Tag, You're It: How Metadata Practices Impact the Discoverability and Use of LibGuides

Stephanie M. Luke, Evie Cordell, Nathan Jeffery, Kate Meyrick, and Thea Soter*

This article discusses the impact that metadata has on the discoverability and usage of Springshare LibGuides at the University of Illinois Urbana-Champaign Library. Using robust linear regression analysis, the authors found that subjects and tags are essential for the discoverability of LibGuides. Guides with more subjects and high-quality tags consistently received greater usage, while the addition of low-quality tags has an overall negative effect on guide visits. These findings highlight the critical role of metadata in resource discovery. The article concludes by recommending the implementation of best practices for LibGuide metadata creation and maintenance. This includes the mandatory use of subjects and tags by guide creators and the standardization of spelling, capitalization, and punctuation in tag creation. The implementation of these practices will help to ensure that LibGuides function as effective, accessible tools for research and instruction.

Introduction

Springshare LibGuides are a popular content management system specifically designed for libraries and educational institutions. LibGuides allow librarians to create, organize, and share customizable resources that support research, learning, and teaching. LibGuides typically include curated materials such as databases, books, articles, websites, videos, and other resources. They aim to provide information on specific courses, subjects, and various research topics. LibGuides are simple to create, easy to use, and act as a central hub for information about the library and university. Additionally, they lower the barriers to sharing resources.

While LibGuides can be a powerful tool for sharing information, several challenges may impact their effectiveness. LibGuides are not static resources, and they require

*Stephanie M. Luke is Assistant Professor and Metadata Librarian at the University of Illinois Urbana-Champaign, email: smluke2@illinois.edu, ORCID: https://orcid.org/0000-0002-5713-2915. Evie Cordell is Assistant Professor and Learning Librarian at the University of Illinois Urbana-Champaign, email: evelync2@illinois.edu, ORCID: https://orcid.org/0000-0001-8446-415X. Nathan Jeffery is Data Analyst and Statistical Consultant at the University of Illinois Urbana-Champaign, email: nathanj3@illinois.edu, ORCID: https://orcid.org/0000-0003-2008-3950. Kate Meyrick is Learning Librarian, Learning Programs and Initiatives, University of Michigan Ann Arbor, email: kmeyrick@umich.edu, ORCID: https://orcid.org/0009-0008-0137-7746. Thea Soter is Rare Books Assistant Curator, J. Willard Marriott Library at The University of Utah, email: theadora.soter@utah.edu, ORCID: https://orcid.org/0009-0006-8607-2233. ©2026 Stephanie M. Luke, Evie Cordell, Nathan Jeffery, Kate Meyrick, and Thea Soter, Attribution-NonCommercial (https://creativecommons.org/licenses/by-nc/4.0/) CC BY-NC.

consistent maintenance as information can quickly become outdated. Proper maintenance is time-intensive and includes updating content, checking links, and ensuring that guides meet evolving accessibility standards. Other challenges include inconsistent quality of guides (in both content and design), insufficient or excessive volume of content, duplication of content, and issues of usability and accessibility (Kamela & Higgins, 2024).

The University of Illinois Urbana-Champaign Library (University Library) is located at a major R1 land-grant institution with multiple libraries throughout campus. The library employs approximately 400 staff, including tenured and tenure-track librarians, academic professionals, civil service employees, graduate students, and undergraduates. In 2023, librarians at the University Library began the process of developing a plan for a thorough review of the library's LibGuides. As of October 2024, when the authors undertook the work described in this article, the University Library had 82 unit accounts and 2,329 guides, including 1,154 published guides, 820 unpublished guides, and 355 private guides.

The review of LibGuides focused on various aspects of revision but also sought to determine the best use of limited resources, particularly the capacity of librarians to enact suggested changes. An area of interest for the authors (a metadata librarian, eLearning librarian, data scientist, and two MSLIS graduate students) was the use of metadata in LibGuides. This was a crucial part of the larger review of LibGuides at the University Library as metadata can affect the discoverability and use of these resources. Upon the initial review of the data from Springshare, the authors noticed that many guides lacked subjects and tags or that the quality of metadata was less than ideal. As a part of this large-scale review of guides, the authors designed and conducted an audit of LibGuide metadata to determine whether review and revision work on legacy guides should focus on improving metadata use or if effort would be better spent focusing on another dimension of improvement.

While past literature on the topic of LibGuide quality has focused primarily on usability testing of the content and appearance of guides (e.g., layout, organization, clutter, and accessibility), there is less research on exploring issues of discovery and access. Springshare's LibGuide platform functions primarily on full-text searching of guides. It also employs subjects and tags, which are designed to expand the discoverability of and access to these resources. In this article, the authors refer to subjects and tags collectively as "metadata," as these are additional, descriptive access points that creators employ to characterize the content of the guides and facilitate discovery. This article shares the findings of the metadata audit, which was designed to assess the use and quality of metadata in LibGuides at the University Library. The project evaluated two dimensions of guide metadata: first, the quantity of metadata and its effects on guide usage (as measured by the number of page views), and, second, the quality of metadata and how the content of tags might be improved to further facilitate discovery. The authors evaluated the quality of tags by reviewing and noting characteristics that would negatively affect their searchability. These characteristics pose a challenge for discoverability (i.e., the degree to which users can locate content in their research) and reuse (i.e., the extent to which tags can be replicated across resources) (Somerville et al., 2012; FAIR GO, 2016). The presence of those characteristics that negatively affected searchability and filtering indicated low-quality metadata. The absence of these characteristics indicated high-quality metadata.

This article describes the methodology of the audit, the statistical analysis of the gathered data, and the findings of this research. The authors conducted a statistical analysis using a robust linear regression model, which allowed them to determine relationships between

multiple explanatory variables and understand how these variables might impact guide usage. They determined that, overall, metadata quantity has a positive relationship to guide usage. In other words, guides with more subjects and high-quality tags tended to have more usage, measured by guide visits. In contrast, the use of low-quality tags has an overall negative effect on guide visits.

Through the evaluation of this data and a discussion of these findings, the authors argue that best practices for LibGuides should include guidelines for metadata use and reuse. Since adopting Springshare in 2007, the University Library has never established practices for the creation and use of metadata in its LibGuides. The findings of this audit and the statistical analysis of the gathered data clearly indicate that ensuring consistent use of quality metadata should be a priority for librarians creating and maintaining guides. This article further discusses the initiative to create these best practices and ensure their implementation across the University Library.

Literature Review

Studies have found that, despite the sheer quantity of LibGuides created by academic libraries, student engagement with these resources is low (Castro-Gessner et al., 2013; Burchfield & Possinger, 2023). Low student use may be the result of several factors, including usability (Dalton & Pan, 2014; Fagerheim et al., 2017), accessibility (Greene, 2020; Hopper, 2021), and discoverability (Sittler & Dobbs, 2016; Reeb & Gibbons, 2004) of LibGuides.

Research on Springshare's LibGuides has extensively explored the learning tool's usability affordances and limitations. The term "usability" has been used broadly to refer to vastly different aspects of use across studies (Neuhaus et al., 2021; Staley, 2007; Burchfield & Possinger, 2023; Almeida & Tidal, 2017; Sonsteby & DeJonghe, 2013; Ray, 2025; Chan et al., 2019; Conerton & Goldenstein, 2017; Bergstrom-Lynch, 2019; Pratt et al., 2004; Pittsley & Memmott, 2012). For instance, one study defines usability as navigation, finding that students prefer side navigation to top navigation and two columns of information (Thorngate & Hoden, 2017). Another study found that students defined usability to include how the information was relayed, preferring multimedia LibGuides and interactive content over plain text (Almeida & Tidal, 2017). Additionally, LibGuide literature has determined that consistency of guides across an institution is integral to student use (Alverson et al., 2015; Fritch & E. Pitts, 2016; Cox, 1996; Vileno, 2007; Sittler & Dobbs, 2016; Duncan et al., 2015). Consistency refers not only to the same color and font choices but also to consistent language use across the platform (Thorngate & Hoden, 2017; Fritch & Pitts, 2016).

Past research has found that students do not browse LibGuides intentionally but rather locate them through using course management systems, entering the guide address, or searching on the library website or through Google (Conrad & Stevens, 2019; Nielsen, 2001; Carey et al., 2020; Griffin & Taylor, 2018; Brush, 2022). As Brush (2022) writes, "To improve discoverability, all guides needed to include metadata like guide and page descriptions, subjects, tags, and 'friendly' URLs. ... These options had always been available in the LibGuides software, but many guide authors had ignored them because they were not thinking in terms of SEO" (p. 193). Springshare's tagging mechanism (in which a creator assigns subjects and tags) seems to be underutilized by guide creators (Hennesy & Adams, 2021; Brush, 2022). Much of library metadata practice has traditionally centered on providing a structured, taxonomic approach to organizing information through shared, community standards (Park, 2009; Park et al.,

2015; Riley, 2017; Zeng & Qin, 2020; Lubas et al., 2013). In contrast, LibGuide creators assign metadata with few (if any) guidelines about how to use tags, subjects, and descriptions, all of which may improve the discovery of these resources. Thus, LibGuide metadata functions more as a folksonomy than a metadata standard. Unlike defined metadata standards, which rely on consistent frameworks, folksonomies depend on user-generated descriptions (Noruzi, 2007; Lund & Hansen, 2011). While folksonomies offer a lower barrier method of metadata creation than traditional taxonomies, they present many challenges to search and discovery, including ambiguity, synonymy, and differences in granularity of terms (Bullard, 2018; Golder & Huberman, 2006; Trant, 2009).

Springshare Content Management System (CMS) provides a LibGuide that outlines a set of standards, best practices, and fundamental usability "ideas" for creating usable guides. The broader ideas are "less is more, fewer clicks, less text, readability, and consistency" (Springshare CMS, 2025). Although Springshare provides these examples of best practices for implementing its LibGuide software, libraries are not required to adopt them; rather, Springshare suggests that libraries use its examples as a template for making their own. Springshare's suggestions include web design elements like branding, color, text, font, and page design, as well as accessibility practices such as headings, hyperlinks, alternate text, textual description, captions, and documentation. Ideally, each library would create similar categories of best practices; however, the literature suggests that many libraries neither create their own set of best practices nor require the implementation of said best practices (Logan & Spence, 2021). Other studies have noted the problem of consistent rules and application of best practices (Logan & Spence, 2021; Fritch & Pitts, 2016; Del Bosque & Morris, 2021). Past literature has indicated that there is not a universal consensus on what best practices should be followed, which contributes to vastly different design practices across academic libraries.

Methodology

The authors began this project by gathering all University Library LibGuide data through Springshare. The authors collected data from August 2023 to October 2024. They chose these dates to include data for an entire academic year, including the summer semester. They also included the two months of the fall 2024 semester when they began planning the project. This period allowed them to capture data from various semesters and gain a complete understanding of how LibGuides were being utilized at their institution. The reports included statistics, general information (e.g., creation date, owner, etc.), status (i.e., published, unpublished, private), and metadata (i.e., subjects and tags). After running the reports, the authors merged the LibGuide statistics (i.e., usage stats and names of guides) with the information report (i.e., creation date, latest update, owner, type of guide, and name of guide). The merging of this information allowed the authors to compare salient data points and identify trends in guides and usage. Some of the data the authors considered was the number of guides, the age of guides, and the status (i.e., published, unpublished, or private).

The authors sought to determine both quantitative and qualitative relationships from this data. First, was there a relationship between the use of subjects and tags and the overall usage (defined as the number of user visits to guides)? Second, was there a relationship between the quality of tags and overall guide usage? The authors focused their quantitative analysis on published guides, meaning those guides that were available to the public. Of the 2,329 guides that existed at the time of analysis, 49.5% were published, leading to a total of

1,154 (N) guides. Notably, the number of guides fluctuates with the creation and deletion of guides, so this dataset represents a snapshot of one point in time; between the time when the data was analyzed and this manuscript was finished, approximately 400 guides had been deleted or significantly edited as part of a larger movement of ensuring the quality of guides and increasing their accessibility.

The authors also pulled data on the metadata (i.e., subjects and tags) used in the University Library's LibGuides. They first focused on data for the use of subjects across guides. Subjects are predetermined by LibGuide administrators and are meant to represent broad subject headings like history, African American studies, math, and global studies. LibGuide creators at the University Library are not required to choose any of the 73 subjects; however, they can assign one or more subjects to their guides. On the library's LibGuide landing page, guides are arranged by subject. Thus, if a creator does not assign a subject, the guide can only be discovered through searching.

The authors then focused on the data for tag use across guides. Like subjects, tags are not required for guide creation. In contrast to subjects, however, tags are creator-supplied and free-text rather than a controlled vocabulary set by administrators; thus, tags have more variation in their creation, use, and reuse. Guide creators add tags to improve searching and provide additional access points for keywords that may not appear in the body of the guide. The authors pulled data on the tags that appear across all guides and created a spreadsheet for the specific qualitative analysis of this metadata. For their analysis, they identified several key aspects to analyze and compiled the data into a spreadsheet (see Figure 1). To ensure the reliability of the data, two of the authors gathered and organized the data and two other authors reviewed the data separately and checked the results against each other. Any errors that were discovered were checked against the original data to ensure accuracy. Any disagreements between reviewers over the assignment of tag characteristics resulted in a separate review by the authors who originally organized the data.

The authors noted several characteristics of tags in their review. These included orphan tags, tags with duplicative content, tags that introduced additional characters that affected searchability, and tags with misspellings (see Table 1). The latter two of these data points—the introduction of additional characters and misspellings—were used to indicate "low-quality" metadata. While there is no single definition of what constitutes "quality metadata," Park (2009) writes that "Quality metadata reflect the degree to which the metadata in question perform the core bibliographic functions of discovery, use, provenance, currency, authentication, and administration" (p. 213). The first two of these aspects of quality—discovery and use—are applicable to LibGuide subjects and tags. Quality metadata is important to the discussion of all digital objects, including LibGuides, because, as Luke et al. (2024) note, metadata "is the primary means through which digital objects are discovered... [and] without quality metadata,

ID	Name	Count	Orphans	Duplicative	Characters	Misspellings
256321	disertations	1	☐	☑	☐	☑
256195	dissertation	2	☐	☑	☐	☐
319955	dissertations	4	☐	☑	☐	☐
256270	dissertations_and_theses	2	☐	☑	☑	☐

Figure 1. A Snapshot of the Spreadsheet the Authors Used to Evaluate the LibGuide Tags.

TABLE 1
Examples of Common Tagging Issues in Guides

Tag Issue	Definition	Example
Orphan tags	Tags that appear in zero active public guides or are remnants from previously deleted guides.	Bibliometrics
Duplicative content	Tags that are variations that seek to convey the same content.	language learning language_learning language learnings
Introduction of additional characters*	Tags that introduce characters that affect their searchability.	asian_american_studies (use of underscores) anti-discrimination (use of hyphens) Asia-Pacific; Japanese Studies; History (use of semicolons) hero's journey (rendering error)
Misspellings*	Tags that contain misspellings.	coronoa virus

Note. Asterisks (*) indicate those tag issues that the authors defined as low-quality tags.

digital objects are lost in a sea of content" (p. 149). As this is the first study (that the authors are aware of) that investigates LibGuide metadata in this way, it is not possible to compare the validity of the current approach to existing measures of metadata quality in these resources. However, future analyses could investigate the effect of duplicative tags or other dimensions of metadata quality on guide usage.

The authors explored the effects of metadata quality on LibGuide usage in two ways. First, by including "high-quality" and "low-quality" tags separately in a linear regression model, they sought to determine if there was a difference in the overall usage of guides based upon the number of high-quality tags (defined as those tags that did not introduce additional characters or have misspellings) versus the number of low-quality tags (defined as tags with these issues). Second, as previously mentioned, the authors performed a qualitative analysis of tags to identify the problems that could affect their searchability and reuse across guides (see Figure 1; Table 1).

The authors used this data first to run an ordinary least squares (OLS) regression model to estimate the effects of type, subject, and tags on the usage of a LibGuide. The model can be expressed by the following equation:

$$Usage = \beta_0 + \beta_1 TypeSubject + \beta_2 TypeCourse + \beta_3 Subjects + \beta_4 HQTags + \beta_5 LQTags + \epsilon$$

Where β_0 is the intercept, the β_i coefficients indicate the effects of each of our predictor variables on the usage of a guide, and ϵ is the error term. In this model, the authors collapsed type into three categories (subject, course, and other) expressed by the two dummy variables, *TypeSubject* and *TypeCourse*. Tags are split into "high-quality" tags (*HQTags*) and "low-quality" tags (*LQTags*) based on the authors' previous discussion of tag quality.

After running the OLS regression, however, the authors found that this regression did not fit the data especially well. Particularly, the overall F-test was nonsignificant, and the diagnostic plots showed that outliers strongly influenced the model estimates. Therefore, they then ran a robust regression on the same variables using Tukey's bisquare estimation, which

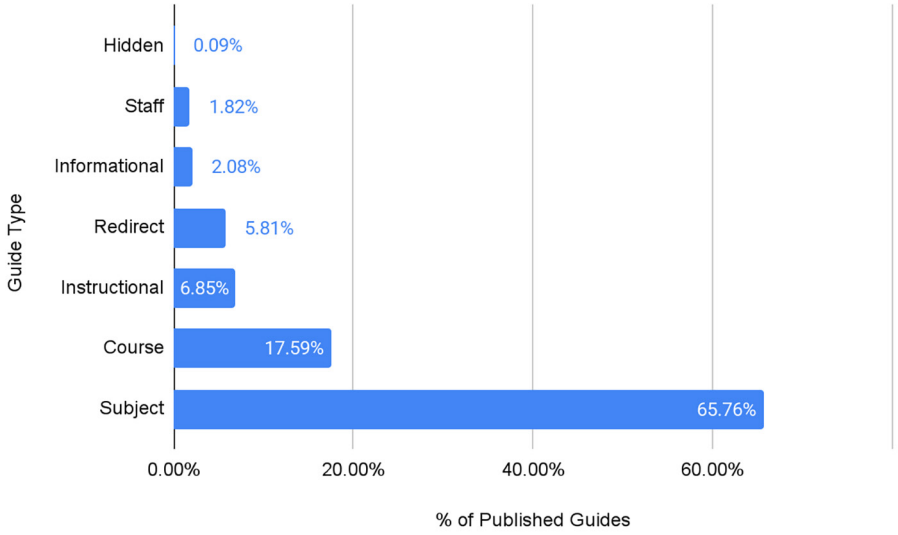


Figure 2. Prevalence of Guide Types Among Published Guides.

decreases the influence of outliers and corrects for the non-normality of the data. The results section includes a discussion of both the OLS regression and the robust regression.

Results

The authors first determined the prevalence of each type of guide (see Figure 2). This was an important step in the statistical analysis because the findings could be used to set priorities for the revision of existing guides, as there are limited resources in the University Library to conduct this work. The LibGuides at the University Library are one of seven types: hidden, staff, informational, redirect, instructional, course, and subject. Hidden guides (0.09% of all guides) are published guides that have been hidden from search and are only available if a user has the URL. Staff guides (1.82%) are intended to be used by University Library staff and often detail internal workflows. Informational guides (2.08%) provide information on an event, specifically exhibits, workshops, conferences, panels, etc. Redirect guides (5.81%) have no content and redirect the user to another guide or webpage. Instructional guides (6.85%) provide instructional frameworks for a specific topic. Course guides (17.59%) are designed for a course that is taught at the University. Finally, subject guides (65.76%) are a general list of resources centering on a specific subject or topic.

In addition to guide type, the authors reviewed tag usage. The limitations of the Springshare report system only allowed them to review guides that had tags. In their review of the tags, the authors noted that there were 2,808 unique tags used 6,032 times across all guides. The average number of times a tag was used was 2.14, but the median and mode of tag use were both one (see Table 2). The outlier of tag usage is "history," which appears across 159 guides.

The authors also made note of tag reuse. Reuse of metadata is important as it facilitates greater discoverability and access across resources. While there is flexibility in creating tags, they function best when

TABLE 2
Data on Tag Use and Reuse

Data Point	Number
Number of unique tags	2,808
Number of times tags are used	6,032
Average times a tag is used	2.14
Median number of tags usage	1
Mode number of tags usage	1

they are reused across multiple guides. The data indicates that creators are largely not reusing tags or are using tags that are similar but not identical. Of the 2,808 unique tags used across all guides, 9% are orphan tags used zero times (likely from deleted guides), 59% of tags are used only one time, 29% of tags are used between two and nine times, and 3% of tags are used 10 or more times.

TABLE 3
Examples of Tags with Duplicative Content

Term 1	Term 2	Term 3
Arms Control	brain_injury	consumer
arms control	traumatic_brain_injury	consumer information
arms_control	ABI	consumerism
	Acquired brain injury	consumers

Because tags are uncontrolled, creators enter similar phrasing into the system, which renders the tags unique. In reviewing tag metadata, the authors noted duplicative tags, where the definition or content of the tag was similar but the tag was phrased differently by creators. The authors found that 919 tags, or 32.7% of all tags, had duplicative content. Table 3 highlights terms that are similar in content but appear as different tags across guides. This has an impact on tag reuse across guides, as duplicative content affects the searchability of the tag and the ability of these tags to be faceted and indexed in the system.

Two of the variables the authors included in the regression analysis related to the quality of tags: misspellings and the introduction of additional characters. The presence of either of these issues in tags characterized the tag as "low-quality." The authors discovered 24 tags included misspellings (e.g., coronoa virus, research proprosal, muisic, and sholarly support). They also discovered that 693 tags introduced additional characters (e.g., hyphens, underscores, ampersands, rendering errors, and semicolons). Thus, more than 25% of all tags were "low-quality," a significant number that could affect the searchability of these tags and the discoverability of their associated guides. The presence of these quality issues also affects the reuse of tags across guides. For instance, if a user selects a tag "arms control," the tags "arms_control," "arms-control," and the misspelled "amrs control" may not appear as a search return. Furthermore, the tags will facet differently in the "filter by tag" feature of Springshare's search system.

Table 4 presents the descriptive statistics for the main variables of interest. Type is described separately in Figure 2.

The presence of outliers, or a few LibGuides that differed significantly from the majority of other LibGuides, had a significant effect on the means and standard deviations of all variables in the model. For all variables in Table 4, the standard deviations are larger than the mean, and the mean is noticeably greater than the median. Even before running the regression model, these descriptive statistics alerted the authors to possible violations of the core assumptions of linear regression, such as the normality of residuals.

TABLE 4
Descriptive Statistics for Variables in the Regression Analysis

Variable	Mean	SD	Median	Min	Max
Usage	1,546.74	15,029.88	272.5	0	498,051
Subjects	1.16	1.69	1	0	30
HQ Tags	2.69	3.27	2	0	37
LQ Tags	0.71	1.34	0	0	10

After the authors ran the linear regression, diagnostic plots (see Figure 3) confirmed that the outliers had a disproportionate effect on the regression estimates. Specifically, all three plots identified the "Anime and Manga" LibGuide as an outlier, and the Cook's distance plot also identified the "Finding Images" LibGuide as an outlier (N.B. In the plots in Figure 3, "Anime and Manga" is the point marked "5" because it was the fifth row in our dataset; "Finding Images" is the point marked "166"). The "Anime and Manga" guide, which has 498,051 views and zero tags and is listed in six subjects, is a notable outlier as this topic is popular. Users may also use the guide relatively easily through subject and full-text search as this is the only guide on this topic on the University Library LibGuides platform. Moreover, the guide seems to garner significant traffic from Google, as it is the first result in a search using the terms "library anime manga." The "Finding Images" guide, which has 1,392 views and 11 tags and is listed in 30 subjects, seems to be receiving significant traffic because it includes the highest number of subjects across all the University Library's guides, approximately 40% of the 73 choices of subjects. Both are legitimate data points that should not be omitted from the dataset, but they should not have such a strong effect on the results.

To better handle the outliers, the authors then ran a robust regression with Tukey's bisquare estimation, which weights observations to lessen the influence of outliers without excluding them entirely. This approach reduces the influence of observations that fall far

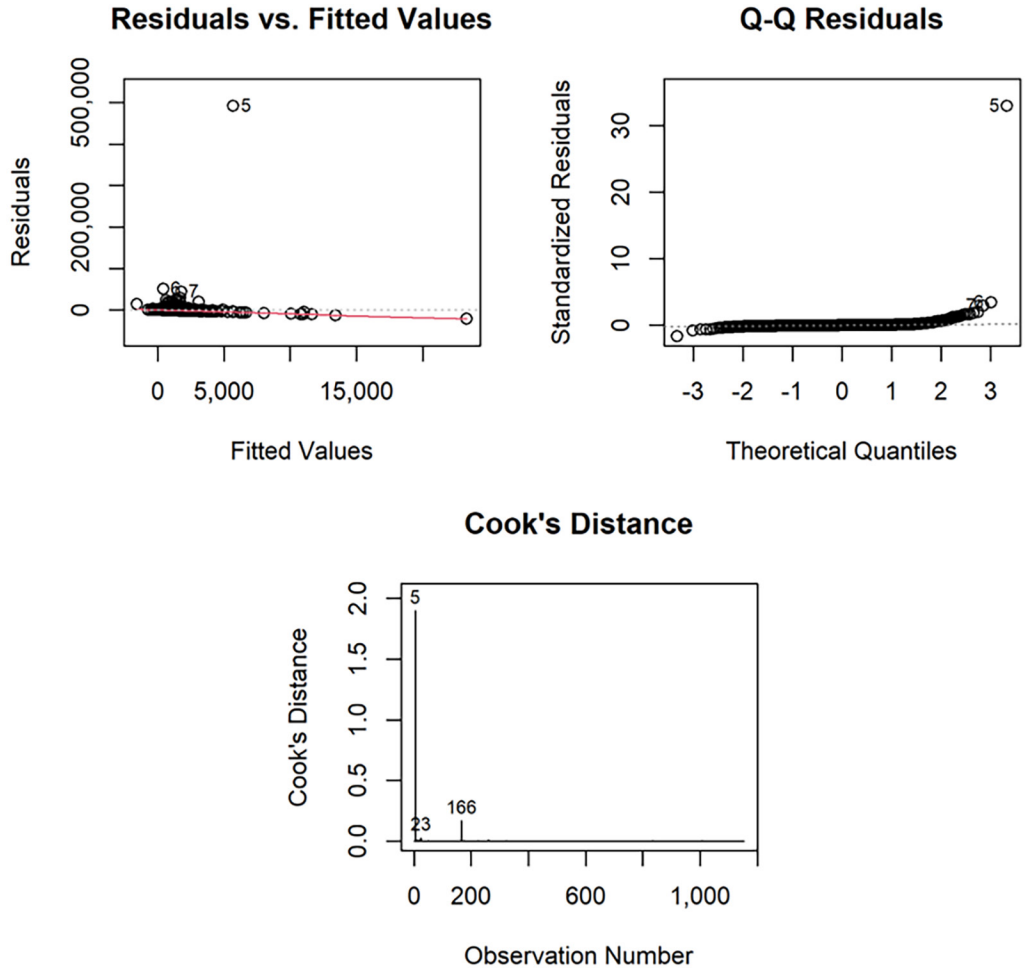


Figure 3. OLS Regression Diagnostic Plots.

TABLE 5
OLS and Robust Regression Results for LibGuide Usage

Variable	OLS Estimate (SE)	Robust Estimate (SE)
Type	–	–
Subject	86.07 (1216.26)	167.77*** (19.96)
Course	-953.22 (1520.43)	64.91** (21.58)
Subjects	769.36** (259.65)	36.75*** (4.53)
Tags	–	–
High quality	-104.44 (140.64)	13.15** (4.27)
Low quality	44.96 (335.45)	-15.34** (5.23)
Observations	1,154	1,154
Adjusted R-squared	0.004	0.185
Test statistic[†]	1.832	180.59***

***p <0.01, **p <0.05, *p <0.1
[†]The OLS regression uses an F statistic; the robust regression uses a chi-squared statistic.

outside the range of most data points and corrects for the non-normality of the data. The robust regression was conducted using the robustbase package in R, which provides appropriate standard errors and significance tests for this method. The robust estimates are provided alongside the OLS estimates in Table 5, and the R code is included in the Appendix.

The OLS regression was not significant overall ($F_{5,1148}$=1.83, p = 0.104), likely due to the outliers discussed above. In other words, the information provided by a LibGuide's metadata did *not* make it possible to predict the usage of that LibGuide any more accurately than could be done using just the mean value of LibGuide usage.

In contrast, the robust regression model was significant overall (χ^2_5 = 180.59, $p < 0.001$), meaning that this model fit the data well. Moreover, all variables in the robust model were significant predictors of LibGuide usage, and all variables except LQTags had positive coefficients, indicating that higher quantities of metadata were associated with more LibGuide usage. When looking at the type of a LibGuide, "subject" and "course" type LibGuides both had higher usage on average than other types of LibGuides.

In the robust model, the coefficient for subjects was 36.75, indicating that adding a subject to a LibGuide is predicted to increase the usage of that guide by roughly 37 views. For tags, each additional high-quality tag was predicted to increase the usage of a guide by 13.15 views, whereas adding a low-quality tag was predicted to decrease the usage by 15.34 views. While unstandardized coefficients in a regression model cannot normally be compared directly, the amount of time it takes to add a subject is comparable to the amount of time it takes to add a tag. Therefore, where practical, adding subjects to LibGuides is predicted to have a larger impact than adding tags.

Discussion

In analyzing guide metadata and usage, the authors discovered that LibGuide creators do not always add metadata to guides or that they format the metadata in a way that does not optimize the searchability of these terms; both are factors that affect how easily these resources can be discovered. Creators are also inconsistently using subjects and tags, which

has a notable impact on the discoverability of guides, sometimes making them difficult to find. While LibGuides are a valuable resource, they are only effective if users can locate them.

The findings of the robust regression model (which fit the data better) supported the authors' hypothesis that a higher quantity of metadata would lead to higher LibGuide usage. As previously mentioned, the robust regression model found that the addition of a subject resulted in an average of 36.75 more views of a guide, while the addition of a high-quality tag resulted in an average of 13.15 more views. In consideration of the limited resources of most libraries, librarians working with LibGuides may want to prioritize remediating subject usage in their guides. This can be accomplished by reviewing existing guides and adding one or more relevant subjects. The addition of subjects is significant to discovery, as guides are most often sorted by subject on the main LibGuides page. This creates notable access points for guides and guides without subjects are discoverable only through search.

Based on the findings of the robust regression analysis, the quality of tags also affected guide usage. As previously mentioned, the addition of a high-quality tag resulted in, on average, an increase of 13.15 guide views. The addition of a low-quality tag had the opposite effect, resulting in, on average, 15.34 fewer guide views. Both results were statistically significant and indicate the importance of considerations of metadata quality in addition to metadata quantity. If a low-quality tag were removed and replaced with a high-quality tag, the result would be an estimated increase of 28.49 views, which is closer to, but still less impactful than, adding a subject. If a library has to prioritize adding subjects or tags, the data suggests that remediating subjects will have a greater effect overall and thus should be addressed first. The authors suggest that librarians first prioritize formalizing workflows for tag use in new guide creation and then later focus on returning to existing guides and remediating and expanding the use of tags.

The authors did an extensive evaluation of how the unique rendering of tags affects their searchability. For misspelled tags, unless a user also misspells a term in their search, these will return zero search results because of the limitations of Springshare's discovery system and the lack of fuzzy searching (or approximate string matching), a search technique that will return results for concepts similar but not exact to a search query (Murty & Biswas, 2004). As mentioned, the presence of additional characters can affect the searchability of the tag as well as its reuse across guides. While two tags may contain the same word, the introduction of additional characters affects the search returns for the phrase. Furthermore, tags with additional characters facet differently in the "filter by tag" feature of Springshare's search features. From communication with Springshare, the authors also learned that hyphens function differently in searches than underscores or spaces. In searching, hyphens function as the Boolean operator "not." Thus, the creator of a guide may input the tag for the course Biology 201 as "biology-201," but searching this exact hyphenated term on the platform will remove all references to 201 from the results.

This study highlights the need for future research that gauges the effect on guide usage of revising low-quality tags to make them high-quality; however, the work of remediating tags is much more time-intensive than subject remediation. The limitations of Springshare's system make the large-scale cleanup of tags challenging. Batch editing tags would allow the authors to make these changes, but this function is currently limited. Batch editing could also assist in the merging of duplicative tags, which may also have a positive impact on the reuse

of tags across guides. The authors can currently only edit tags either through the editing of individual guides or through remediating tag by tag. In the case of the latter, however, many edits, such as merging tags and removing additional characters like hyphens and underscores, were not allowed by the system. The authors intend to request additional search functionality in the Springshare platform. Springshare's lack of fuzzy searching, advanced searching, and controlled vocabularies diminish the discoverability of guides.

While LibGuides are often among the most used and extensive forward-facing publications for a library, there is a lack of consistent peer review in their development (Logan & Spence, 2021). The results of this analysis show that the lack of shared best practices in creating LibGuides metadata is a significant obstacle to their discovery. These issues underscore the need for establishing best practices in metadata creation and providing training for guide creators on optimal subject and tag use. The LibGuide team housed within the Teaching, Learning, and Academic Support (TLAS) unit is responsible for LibGuide maintenance and training at the University Library. Based on the findings of this metadata analysis, the team is developing a comprehensive plan to improve LibGuide organization and discoverability. The TLAS unit will review and revise its existing best practices and current LibGuide instruction workshops. The unit is also planning to collaborate with the metadata librarian to create standardized metadata requirements for all new guides, including mandatory use of subjects and tags, along with consistent capitalization and punctuation rules. These standards will help reduce duplicative content, encourage tag reuse, and enhance the discoverability of similar guides across the library's collection.

To implement these changes across the library, the LibGuide team plans to establish a peer-review process for guides and will implement the publishing workflow feature in Springshare. The team is also developing an accessibility and quality control checklist that all guides will need to meet before they can be published. When the publishing workflow feature in Springshare is activated, all unpublished guides are sent to designated reviewers before they are published. Reviewers will assess the guides to ensure they meet each of the requirements. If all the requirements are met, then the guide will be published. If they are not, the guides are sent back to the creator with a detailed description of what needs to be corrected for publication.

Setting up the peer review process will not only improve metadata and increase the discoverability of LibGuides but will also help meet recently expanded Title II requirements and updated web content accessibility guidelines (W3C, 2024; University of Illinois Urbana Champaign Office of the Provost, 2024). Compliance with these guidelines will influence best practices for creating and revising LibGuides. The TLAS unit will collaborate with library departments to revise and update existing guides. This process will involve addressing issues related to subject and tag usage. These efforts will ensure that future guides comply with all Title II requirements while also improving their discoverability.

Conclusion

While metadata in LibGuides may seem like a minor consideration, it plays a significant role in guide discoverability. By creating and implementing best practices around subject use and tagging, libraries can ensure that users discover relevant resources when searching their websites. Moving forward, the authors intend to establish clear guidelines for metadata practices, including developing a database and integrating it into LibGuide instruction sessions.

Providing this will ensure that LibGuide creators are consistent in their naming practices. Future research will assess the impact of these improvements through user testing.

The authors also found LibGuides data difficult to manage because they are not static. As previously mentioned, the LibGuides platform at the University of Illinois Urbana-Champaign Library is constantly changing, with multiple librarians across 82 departmental accounts adding, deleting, and editing guides at any time. While the data in this article was collected in October 2024, there has since been a focused effort to review existing LibGuides in preparation for a large-scale Title II review for compliance. As a result, about 400 guides have been deleted or merged with others. The authors plan to gather data on LibGuides before starting both the formal work for accessibility compliance and the implementation of best practices for LibGuide creation that are currently under review.

The authors recognize that these findings are specific to their institution. Future research could involve analyzing these variables using data from other universities to see if their results are consistent elsewhere. They have included an Appendix, which they used for data analysis, and encourage others to explore this topic as well.

Roles from CRediT

Stephanie M. Luke: conceptualization, data curation, investigation, methodology, formal analysis, writing - original draft preparation, writing - reviewing and editing. Evie Cordell: conceptualization, data curation, investigation, methodology, formal analysis, writing - original draft preparation, writing - reviewing and editing. Nathan Jeffrey: methodology, formal analysis, writing - reviewing and editing. Kate Meyrick: data curation. Thea Soter: data curation.

References

Almeida, N., & Tidal, J. (2017). Mixed methods not mixed messages: Improving LibGuides with student usability data. *Evidence-Based Library and Information Practice, 12*(4): 62–77. https://doi.org/10.18438/B8CD4T

Alverson, J., Schwartz, J., Brunskill, A., & Lefager, J. (2015). Creating audience and environment-friendly research guides: Findings from a user study. *Association of College and Research Libraries 2015 Conference*, 125–133. http://hdl.handle.net/11213/17841

Bergstrom-Lynch, Y. (2019). LibGuides by design: Using instructional design principles and user-centered studies to develop best practices. *Public Services Quarterly, 15*(3), 205–223. https://doi.org/10.1080/15228959.2019.1632245

Bullard, J. (2018). Curated folksonomies: Three implementations of structure through human judgment. *Knowledge Organization, 45*(8), 643–652. https://doi.org/10.5771/0943-7444-2018-8-643

Burchfield, J., & Possinger, M. (2023). Managing your library's LibGuides: Conducting a usability study to determine student preference for LibGuide design. *Information Technology & Libraries, 42*(4), 1–18. https://doi.org/10.5860/ital.v42i4.16473

Brush, D. (2022). Case study: Does following best practices increase LibGuides usage? *Journal of Web Librarianship, 16*(4), 189–199. https://doi.org/10.1080/19322909.2022.2112798

Castro-Gessner, G., Wilcox, W., & Chandler, A. (2013). Hidden patterns of LibGuides usage: Another facet of usability.

Carey, J., Pathak, A., & Johnson, S. C. (2020). Use, perceptions, and awareness of LibGuides among undergraduate and graduate health professions students. *Evidence Based Library and Information Practice*, *15*(3), 157–172. https://doi.org/10.18438/eblip29653

Conerton, K., & Goldenstein, C. (2017). Making LibGuides work: Student interviews and usability tests. *Internet Reference Services Quarterly, 22*(1), 43–54. https://doi.org/10.1080/10875301.2017.1290002

Conrad, S., & Stevens, C, (2019). 'Am I on the library website?': A LibGuides usability study. *Information Technology and Libraries, 38*(3): 49–81. https://doi.org/10.6017/ital.v38i3.10977

Cox, A. (1996). Hypermedia library guides for academic libraries on the world wide web. *Program, 30*(1), 39–50. https://doi.org/10.1108/eb047216

Dalton, M., & Pan, R. (2014). Snakes or ladders? Evaluating a LibGuides pilot at UCD Library. *The Journal of Academic Librarianship*, *40*(5), 515–520. https://doi.org/10.1016/j.acalib.2014.05.006

Del Bosque, D., & Morris, S. E. (2021). LibGuide standards: Loose regulations and lax enforcement. *The Reference Librarian, 62*(1), 1–22. https://doi.org/10.1080/02763877.2020.1862022

Duncan, V., Lucky, S., & McLean, J. (2015). Implementing LibGuides 2: An academic case study. *Journal of Electronic Resources Librarianship, 27*(4), 248–258. https://doi.org/10.1080/1941126X.2015.1092351

Fagerheim, B., Lundstrom, K., Davis, E., & Cochran, D. (2017). Extending our reach: Automatic integration of course and subject guides. *RUSQ: A Journal of Reference and User Experience*, *56*(3), Article 3. https://doi.org/10.5860/rusq.56n3.180

Fritch, M., & Pitts, J. E. (2016). Adding bite to the bark: Using LibGuides2 migration as impetus to introduce strong content standards. *Journal of Electronic Resources Librarianship, 28*(3), 159–171. https://doi.org/10.1080/1941126X.2016.1200926

GO FAIR (2016). Fair Principles. https://www.go-fair.org/fair-principles

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science, 32*(2), 198–208. https://doi.org/10.1177/0165551506062337

Greene, K. (2020). Accessibility nuts and bolts: A case study of how a health sciences library boosted its LibGuide accessibility score from 60% to 96%. *Serials Review*, *46*(2), 125–136. https://doi.org/10.1080/00987913.2020.1782628

Griffin, M., & Taylor, T. I. (2018). Employing analytics to guide a data-driven review of LibGuides. *Journal of Web Librarianship, 12*(3), 147–159. https://doi.org/10.1080/19322909.2018.1487191

Hennesy, C., & Adams, A. L. (2021). Measuring actual practices: A computational analysis of LibGuides in academic libraries. *Journal of Web Librarianship, 15*(4): 219–242. https://doi.org/10.1080/19322909.2021.1964014

Hopper, T. L. (2021). Accessibility and LibGuides in academic libraries. *The Southeastern Librarian*, *68*(4). https://doi.org/10.62915/0038-3686.1904

Kamela, M., & Higgins, A. (2024) Tips and trends: Optimize your guides: Creating and maintaining user-friendly LibGuides. ACRL's Instruction Section Website. https://acrl.ala.org/IS/tips-and-trends-optimize-your-guides-creating-and-maintaining-user-friendly-libguides

Logan, J., & Spence, M. (2021). Content strategy in LibGuides: An exploratory study. *The Journal of Academic Librarianship, 47*(1), 102282. https://doi.org/10.1016/j.acalib.2020.102282

Lubas R., Jackson A., & Schneider I. (2013). *The metadata manual: A practical handbook*. Chandos Publishing.

Luke, S. M., Han, M.-J., & Stratton, T. (2024). Changes in digital collections and their metadata: A longitudinal study of UIUC digital library. *Journal of Library Metadata, 24*(3), 147–164. https://doi.org/10.1080/19386389.2024.2338015

Lund, W. B., & Hansen, C. (2011). You can lead them to water, but you can't make them drink: Using crowd sourcing to lead library patrons to extended library services relevant to their search criteria. *ACRL Conference Proceedings*, 504–512.

Murty, S. G. K., & Biswas, R. N. (2004). A fuzzy logic based search technique for digital libraries. *DESIDOC Bulletin of Information Technology, 24*(6), 3–10. https://doi.org/10.14429/dbit.24.6.3640

Nielsen, J. (2001). *Search: Visible and simple*. Jakob Nielsen's Alertbox.

Neuhaus, C., Cox, A., Gruber, A. M., Kelly, J., Koh, H., Bowling, C., & Bunz, G. (2021). Ubiquitous LibGuides: Variations in presence, production, application, and convention. *Journal of Web Librarianship*, *15*(3), 107–127. https://doi.org/10.1080/19322909.2021.1946457

Park, J. R. (2009). Metadata quality in digital repositories: A survey of the current state of the art. *Cataloging & Classification Quarterly, 47*(3–4), 213–228. https://doi.org/10.1080/01639370902737240

Park, J. R., Brenza, A., & Lu, C. (2015). A comparative analysis of metadata best practices and guidelines: Issues and implications. *International Journal of Metadata, Semantics and Ontologies*, *10*(4), 240–260. https://doi.org/10.1504/IJMSO.2015.074751

Pittsley, K. A., & Memmott, S. (2012). Improving independent student navigation of complex educational web sites: An analysis of two navigation design changes in LibGuides. *Information Technology and Libraries*, *31*(3), Article 3. https://doi.org/10.6017/ital.v31i3.1880

Pratt, J. A., Mills, R. J., & Kim, Y. (2004). The effects of navigational orientation and user experience on user task efficiency and frustration levels. *The Journal of Computer Information Systems*, *44*(4), 93–100.

Ray, T. (2025). Usability, visibility, and style: LibGuides usability testing to support students' needs. *The Southeastern Librarian, 72*(4). https://doi.org/10.62915/0038-3686.2108

Reeb, B., & Gibbons, S. (2004). Students, librarians, and subject guides: Improving a poor rate of return. *portal: Libraries and the Academy*, *4*(1), 123–130.

Riley, J. (2017). *Understanding metadata*. Washington DC, United States: National Information Standards Organization. http://www.niso.org/publications/press/UnderstandingMetadata.pdf

Sittler, R. L., & Dobbs, A. W. (2016) *Innovative LibGuides applications: Real world examples*. Rowman & Littlefield.

Somerville, M. M., Schader, B. J., & Sack, J. (2012) Improving the discoverability of scholarly content in the twenty-first century: Collaboration opportunities for librarians, publishers, and vendors, A SAGE White Paper. SAGE Publications Ltd.

Sonsteby, A., & DeJonghe, J. (2013). Usability testing, user-centered design, and LibGuides subject guides: A case study. *Journal of Web Librarianship, 7*(1), 83–94. https://doi.org/10.1080/19322909.2013.747366

Staley, S. M. (2007). Academic subject guides: A case study of use at San José State University. *College & Research Libraries, 68*(2), Article 2. https://doi.org/10.5860/crl.68.2.119

Thorngate, A., & Hoden, A. (2017) Exploratory usability testing of user interface options in LibGuides 2. *College & Research Libraries, 78*(6). https://doi.org/10.5860/crl.78.6.844

Trant, J. (2009). Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information, 10*(1), 1–44.

University of Illinois Urbana Champaign Office of the Provost. (2024). Digital Accessibility and Excellence Steering Committee. https://provost.illinois.edu/about/committees/education-2/digital-accessibility-and-excellence-advisory-committee

Vileno, L. (2007). From paper to electronic, the evolution of pathfinders: A review of the literature. *Reference Services Review, 35*(3), 434–451. https://doi.org/10.1108/00907320710774300

W3C. (2024). Web content accessibility guidelines (WCAG) 2.2. https://www.w3.org/TR/2024/REC-WCAG22-20241212

Zeng, M. L., & Qin, J. (2020). *Metadata*. American Library Association.

Appendix

R code

```
# Load R libraries
library(ggplot2)
library(dplyr)
library(stringr)
# Load the data ————————-
libguides <- readxl::read_excel(
 'Libguide Metadata Spreadsheet.xlsx'
)
tags <- readr::read_tsv(
 "Libguides Tags Redundancies.tsv",
)

# Data cleaning ————————
## for Type, compare 'Subject' and 'Course' to all other types
libguides$TypeSubject <- libguides$Type == 'Subject'
libguides$TypeCourse <- libguides$Type == 'Course'
## for `tags` data: filter out Orphan, create 'Low' and 'High' quality tag categories
tagsdata <- tags |>
 filter(Orphan == FALSE)
tagsdata$Quality <- if_else(
 tagsdata$Characters | tagsdata$Misspellings,
 'Low',
 'High',
 missing = NA
)
## remove whitespace
tagsdata$Name <- str_squish(tagsdata$Name)
## use tags data to classify the 'Tag Names' column in libguides data
## into 'High Quality' and 'Low Quality' tags
## initialize new variables
libguides$LowQualityTags <- 0
libguides$HighQualityTags <- 0
## Split the 'Tag Names' field into separate entries and remove whitespace
libguides$TagNameList <- str_split(libguides$`Tag Names`, ', ')
libguides$TagNameList <- purrr::map(libguides$TagNameList, ~purrr::map_chr(.x,
str_squish))
## Loop through the 'tagsdata' table.
## (note: this loop takes ~1 min to run)
for (tag in 1:nrow(tagsdata)) {
 # For each tag in `tagsdata`, count the entries in `Tag Names` that match
 matches <- purrr::map_vec(
  libguides$TagNameList,
  ~sum(
   as.numeric(
    str_equal(.x, tagsdata$Name[tag])
   )
  )
 )
```

```
 # set NA's to 0 for easier addition
 matches[is.na(matches)] <- 0
 # Add the matches to LQ tags
 if (tagsdata$Quality[tag] == "Low") {
  libguides$LowQualityTags <- libguides$LowQualityTags + matches
 }
}
# all other tags are defined as high quality
libguides$HighQualityTags = libguides$Tags - libguides$LowQualityTags

# Exploring the data ———————-
## Descriptive stats
psych::describe(libguides$Usage, fast = TRUE)
psych::describe(libguides$Subjects, fast = TRUE)
psych::describe(libguides$HighQualityTags, fast = TRUE)
psych::describe(libguides$LowQualityTags, fast = TRUE)
table(libguides$Type)
## Bar chart for Type ——
libguides |>
 count(Type) |>
 # Calculate percent
 mutate(Percent = round(n / sum(n) * 100, 2)) |>
 # Bar chart
ggplot(
 aes(x = Percent, y = Type, fill = Type)
) +
 geom_col() +
 geom_text(
  aes(x = Percent / 2 + 2, label = paste0(Percent, '%')
  )
 ) +
 theme_minimal() +
 scale_x_continuous(labels = ~scales::percent(.x, scale = 1)) +
 scale_color_discrete() +
 labs(x = "", y = "") +
 theme(
  legend.position = "none",
  panel.grid.minor = element_blank(),
  panel.grid.major.y = element_blank(),
 )
ggsave(
 "types-of-libguides.png",
 width = 6,
 height = 4,
 units = "in"
)

# Regression ———————-
## Linear regression
model1 <- lm(Usage ~ TypeSubject + TypeCourse + Subjects + Tags, data = libguides)
summary(model1)
### Diagnostic plots
```

```
png(
 'OLS-diagnostic-plots.png',
 height = 6,
 width = 6,
 units = "in",
 res = 320
)
par(mfrow = c(2, 2))
plot(model1, 1) # Residuals vs fitted: shows outliers
plot(model1, 2) # Q-Q plot: shows non-normality
plot(model1, 4) # Cook's distance: shows influential observations
dev.off()
### look at the outliers
libguides[c(5,166), ] |> View()
## Robust regression
library(robustbase)
model2 <- lmrob(Usage ~ TypeSubject + TypeCourse + Subjects + Tags,
 data = libguides)
summary(model2)

# Chi-square test for model fit
model2_null <- lmrob(Usage ~ 1, data = libguides)
anova(model2, model2_null, test = "Deviance")
```