that something has likely occurred among the predictor variables and the UNCL locations, the next step would be to continue the MLR analysis to pinpoint what has happened and where. When perusing the output in Table 6c (Parameter Estimates), the reader should, again, keep in mind that UNK is the reference category for the dependent variables (i.e., the UNCL locations) and that LC class *Q – Science* is reference class for the predictors.

But, before moving on to the parameter estimates table, the author though that it might be helpful to provide a summary table of results for the reader not interested in reading through a long, crowded statistical table.

Table 6c (Output Summary): UNCL Member Relative Preferences			
LC Class (UNK)[a]	UNL	UNMC	UNO
B	--	--	+
G	-	--	-
H	=	--	++
L	=	--	++
R	--	++	-
S	+	--	--
T	++	--	++
Z	-	--	++
MinorLC	=	--[c]	++
Q[b]	=	=	=
KEY (relative enthusiasm) much less --, less -, equal =, more +, much more ++			
a. UNK is the UNCL reference group b. LC class Q is of equally high interest to all UNCL c. A double minus does not adequately express how unenthusiastic UNMC would be for a MinorLC collection (see the parameter estimates)			

The rough results of the multinomial logistic regression analysis, presented in the Output Summary table, suggest that for the UNCL members, the prospect of dividing the Freedom Collection into smaller subject collections might not be entirely agreeable. With UNK as the reference group for the UNCL members and LC classification Q – *Science* as the reference group for the subjects on offer, the UNCL members' relative enthusiasm for each potential subject collection appeared to be quite varied.

In reading the table, one should keep in mind that it indicates the UNCL members' relative, rather than the absolute, interest in each potential subject collection comprised of the journals of the listed LC classes. The previous report, Part 1, would provide a better sense of UNCL members' absolute interests.

Of the UNCL members, the easiest to interpret would be UNMC. UNMC utilizes content from *Q – Science* quite a bit, and its proportional utilization is in line with the other UNCL members'. UNMC's utilization of content from *R – Medicine*, unsurprisingly, is large and is relatively outsized when compared to the other UNCL members'. Its utilization of the rest of the FC content provided in the remaining subjects is comparatively negligible.

After one acknowledges the consortium-wide fondness for content from *Q – Science*, the picture for the rest of UNCL's membership becomes more complicated. The rest of the table shows that UNL's and UNO's relative utilization of the content from the remain LC classes was quite varied. One would expect that UNK, UNL, and UNO would not come to easy agreement were the FC broken up into smaller, subject-specific mini-packages.

If the above summary is insufficient and the reader would like to review the actual results of the MLR analysis, they are presented in the Parameter Estimates table below. To interpret these results, the reader will need to be familiar with three of the outputs: Sig (p value); β, the regression coefficient; and Exp[β], the adjusted odds ratio:

- if a Sig (p value) is > .05, then the Wald Chi-Square value did not meet or exceed the necessary critical value, and the apparent differences in performance, if any, between an LC class's downloads at the listed location and at UNK cannot confidently be distinguished from variation attributable to chance and/or to measurement error. Thus, there should be no real differences between UNK's enthusiasm for the LC class in question and the listed UNCL member's (see: UNL's L);
- if a Sig value is ≤ .05, the regression coefficient (β) is positive, and the adjusted odds ratio (Exp[β]) with its 95% CI is above 1.0, then the likelihood of the outcome (i.e., downloads) increases as one moves from UNK to the location in question. Thus, the listed UNCL member should be more enthusiastic than UNK about the listed LC class;
- if the Sig value is ≤ .05, the regression coefficient (β) is negative, and the adjusted odds ratio (Exp[β]) with its 95% CI is below 1.0, then the likelihood of the outcome decreases as one moves from UNK to the location in question. Thus, the UNCL member should be less enthusiastic than UNK about the listed LC class.

If the reader wishes to use the Parameter Estimates table to make estimates, it will be important to keep in mind that with this type of regression the coefficients (β) are used to estimate movement in the log of the odds for the outcome variable, log(y), rather than to estimate the movement of the outcome variable per one-unit increase in the predictor variable, as is usually the case with linear regression. The log is employed to encourage linearity (Orme & Combs-Orme, 2009). *

If the reader wishes to predict the outcome variable, the reader instead will need to employ the adjusted odds ratio (Exp[β]) (sometimes Incidence Rate Ratio or IRR). To do so, multiply the

---

*The author feels that he should note here that he never employs this particular approach, Multinomial Logistic Regression, in his regular work, so if any of the explanations or analyses appear to be incorrect, the reader should feel free to contact the author, and he will gladly correct whatever errors may appear in this section of the report. It has been quite a while since the author read up on MLR, so he would readily confess to being no expert.

LC class value (Exp[β]) for the UNCL member by UNK's download count in Table 6b. For example: UNK LC B = 60 downloads and UNL's LC B is Exp[β] = .453, so UNL's downloads should equal roughly 60 x .453 = 27.633. The actual UNL value in Table 6b for LC B = 30, so the model estimate was pretty close. The performances of the LC classes by UNCL location were as follows:

Table 6c (Parameter Estimates): Multinomial Logistic Regression
(N = 6,000; n = 1,500/location)

								C.I. for Exp[β]	
UNCL[a]	LC[b]	β	S.E.	Wald[c]	df	Sig[d]	Exp[β]	Lower	Upper
UNL	α	.081	.066	1.502	1	.220			
	B	-.791	.233	11.559	1	.001	.453	.287	.715
	G	-.567	.234	5.854	1	.016	.567	.358	.898
	H	-.261	.137	3.613	1	.057	.770	.589	1.008
	L	-.251	.270	.868	1	.352	.778	.459	1.319
	R	-.974	.104	88.244	1	.000	.378	.308	.463
	S	.317	.150	4.451	1	.035	1.373	1.023	1.843
	T	1.250	.124	102.087	1	.000	3.489	2.738	4.446
	Z	-.592	.733	.652	1	.419	.553	.131	2.328
	Minor	-.282	.454	.385	1	.535	.754	.310	1.838
	Q	0[e]			0				
UNMC	α	.016	.067	.056	1	.814			
	B	-2.047	.382	28.722	1	.000	.129	.061	.273
	G	-2.869	.598	23.042	1	.000	.057	.018	.183
	H	-1.995	.242	67.775	1	.000	.136	.085	.219
	L	-1.690	.450	14.105	1	.000	.185	.076	.446
	R	.564	.086	42.889	1	.000	1.758	1.485	2.081
	S	-2.746	.427	41.412	1	.000	.064	.028	.148
	T	-1.032	.192	28.797	1	.000	.356	.244	.519
	Z	-.932	.839	1.233	1	.267	.394	.076	2.040
	Minor	-21.383	.000	-	1	-	5.17E-10	5.17E-10	5.17E-10
	Q	0[e]			0				
UNO	α	-.016	.068	.057	1	.812			
	B	.348	.181	3.694	1	.055	1.416	.993	2.019
	G	-.677	.250	7.356	1	.007	.508	.312	.829
	H	.598	.122	24.055	1	.000	1.819	1.432	2.310
	L	.593	.231	6.595	1	.010	1.810	1.151	2.847
	R	-.462	.097	22.623	1	.000	.630	.521	.762
	S	-1.138	.224	25.934	1	.000	.320	.207	.496
	T	.652	.133	23.854	1	.000	1.919	1.477	2.492
	Z	1.351	.507	7.097	1	.008	3.862	1.429	10.436
	Minor	.914	.364	6.303	1	.012	2.495	1.222	5.092
	Q	0[e]			0				

a. The reference category is UNK (β = 0; Exp[β] = 1.00)
b. α = Intercept
c. Wald = Wald Chi-Square (a measure of the strength of the "signal" present)
d. Sig = .000 indicates p < .0005
e. This parameter is set to zero because it is redundant (i.e., Q is the baseline predictor category)

References

Bell, W. D., & Sanders, M. S. (2000). *Understanding multivariate research*. Westview.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd ed. Erlbaum.

---. (1992). A power primer. *Psychological Bulletin*, *112*(1): 155-159.

---. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*, 2nd ed. Erlbaum.

Deacon, J. (n.d.). Statistical tests for significance. In *The really easy statistics site*. http://archive.bio.ed.ac.uk/jdeacon/statistics/tress4.html#Transformation%20of%20data

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*, 4th ed. Sage.

Freud, S. (2002). *Civilization and its discontents*. Penguin.

Gelman, A., Hill, J., & Vehtari, A. (2021). *Regression and other stories*. Cambridge UP.

*Gomez, K. A. & Gomez, A. A. (1984). *Statistical procedures for agricultural research*, 2nd ed. Wiley.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown.

Mangiafico, S. S. (2016). Beta regression for percent and proportion data. In *Summary and analysis of extension program evaluation in R* (version 1.19.10). New Brunswick, NJ: Rutgers Cooperative Extension. https://rcompanion.org/handbook/J_02.html

McDonald, J. H. (2014). *Handbook of biological statistics*, 3rd ed. Sparky House Publishing. https://www.biostathandbook.com/

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

Samuelson, P. A. (1938). A note on the pure theory of consumers' behaviour. *Economica*. New Series. *5*(17): 61–71. doi:10.2307/2548836. JSTOR 2548836.

___. (1948). Consumption theory in terms of revealed preference. *Economica*. New Series. *15*(60): 243–253. doi:10.2307/2549561. JSTOR 2549561.

*Sokal, R.R., & Rohlf, F.J. (1995). *Biometry*, 3rd ed. W.H. Freeman.

*---. (1981). *Biometry*, 2nd ed. W.H. Freeman.

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

Stuart, M. (Director). (1971). *Willy Wonka & the chocolate factory* [Film]. Wolper Pictures, Ltd.

*Warton, D.I., and F.K.C. Hui. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology, 92*: 3–10.

---

*Citations marked with an asterisk were cited in an internal version of the report that included an appendix that explained in more detail why Analysis of Variance (ANOVA) was an inappropriate technique for the subject-profile question.