Evidence Based Library and Information Practice

The Socioeconomic Profile of Well-Funded Public Libraries: A Regression Analysis

Michael Carlozzi

Explore this journal

Cite this article

Article abstract

Objective – This study aimed to explore the well-established link between public library funding and activity, specifically to what extent socioeconomic factors could explain the correlation.

Methods – State-level data from the Massachusetts Board of Library Commissioners were analyzed for 280 public libraries using two linear regression models. These public libraries were matched with socioeconomic data for their communities.

Results – Confirming prior research, a library's municipal funding correlated strongly with its direct circulation. In terms of library outputs, the municipal funding appeared to represent a library's staffing and number of annual visitations. For socioeconomic factors, the strongest predictor of a library's municipal appropriation was its "number of educated residents." Other socioeconomic factors were far less important.

Conclusion – Although education correlated strongly with library activity, variation within the data suggests that public libraries are idiosyncratic and that their funding is not dictated exclusively by the community's socioeconomic profile. Library administrators and advocates can examine what libraries of similar socioeconomic profiles do to receive additional municipal funding.

Evidence Based Library and Information Practice

The Socioeconomic Profile of Well-Funded Public Libraries: A Regression Analysis

Michael Carlozzi
Library Director
Wareham Free Library
Wareham, Massachusetts, United States of America
Email: carlotsee@gmail.com

Abstract

Objective – This study aimed to explore the well-established link between public library funding and activity, specifically to what extent socioeconomic factors could explain the correlation.

Methods – State-level data from the Massachusetts Board of Library Commissioners were analyzed for 280 public libraries using two linear regression models. These public libraries were matched with socioeconomic data for their communities.

Results – Confirming prior research, a library's municipal funding correlated strongly with its direct circulation. In terms of library outputs, the municipal funding appeared to represent a library's staffing and number of annual visitations. For socioeconomic factors, the strongest predictor of a library's municipal appropriation was its "number of educated residents." Other socioeconomic factors were far less important.

Conclusion – Although education correlated strongly with library activity, variation within the data suggests that public libraries are idiosyncratic and that their funding is not dictated

exclusively by the community's socioeconomic profile. Library administrators and advocates can examine what libraries of similar socioeconomic profiles do to receive additional municipal funding.

Introduction

I once noticed staggeringly high circulation numbers coming from a particular public library and pointed it out to a senior library director I knew. The notable library served a population almost identical to my own as well as the director's, roughly 22,000 residents. Yet this library circulated over 173 items per hour open in contrast to my library (64) and his (112). I asked the director why he thought this library circulated such volume.

This was his verbatim email reply: "$$$$$$$$$$$$$$$$$$$$$$$$$$$$$"

The light-hearted response turned out to be well-grounded: all three circulation totals corresponded to our ranking in municipal funding. More generally, the Pew Research Center's survey data suggest that wealth correlates with library usage (Rainie, 2016). These data were corroborated by the Institute of Museum and Library Services' (IMLS) Fiscal Year 2011 report, which used statistical modeling to show that in "most cases . . . when investment increases, [library] use increases, and when investment decreases, use decreases" (Swan et al., 2014, p. 1). A subsequent IMLS (2016) report drew similar conclusions, supporting what librarians had long suspected: libraries succeed with financial commitment.

But these analyses cannot determine the extent to which financial investment impacts library usage. IMLS's multilevel growth models, for instance, showed that library use corresponded to differences in financial investment. Yet financial investment might merely measure the size and scope of a library's service population; larger libraries receive more funding to support larger communities. Financial investment also

might just reflect a community's socioeconomic profile. The Pew Research Center's surveys consistently find that wealthier and more educated people use libraries more often than those with lower income and education levels (Geiger, 2017; Rainie, 2016). Thus, library funding and usage might both be effects of the community's overall characteristics.

To try to address these concerns, I analyzed library data from 280 public libraries and confirmed that municipal appropriation strongly correlated with direct circulation. I then included socioeconomic factors for the communities of these libraries to find that the number of a community's "educated residents" significantly affected a library's municipal appropriation, far more than any other socioeconomic factor. However, enough variation existed within the data to reject any "demographics are destiny" arguments—library funding and library usage are not necessarily governed by uncontrollable, socioeconomic factors.

Literature Review

Around the turn of the century, library researchers sharpened focus on library-based assessments. Dugan and Hernon (2002) attribute the change in academic libraries to a shift in priorities as the traditional role of libraries was to "meet the needs of the academic community's information needs" (p. 377). For example, traditional assessment measures (outputs) concerned operating hours and collection space. Given the increase in information literacy demands, however, Dugan and Hernon argue that traditional outputs could not capture the scope, or even existence of, student learning and were even misaligned with assessments; they argue that traditional outputs belong to an

evaluative, not assessment, framework. Thus were born library-based outcomes, which focused on the measurable results of library-based participation (e.g., information literacy gain scores on a pre/post-test).

Public library outcomes tend to focus not so much on learning as on economics. Considerable research has attempted to approximate these economic benefits, with consensus reaching a cost-benefit ratio of around $4 to $1 USD (Aabø, 2009; Bureau of Business Research, 2017; Howard Fleeter & Associates, 2016; Ward, 2008). Similar benefits were found internationally as well (Bundy, 2009). Of course, such a narrow view of "value" cannot capture all of the public library's benefits. Jaeger et al. (2011) summarize several alternative ways to assess value, and McMenemy (2007) argues that an explicitly economic focus ignores the public library's other cultural and societal contributions.

Public libraries in the United States report data either directly to the IMLS's Public Libraries Survey (PLS) or to their state agencies, themselves collectors of data in formats very similar to the PLS. The PLS collects outputs such as a library's circulation, visitations, reference transactions, computer usage, collection size, staffing levels, financial expenditures, and operating hours. These outputs only indirectly measure value; as Holt and Elliott (2003) argue, they "do not represent equal consumption of services or equal value to the library customer" (p. 425). Nevertheless, as Holt and Elliott acknowledge, politicians and stakeholders tend to regard libraries with greater numbers of these outputs as "the best libraries" (p. 425). Much library research, then, focuses on these outputs. The IMLS's own research analyzes circulation, visitations, staffing, financial expenditures, collection size, computer usage, programming, and reference transactions (IMLS, 2016; Swan et al., 2014). Economic analyses of public libraries use the same outputs (e.g., Bureau of Business Research, 2017).

Some research has established a strong correlation between a library's activity, as approximated by the above outputs, and a library's financial investment (Swan et al., 2014). Although academic researchers avoid inferring causation from correlation, non-researchers might not be so prudent, as in Meyer (2016), who argued from an IMLS report that "if libraries receive more public funds, more people use them. . . . If the public wants to reverse the [downward usage] trend and make the local library more useful, it should do the one thing evidence supports: Fund it better" (para. 12). This is a reasonable inference since financial investment facilitates service. As libraries receive more funding they "can have more staff, more classes, more copies of the latest bestseller, and—maybe most importantly—longer hours" (Meyers, 2016, para. 14). McQuillan (2003) drew a similar observation: "more money means more librarians, more books, more magazines, and more open hours" (p. 46).

On the other hand, the theory of public choice, especially Tiebout's model, might posit that library funding reflects community demand rather than causal relationships. Developed by Charles Tiebout (1956), this model imagines "consumer-voters" who choose "the community which best satisfies [their] preference pattern for public goods" (p. 418). The model attempts to explain the economics of public goods by arguing that this "preference pattern" leads to people voting with their feet. While little attention has been given to the theory of public choice in the library literature, Bryce (2003) describes the Tiebout model as allowing for residents to "decide the kind of community they want to live in" (p. 416). Residents who want, for example, excellent library services may vote to raise taxes to support such services. Research in Massachusetts (e.g., Snow, Gianakis, & Haughton, 2015) shows that this effect occurs at the local level. Tiebout's model reflects population shifting; as public expenditure decisions occur, "populations shift and property prices reflect the public choice of the community" (Bryce, 2003, p. 416).

In the Tiebout model, then, financial investments do not necessarily boost library outputs. Instead, higher outputs reflect the desires and voting patterns of specific communities. Residents who disagree with raising taxes to support public libraries will, in theory, oppose such raises or, if they occur, move elsewhere. Bryce (2003) studied this subject in the context of public libraries, surveying American adults about their attitudes toward public library services and attempting to connect these responses to library funding through respondents' zip codes. He found "modest levels of association between demand for library services and library funding support" (p. 422) but largely rejected Tiebout's model. Despite this rejection, Bryce's research has been used to make bold claims regarding the theory of public choice; based on Bryce's work, Stenstrom and Haycock (2015) claim that "the theory of public choice has shown increased use does not correlate to increased funding" (para. 6).

One way to further previous research would be to examine community dynamics directly alongside library activity. The IMLS's reports omit "population demographics, poverty, and community characteristics" (Swan et al., 2013, p. 13). These characteristics might offer insights on library funding and activity. Education level, defined often and in this paper as "the percentage of residents with a Bachelor's degree or higher," shows particular promise. Survey data from the Pew Research Center suggest a connection between education and library usage (Rainie, 2016); college graduates were significantly more likely to report using libraries than non-college graduates by a difference of 17 percentage points (Geiger, 2017).

Political affiliation may also be a useful characteristic, but it shares a complicated relationship with wealth. Gelman et al.'s (2007) multilevel analysis in America, for example, shows that "richer states" support liberal candidates while "richer voters" support conservative candidates, i.e. wealthier voters

within states, *regardless of those states*, tend to vote conservatively. What about voters within local communities? Brett Benson (2012) analyzed and collated the voting patterns of every municipality in Massachusetts from 2006 to 2012 and generated an average margin of victory for liberal or conservative candidates. A score of zero means that the community demonstrated no preference for liberal or conservative candidates across 2006 to 2012. Positive scores indicate a "more liberal" preference and negative scores a "more conservative" preference. In Provincetown, for example, the average score of +73% means that, on average, liberal candidates received 73% *more* of the vote (not 73% *of* the vote) over conservative candidates. Lynnfield, in contrast, scored -28%, indicating that conservative candidates received 28% more of the vote, on average, over liberal candidates.

Data provided by a state-level agency can help further current research lines. Entering community data for individual states creates both a manageable dataset and a simplified analysis, as multilevel modeling will not be necessary to control for unique statewide dynamics. Community data, then, may validate other measures such as the Pew Research Center's surveys. Because state-level library agencies use the IMLS's Public Libraries Survey, intrastate analysis may generalize across at least the United States, if not internationally. As Holt and Elliott (2003) indicate, states hire "staff whose principal tasks . . . are to collect library input and output statistics" (p. 425). The Massachusetts Board of Library Commissioners (MBLC) is one such state-level agency. Turning to the MBLC's dataset, I asked the following research questions:

1) To what extent does a library's funding, specifically its municipal appropriation, account for variation among direct circulation after controlling for library-related variables?

2) To what extent do these library-related variables explain variation among direct circulation?
3) To what extent do community variables used as proxies of library usage (income, education level, age, and political affiliation) correlate with library activity and funding?

Methods

Data Collection

To analyze the relationship between financial investment and library outputs, I relied on data from the Massachusetts Board of Library Commissioners' Fiscal Year 2015 report. Every year, the MBLC releases an extensive report on all Massachusetts public libraries. The data come from Annual Report Information Surveys (ARIS), which library directors must submit to qualify for the statewide certification program. For the MBLC's FY 2015 dataset, 369 separate ARIS reports were released.

Based on the IMLS's Public Libraries Survey, the MBLC's dataset includes all of the usual outputs, e.g., circulation, visitations, and operating hours. Data include financial information such as the library's total operating income, its expenditures, and its Total Appropriated Municipal Income (TAMI), which is the amount of municipal funding received. Overwhelmingly, Massachusetts' public libraries in FY 15 operated from municipal income, as represented by the TAMI as a percent of total operating income (median = 91.8%; mean = 86.2%). This mean closely resembled the national average of 85.7% as reported in the IMLS's FY 13 report.

To represent the library's financial variable, I chose municipal appropriation over total operating income for several reasons. First, municipal appropriation contains fewer potential errors; it is the amount of funding that a municipality apportions its library, appearing in public documents as the library's "line-item"

funding. Total operating income, by contrast, is more of an estimate, meant to include all of a library's income as generated from small donations to large bequests and requires consideration of all grants, donations, and miscellaneous funds bestowed during the fiscal year. Second, within the MBLC's dataset, operating income did not correlate as strongly as municipal appropriation with direct circulation; operating income's $r = .76$ whereas municipal appropriation's $r = .93$. Third, the appropriation represents a municipality's financial commitment irrespective of a library's good fortune, i.e. which libraries have generous individual donors, deep endowments, or vigorous fundraising groups. Appropriation ostensibly measures overall community support better than total operating income.

Not all data reported by the MBLC were used in this analysis. Roughly 80% of public libraries in Massachusetts serve between 2,000 and 99,999 residents. This analysis examined only these libraries because very small and very large libraries skewed results or bore non-generalizable community dynamics. Consider that the average municipal allotment in the entire dataset was $707,882 (median = $368,152) and then consider the Boston Public Library's municipal allotment ($33,416,127). This astronomically high figure would skew the dataset. Furthermore, tiny communities may feature high socioeconomic measures because they are populated by wealthy residents ostensibly uninterested in social services. Alford's population of 474, for instance, has a median household income of $95,313, but with a median age of 57 years, Alford does not represent a typical community. I removed some other libraries from the original dataset because they were presented as independent libraries in a larger municipality. I also removed one municipality, a college town, for its abnormally low median age. The final number of public libraries (*N*) was 280.

Models

I built two linear regression models to analyze the impacts of (1) library outputs on direct circulation and (2) community variables on municipal funding. Regression models are presented alongside their coefficient of determination (R^2) and standard error of the estimate. R^2 refers to the amount of variation within the data explained by the model. All reported R^2 values are the adjusted figures so as to minimize the impact of adding variables. The standard error of the estimate refers to the average amount a model's predictions are "off," or the average distance from an actual value to its estimated value on the regression line.

Selecting independent variables for linear regression model 1 (dependent variable = direct circulation) required some consideration. I could not select variables based solely on the strength of correlation because virtually all library outputs correlated strongly with direct circulation (Pearson's zero-order correlations). This was largely because of confounding variables and collinearity. For example, director's salary correlated with circulation (r = .63) despite having no logical connection to it. When controlling for municipal allotment, i.e. adding it into the model, director's salary becomes nonsignificant (p = .47), and its *partial* correlation—so named because the impact of municipal appropriation is "partialled out"— becomes .001.

Collinearity refers to the correlation between predictors in a model, not between predictors and dependent variables. With high collinearity between variables, the contribution of each variable becomes unclear. One way to measure collinearity is the variance inflation factor (VIF), which estimates the increase in a coefficient's variance from collinearity, where a VIF value of one means "no collinearity." Some collinearity, especially with observational data, is unavoidable. But how much is too much? Convention suggests that VIF values up to five indicate a small-modest level of collinearity but

higher values are more problematic (Stine, 1995). Given the nature of these data, however, modest-high collinearity is unavoidable; an increase in one measure tends to indicate an increase in another. This makes sense. As libraries receive more funding they add more staff, field more reference questions, circulate more items, pay their directors higher wages— essentially, they do more of everything, as both Meyer (2016) and McQuillan (2003) noticed.

I selected variables, then, which were used by the IMLS and other researchers, were logically linked with circulation, and which had low collinearity. These variables represented activities that might realistically affect circulation. The final list of variables for model 1, which met the above criteria, included programs offered (adult and children, annually), total visitors (annually), staff hours (total annually), and physical holdings (total). I did not include electronic holdings since, in Massachusetts, these are often managed at the consortium level.

Despite having a logical connection to circulation and being included in previous research, operating hours were excluded from this model because of their non-linear relationship to circulation. The MBLC awards state aid partially in proportion to the number of hours opened, but state aid is capped. For example, libraries with service populations between 15,000 and 24,999 must open 50 hours per week for maximum state aid, with additional hours yielding no more aid. Libraries lack financial incentive, then, to open more hours than this threshold as suggested by Figure 1.

Linear regression model 2 examined the impact of community characteristics on municipal appropriation (dependent variable), following Swan et al.'s (2014) suggestion that "more could be learned by incorporating other contextual data, such as information on poverty and community characteristics" (p. 13). I added data on these community characteristics based on the
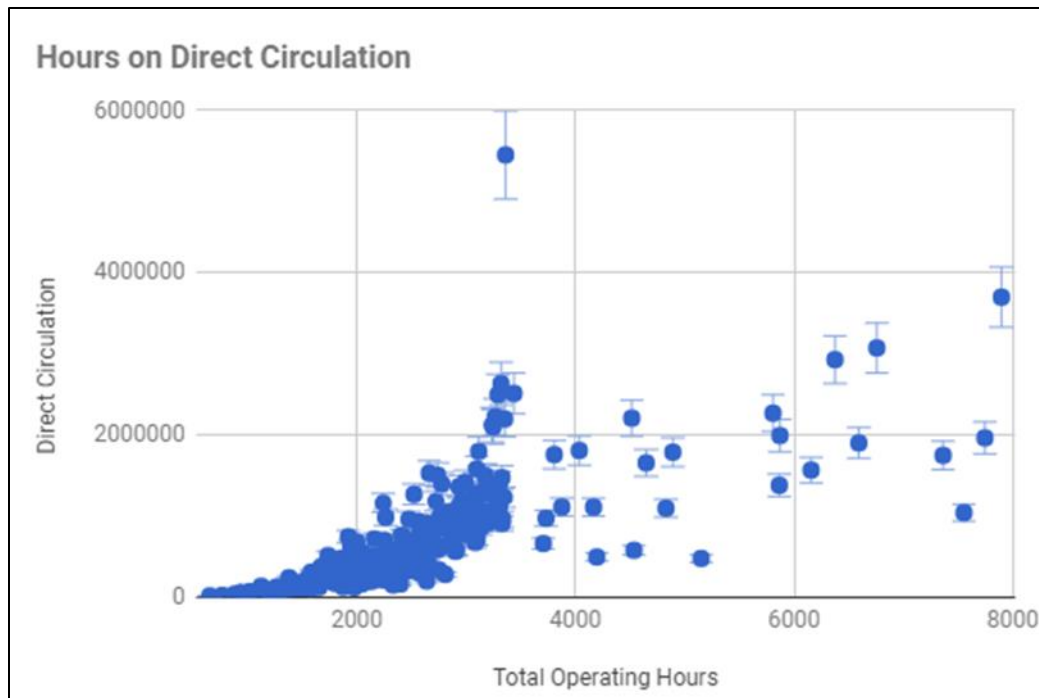
Figure 1
Total operating hours on direct circulation. Note the "wall" created as most libraries reach the threshold to receive the maximum amount of state aid.

latest available census data, either the 2010 U.S. Census or the 2011 or later American Community Survey (ACS), from the American Fact Finder online. Age is represented by the community's median age. Population is the latest available estimate from the ACS. I estimated political affiliation using Benson's (2012) dataset on municipal Massachusetts' voting trends. I chose median family income over median household income because they measured essentially the same construct but median family income correlated better with both municipal allotment and direct circulation; per capita income correlated poorly with both measures.

Education level requires some explanation. Education level (percentage of residents with a Bachelor's degree or higher) and population shared an interaction effect. A model of just population and education level yielded an R^2 of .60, with moderate partial correlations to municipal funding (population $r = .77$ and

education $r = .32$). I suspected, however, that population interacted with education, i.e. gains from population differed depending on education levels. I first centered these two variables around their means and then subtracted the mean from each value to avoid complications from collinearity (Afshartous & Preston, 2011). I then multiplied population by education level to create the interaction term. With the interaction term in the model, substantially more variance was explained ($R^2 = .82$). To simplify model 2, I measured education level by generating a statistic called the "number of educated residents," calculated by multiplying a community's estimated population by its estimated educational attainment (percentage of residents with a Bachelor's degree or higher). This statistic alone explained almost as much variance as the above model ($R^2 = 0.80$), and I used it for model simplicity.

19

Results

As previous research had suggested might happen, municipal appropriation strongly correlated with direct circulation ($r = .93$), by far the strongest individual effect of any variable. Table 1 presents the results of Model 1: library outputs (total visitors, physical holdings, staff hours, number of total programs offered) on direct circulation. Table 2 presents a correlation matrix.

This model explained a considerable amount of variance ($R^2 = .87$) with a modest standard error of the estimate (69,066). Visitors, staff hours, and holdings were all significant predictors. Programs offered was the only nonsignificant predictor on circulation ($p = .13$). It is possible, however, that the effect of programming is so slight that a larger sample size would be required to detect significance. This make sense, as a library's programs reasonably cannot be expected to influence circulation as much as, say, the number of visitors.

The largest effect on direct circulation was the number of staff hours worked (partial $r = .41$). The total number of annual visitors came close (partial $r = .37$). Municipal appropriation and total staff hours correlate extremely well and have high collinearity ($r = .97$; VIF = 15.6), suggesting that they measure a similar construct, although when in the same model, municipal appropriation retains a higher partial correlation ($r = .48$) than staffing ($r = .12$). That may be because staff hours have an empirical limit whereas appropriation does not; even very large libraries eventually reach a critical mass of staff members.

Table 1
Output Variables on Direct Circulation

	Unstandardized B	***P* Value**	**95% Confidence Interval**	**Partial Correlation**
Constant	-45860	<.01	-62434 – -29286	--
Visitors	.53	<.01	.36 – .71	.37
Holdings	.28	.03	.03 – .53	.14
Programs	35.35	.13	-10.94 – 81.65	.10
Staff Hours	279.67	<.01	205.88 – 353.46	.44

$M = 176,544$. $N = 236$. Some libraries were removed for not having submitted data for all included variables.

Table 2
Correlation Matrix of Output Variables and Direct Circulation

	Circulation	**Staff Hours**	**Programs**	**Holdings**	**Visitors**
Circulation	1.0	.92	.67	.83	.89
Staff Hours	.92	1.0	.69	.87	.89
Programs	.67	.69	1.0	.59	.64
Holdings	.83	.87	.59	1.0	.80
Visitors	.89	.89	.64	.80	1.0

Table 3

Socioeconomic Variables on a Library's Municipal Appropriation

	Unstandardized B	**P Value**	**95% Confidence Interval**	**Partial Correlation**
Constant	-23098.23	.23	-607048 – 145093	--
Family Income	.44	.52	-.90 – 1.77	.04
Education	73.15	<.01	67.72 – 78.57	.85
Political	2111.38	.03	213.52 – 4009.24	.13
Age	7446.46	.01	765 – 15658	.15

M = $700,428. N = 280.

Table 4

Correlation Matrix of Socioeconomic Variable and Municipal Appropriation

	TAMI	**Education**	**Family Income**	**Age**	**Political**
TAMI	1.0	.89	.23	-.30	.28
Education	.89	1.0	.26	-.39	.25
Family Income	.23	.26	1.0	.01	-.26
Age	-.30	-.39	.01	1.0	-.08
Political	.28	.25	-.26	-.08	1.0

Table 3 presents results from model 2, and Table 4 presents a correlation matrix on the effects of community dynamics on municipal appropriation. This model explained considerable variance (R^2 = .85) but contained a relatively high standard error of the estimate ($259,768). The number of educated residents had the strongest impact by far (partial r = .85); for every additional "educated resident," the model predicted a $73.15 increase in municipal appropriation. The 95% confidence interval was also fairly narrow, ranging from $67.72 to $78.57.

As with population, I suspected that age might have interacted with education level. Without the interaction effect, age was negatively correlated with appropriation (r = -.30), suggesting that older communities were not as generous as younger ones. (The effect was nonsignificant with other variables in the model, however.) But with the interaction effect in the model, age retained a significant and positive effect (partial r = .15). This measure was not precise, however, with a very wide 95% CI. Income level was insignificant (p = .52) after controlling for education.

Political affiliation was also a significant (p = .03) but with a very wide 95% CI. It did not have a clear interaction effect with education or any other variable. Such imprecision might suggest problems with the dataset. Although Benson's (2012) dataset was extensive, it was not necessarily rigorous; it simply averaged margins of victory across several elections. This might not be a valid way to approximate voting patterns.

Discussion

Previous research has demonstrated a strong correlation between funding and library activity, at least as measured through the variables of circulation and annual visitations. As Swan et al. (2013) found, "[Library] revenue was a positive predictor for visitation, circulation, and program attendance" (p. 13). Drawing on the MBLC's data, I analyzed library usage statistics, extending previous research by including community characteristics. This analysis aimed to learn what municipal allotment might actually measure, for example, a community's income or education level.

In terms of library outputs, direct circulation strongly correlated with both staffing and visitations. Other variables previously studied by the IMLS (e.g., reference transactions and programs offered) indicated little to no correlation after controlling for municipal appropriation or other variables. But this insight, unfortunately, lacks utility. The high VIF (15.6) between staffing and municipal allotment suggests that they may measure the same construct. Advising library administrators to add more staff provides neither clarity nor guidance. We can reasonably infer that libraries hire more staff in reaction to financial increases, something already well known. And, like staffing, visitations are uninformative. We are interested in *why* people visit libraries not that they *do*. Obviously, visitations correlate with circulation totals—as more people visit libraries, more materials circulate.

As the strongest effect on a library's activity was its municipal appropriation, it makes sense to determine what affects this appropriation. This analysis suggests that a library's municipal allotment stems largely from its community's education level; about 80% of the data's variation could be explained by the number of a community's educated residents alone, even after controlling for other influences. Model 2 predicted that each additional educated resident might be expected to increase library funding by

about $73 while holding other variables constant. Interestingly, median family income was found to be nonsignificant when controlling for education level. This may relate to the fact that the examined state was Massachusetts, which is historically the highest-ranking state in terms of educational attainment (Ogunwole et al., 2012). Older or liberal communities were also more likely to receive library funding. These effects were slight, however, and, at least in the case of age, related to education level. Political affiliation may also interact with education level, but this analysis may not have been able to pick it up due to methodological issues (e.g., sample size and limitations of Benson's dataset).

That education influences municipal allotment so strongly suggests that municipal allotment reflects the community's demand for library services, lending indirect and admittedly strictly correlative support for the theory of public choice. Had an income measure been the dominant influence instead of education level, then another explanation may have been more plausible, i.e. public libraries simply benefit from the largesse of their communities. Yet, when controlling for education, median family income did not predict direct circulation. Even without controlling for education, income was a relatively weak predictor ($r = .23$). Many wealthy communities appeared to fund their libraries (relatively) poorly and vice versa. Simply put, the more educated people in a community (in this dataset at least), the higher its public library's funding tended to be, corroborating survey data from the Pew Research Center (Geiger, 2017; Rainie, 2016).

Limitations and Future Research

It should be noted that this analysis relied exclusively on data from one Northeastern, highly educated state. As Swan et al. (2013) indicated, interstate analyses should use multilevel models to consider dynamics unique to each state. Such dynamics may affect the generalizability of these findings. Other researchers could apply socioeconomic analysis

to other states and countries. Furthermore, this research analyzed correlations and thus cannot establish causation. While the data suggest that educated communities drive library funding, this conclusion cannot be drawn and further research would have to examine its feasibility. Previous research by Bryce (2003) found a lack of support for the theory of public choice in public libraries, although Bryce labels his findings as "too preliminary in nature" (p. 423). To further this research line, one might be interested in examining within-subject funding and circulation levels across several years.

Furthermore, the seemingly high R^2 values in these models obscure the correspondingly high standard errors of the estimate. Just because two values correlate does not mean that individual predictions based on the regression line will be accurate. This is a well-documented shortcoming of R^2; Hahn (1973), for example, noted that "unlike the standard error of the estimate . . . R^2 alone does *not* provide direct information as to how well the regression equation can be used for prediction" (p. 611). Indeed, when the socioeconomic regression model predicted municipal appropriation, the average estimate was off by $259,768. That is a very high standard error considering that the average value in this dataset was $700,428. Circulation values similarly had high standard errors of the estimate; in the model of only library outputs, the error was 69,066. Of course, these are average values—some estimates were way off and others were almost perfect—but given that the average circulation total was 176,544, this error comes across as quite high.

However, these high standard errors may matter only insofar as we interpret the data continuously, when perhaps it should be understood as ordinal, similar to a Likert scale. In continuous data, all unit increases are treated equally, justifying the calculation of an average. But this approach may be inappropriate here. To illustrate this concern, consider a public library in Massachusetts with a service population of 23,000 residents. A funding increase from

$200,000 to $400,000 would essentially create a viable public library; $200,000 cannot satisfy statewide certification requirements for a service population of that size. An increase from $400,000 to $600,000, while improving services, would not have the same level of impact as the initial increase from $200,000. And an increase from $1,700,000 to $1,900,000 means even less, given diminishing returns. The high standard errors of the estimate may be deceptive; perhaps what matters is that libraries hit a certain threshold of funding and any variation above that level matters less than variation below that level. Therefore, libraries may be better understood as belonging to certain categories. For example, the difference between $676,076 and $2,127,001 is certainly *numerically* large, but the former library can likely deliver an effective level of public service in a way that even a $400,000 library might not. Further research could explore this relationship in detail.

Nevertheless, all of the data's variation demonstrates the idiosyncrasies of public libraries. In spite of the strong correlations found here, these regression models leave considerable "wiggle room" for librarians, administration, and advocates to impact their communities. Regarding municipal appropriation, community characteristics could not explain almost 15% of the variance—and that 15% appears significant. Swan et al. (2013) reached similar conclusions when arguing that "although revenue is an important piece of the puzzle, it is by no means the only investment that explains changes in library use" (p. 13). These data reaffirm their claim. Poorly funded libraries may try comparing their own communities to communities of similar educational levels and reach out to those libraries to understand how they develop, promote, and deliver services. For instance, two libraries in this dataset have an almost identical number of educated residents (16,453 to 16,936) yet extremely divergent municipal appropriations ($676,076 to $2,127,001). The poorer library could try to discover any notable systemic differences (e.g., a form of

government), and if the poorer library finds nothing substantive, it could contact the wealthier library to try to understand its good fortune and perhaps implement some of the wealthier library's services or approaches.

Conclusion

Municipal allotment appears to operate as a sort of proxy variable, i.e. a variable that approximates some real phenomenon such as a community's interest in its library. This proxy variable is likely the result of many idiosyncratic factors, but the strongest factor was the number of a community's educated residents. More educated communities were more likely to have greater municipal allotments and, in turn, to circulate more materials. However, library advocates should take heart knowing that enough variation existed within the data to allow libraries an opportunity to escape any "demographics are destiny" conclusions. Financial investment appears to be just one part of a large, mysterious puzzle.

References

Aabø, S. (2009). Libraries and return on investment (ROI): A meta-analysis. *New Library World, 110* (7/8), 311–324.

Afshartous, D., & Preston, R. A. (2011). Key results of interaction models with centering. *Journal of Statistics Education*, *19*(3). https://dx.doi.org/10.1080/10691898.2011.11889620

Benson, B. (2012, November 15). How demographic or republican is my town? A partisan ranking of MA municipalities from P-Town to Lynnfield. [Web log post]. Retrieved from http://massnumbers.blogspot.com/2012/11/how-democratic-or-republic-is-my-town.html

Bryce, A. (2003). Public opinion and the funding of public libraries. *Library Trends*, *51*(3), 414–423.

Bundy, A. (2009). Public libraries: It's their funding, stupid. *Australian Public Libraries and Information Services*, *22*(3), 95–96.

Bureau of Business Research IC² Institute & University of Texas at Austin. (2017). Texas public libraries: Economic benefits and return on investment. Retrieved from the Texas State Library website at https://www.tsl.texas.gov/sites/default/files/public/tslac/pubs/ROI_Final.pdf

Dugan, R. E., & Hernon, P. (2002). Outcomes assessment: Not synonymous with inputs and outputs. *The Journal of Academic Librarianship*, *28*(6), 376–380. https://dx.doi.org/10.1016/S0099-1333(02)00339-7

Geiger, A. (2017, June). Millennials are the most likely generation of Americans to use public libraries. *Pew Research Center*. Retrieved from http://www.pewresearch.org/fact-tank/2017/06/21/millennials-are-the-most-likely-generation-of-americans-to-use-public-libraries/

Gelman, A., Shor, B., Bafumi, J., & Park, D. (2008). Rich state, poor state, red state, blue state: What's the matter with Connecticut? *Quarterly Journal of Political Science*, *2*, 345–367. https://dx.doi.org/10.1561/100.00006026

Hahn, G. J. (1973). The coefficient of determination exposed! *Chemical Technology, 3*(10), 609–612.

Holt, G. E., & Elliott, D. (2003). Measuring outcomes: Applying cost-benefit analysis to middle-sized and smaller public libraries. *Library Trends*, *51*(3), 424–440.

Howard Fleeter & Associates. (2016). *The return on investment of Ohio's public libraries & a comparison with other states*. Columbus, OH: Ohio Library Council. Retrieved from http://olc.org/wp-content/uploads/documents/post-id_2060/2016/04/Ohio-Public-Libraries-ROI-Report.pdf

Institute of Museum and Library Services. (2016). *Public libraries in the United States survey: Fiscal year 2013*. Retrieved from https://www.imls.gov/sites/default/files/publications/documents/plsfy2013.pdf

Jaeger, P. T., Bertot, J. C., Kodama, C. M., Katz, S. M., & DeCoster, E. J. (2011). Describing and measuring the value of public libraries: The growth of the Internet and the evolution of library value. *First Monday: Peer Reviewed Journal on the Internet*, *16*(11). https://dx.doi.org/10.5210/fm.v16i11.3765

McMenemy, D. (2007). What is the true value of a public library? *Library Review*, *56*(4), 273–277. https://dx.doi.org/10.1108/00242530710743471

McQuillan, J. (2003). More money, more librarians, more reading: Evidence on funding public libraries. *Knowledge Quest*, *31*(3), 46.

Stine, R. A. (1995). Graphical interpretation of variance inflation factors. *The American Statistician*, *49*(1), 53–56. https://dx.doi.org/10.1080/00031305.1995.10476113

Meyer, R. (2016, April). Fewer Americans are visiting local libraries—and technology isn't to blame. *The Atlantic*. Retrieved from http://www.theatlantic.com/technology/archive/2016/04/americans-like-their-libraries-but-they-use-them-less-and-less-pew/477336

Ogunwole, S. U., Drewery, Jr., M. P., & Rios-Vargas, M. (2012). The population with a Bachelor's degree or higher by race and Hispanic origin: 2006–2010 (ACSBR/10-19). *American Community Survey Briefs*. Retrieved from https://www.census.gov/prod/2012pubs/acsbr10-19.pdf

Rainie, L. (2016, April). Library users and learning. *Pew Research Center*. Retrieved from http://www.pewinternet.org/2016/04/07/library-users-and-learning

Snow, D., Gianakis, G., & Haughton, J. (2015). The politics of local government stabilization funds. *Public Administration Review*, 75(2), 304–314. https://dx.doi.org/10.1111/puar.12317

Stenstrom, C., & Haycock, K. (2015, September). Public library advocacy: An evidence-based perspective on sustainable funding. *Public Libraries Online*. Retrieved from http://publiclibrariesonline.org/2015/09/public-library-advocacy-an-evidence-based-perspective-on-sustainable-funding/

Swan, D. W., Grimes, J., Owens, T., Miller, K., Arroyo, J., Craig, T., Dorinski, S., Freeman, M., Isaac, N., . . . P.& Scotto, J. (2014). *Public libraries in the United States survey: Fiscal year 2011 (IMLS-2014-PLS-01)*. Washington, DC: Institute of Museum
and Library Services. Retrieved from https://www.imls.gov/sites/default/files/publications/documents/pls2011.pdf

Tiebout, C. M. (1956). A pure theory of local expenditures. *The Journal of Political Economy*, *64*(5), 416–424.