FSU Libraries Spring 2022 Research Data Services Report

I: Introduction

About Research Data Services @ FSU Libraries

Data science skills have become increasingly vital for college graduates for the past decade. At the same time, there has been a notable shortage of college graduates with these skills. Two articles in the Harvard Business Review have outlined the shortage of people in the field: one in 2012, with a follow-up article in 2022. (Davenport and Patil 2012) (Davenport and Patil 2022). Around these same times, two executive memorandums from the Office of Science and Technology Policy in 2013 and 2022 opened publicly funded research, thus expanding the amount of publicly available data for analysis. (Holdren 2013) (Nelson 2022)

Even further, data analysis skills are needed beyond academic and government research. In collaboration with IBM and Burning Glass Institute, The Business Higher Education Forum released a report in 2017 titled “The Quant Crunch” that highlights the need for certain data science skills *outside* of the title of data scientist. Per their report, the demand for employees to have data science skills increased by 40%, as did clinical data analysis [54%], data visualization skills [31%], A/B Testing [22%], and Machine Learning [17%], to name a few. Further, that report has extensive lists of occupations that need data science skills: from Epidemiologist to UI/UX Designers, the range of potential uses is no longer limited to statisticians and computer scientists alone.

In response to academic, government, and private sector demand for people with these skills, FSU Libraries has developed a range of new and established research data services since 2014. Among what has been established to date,

- Research data management consultations with 1:1 training
- Training in data analysis and visualization via workshops
  - Data @ Your Desk series (past workshops have included Python, MATLAB, R, QGIS, Jupyter Notebooks, SPSS, NVivo, etc.)
- Digital humanities data analysis and interpretation
- Preservation and access to data collections
- Academic health data management, infrastructure, and software training
- Advocacy for open data, open science, and data literacy

In addition to the above services, FSU Libraries established the Data Fellowship Program within STEM Libraries in 2021. This initiative's primary aim was to develop a peer-to-peer teaching, research support, and outreach model where student data fellows can contribute to research data services at FSU Libraries. Furthermore, the data fellowship program also provides opportunities for independent data science related projects based on a given data fellow’s previous research and/or technical skill, which is helpful for their own professional development. This said, one previous data fellow had a previous background in survey research

and R programming, giving FSU Libraries tools to follow up with student engagement around the 2021-2022 academic year.

Why We Did This Survey

While research data services have been available at FSU Libraries for several years, the question of tangible student engagement arises. First and foremost, FSU Libraries did not previously have data on undergraduate student engagement with research data services. While there is data for how often and how long students visit an FSU Libraries building, that data does not provide insights into what a student proceeds to do once they visit. Further, remote engagement is harder to account for; with the existence of remote/asynchronous education, there are now multiple modes of student engagement.

Additionally, undergraduates may be unaware of everything an academic library can provide. While undergraduate students may visit libraries on a regular basis, their engagement with the library's research data services might be affected by their awareness of what research data services exist. Moreover, knowing what services students are aware of is useful for knowing what outreach efforts are working and identifying knowledge gaps students may have for these services. In addition to marketing strategies implemented by FSU Libraries in the past, one aspect of the data fellowship program involves outreach. Thus, having insight into what students use and have knowledge of allows us to ascertain what engagement methods are/are not working, and develop additional strategies to boost student usage of these services.

II: Methodology

Survey Design

We designed an eight-question survey to address core pieces of information that we would need to glean insight into student engagement with FSU Libraries' research data services. These questions were related to (but not limited to) major types, class level, whether students had heard of and/or used research data services, and student perceptions on the importance of data literacy, data analysis, and data visualization for their academics and careers. The final question was an open-ended text response question where students could input ideas and suggestions for research data services at FSU Libraries.

Two versions of this survey were distributed. The first version was a beta version that was distributed to a targeted sample of student workers and OPS staff within FSU Libraries. Conducting a beta test of the survey allowed us to check for survey flow errors, double-check for unclear language (e.g. what data services are), and make sure that the average survey respondent could be finished within 2 to 5 minutes. Insights gleaned from beta testing were incorporated into a second version of the survey, which was disseminated to the general student population. However, it is important to note that the general content and structure of the survey remained consistent between both versions.

The final survey was disseminated via QR code flyers that lead to a Qualtrics survey. Both versions of our survey were anonymous unless a respondent elected to enter an incentive raffle pool, with survey confidentiality maintained for all respondents. Our final survey sample consisted of n=56 total survey respondents who opened the survey, and n=47 completed survey responses.

Survey Analysis

As a first step, we looked at the distribution, mean, and median of each individual question to identify where the majority of responses fell. Then, we performed One-Way and Two-Way ANOVA tests in order to discern whether years in school and/or major type would impact a respondent's overall engagement with library research data services, as well as their general sentiments for whether or not data literacy, data analysis, and data visualization were important to their chosen career path.

Logistic regression analyses were performed to determine the likelihood of a respondent using a research data service in comparison to their answers to the three statements in question 7, as well as their academic standing. Finally, we ran tests for multicollinearity and autocorrelation to check for the usefulness of this data and the potential models (i.e; making sure that variables were not correlated to each other, and that our models did not show redundancy.)

III: Results

Basic Demographic Breakdown of Sample

For question one, we found that the median respondent was a third-year student, with a mean of 2.82 years. Thus, we found that our sample skewed slightly towards undergraduate students in the middle of their academic careers, as opposed to earlier undergraduate students *or* graduate students and students past five years. We also opted to collect opinions from the 10 graduate students in the sample instead of screening those participants out: while the combined total of 5+ year students and graduate students was n=10, they were not the majority of the sample.

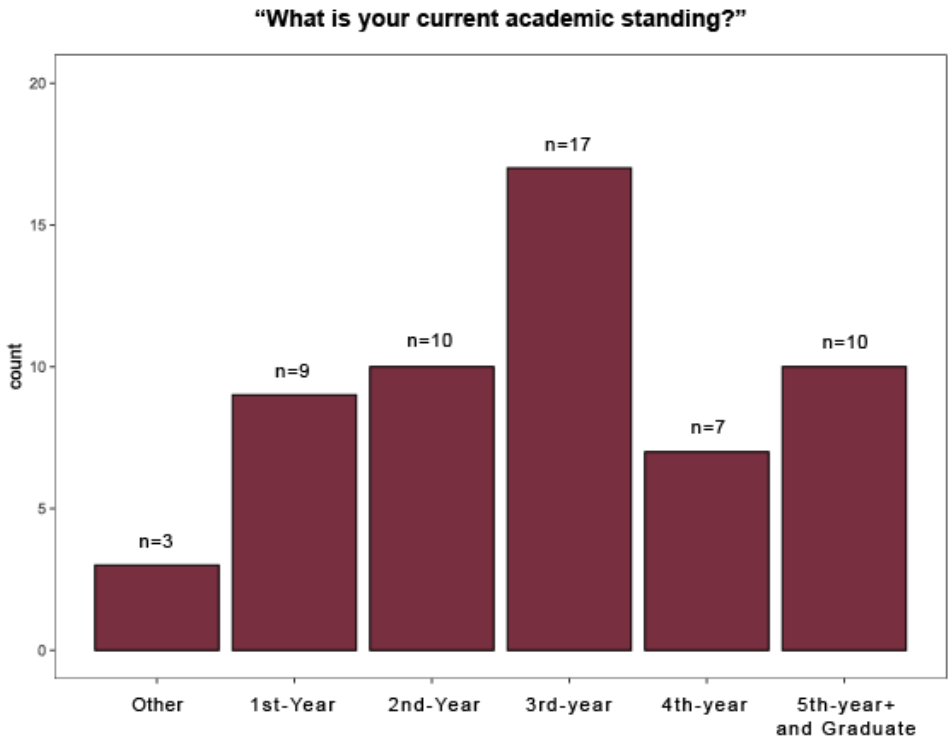


Figure 1: Distribution of survey respondents by academic level

Furthermore, there was a generally even distribution of students across the macro-disciplinary areas of Arts and Humanities, Social Sciences, Business, and STEM (Science, Technology, Engineering, Mathematics). Notably, when responses are categorized into three categories – Arts and Humanities, Social Sciences and/or Business, and STEM, we see that the majority of respondents fell within the second category. However, further ANOVA (Analysis of Variance) that separated Social Sciences from Business did not yield any significantly different results than ANOVA that grouped those two major types into the same category.

Additionally, there were significantly fewer respondents from Arts and Humanities related disciplines than all other fields. One potential explanation for this outcome may be that Arts and Humanities related majors do not traditionally perceive data as relevant to their fields. Thus, it would follow that respondents may have self-screened and decided against taking our survey under the presumption that their opinion and/or the survey itself would not be relevant to them.

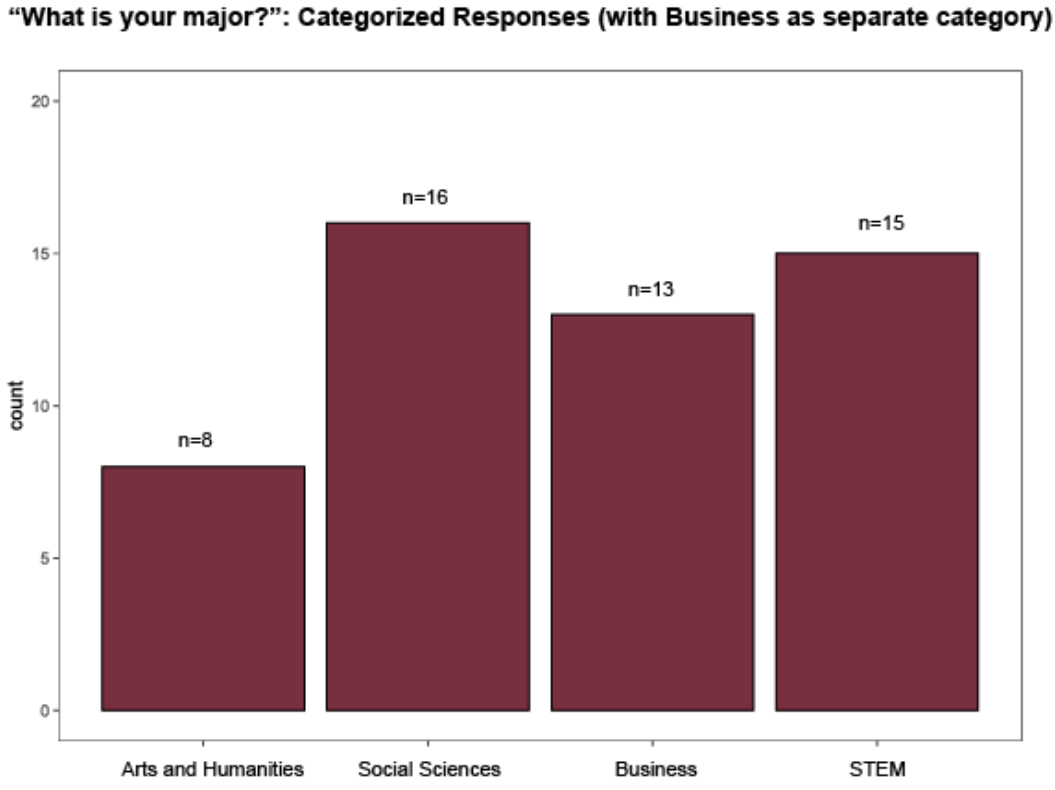


Figure 2: Distribution of major types by category when split into four groups

Discrepancies in Awareness versus Utilization of Research Data Services

For question three, we can note that most respondents (n=43, comprising 76.78% of the entire sample) had heard of at least one type of research data service provided by FSU Libraries. However, the types of data services respondents were aware of were not uniform. Notably, in most cases, more respondents had not heard about a specific service than those who had. The only exception to this trend was workshops and training opportunities, wherein many respondents had heard of this research data service. Interestingly, both research data management and statistical repositories/databases were tied for the least number of respondents who had heard about those specific research data services before.

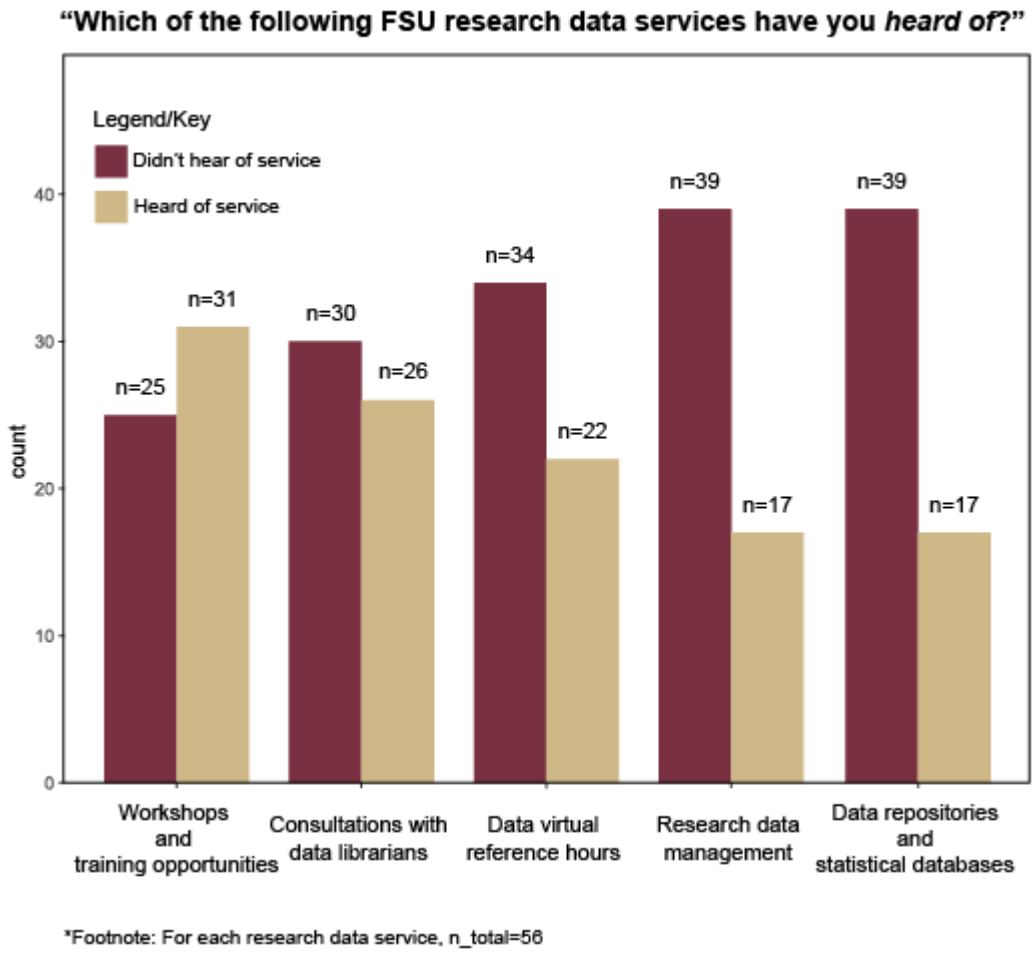


Figure 3: Grouped bar graph of which services respondents heard about and did not hear about

For question four, we found that someone hearing about one or more research data services did not often translate to using those services (with only n=11 for 19.64% of the sample using a research data service.) The results on specific research data services from Question 5 show that the trend of specific services being used also followed the general trend of being heard about far more often than they were used. However, the specific data services that were used most often were not necessarily the research data services that were most frequently heard about. Although workshops were the most frequently heard about data service with a total of 31 respondents hearing about it, only four respondents proceeded to use this data service. Conversely, the most used service was data repositories and statistical databases, despite being one of the two least heard-about research data services among the entire sample.

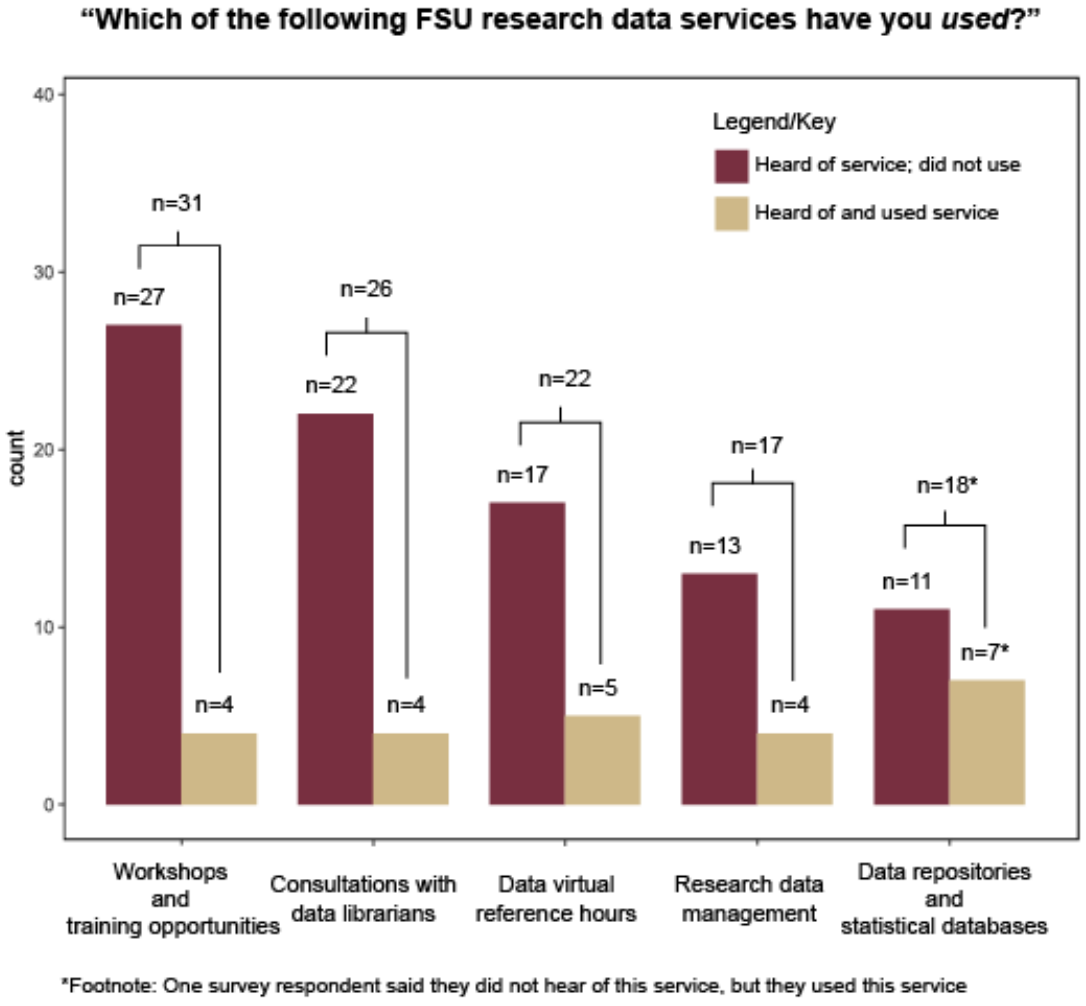


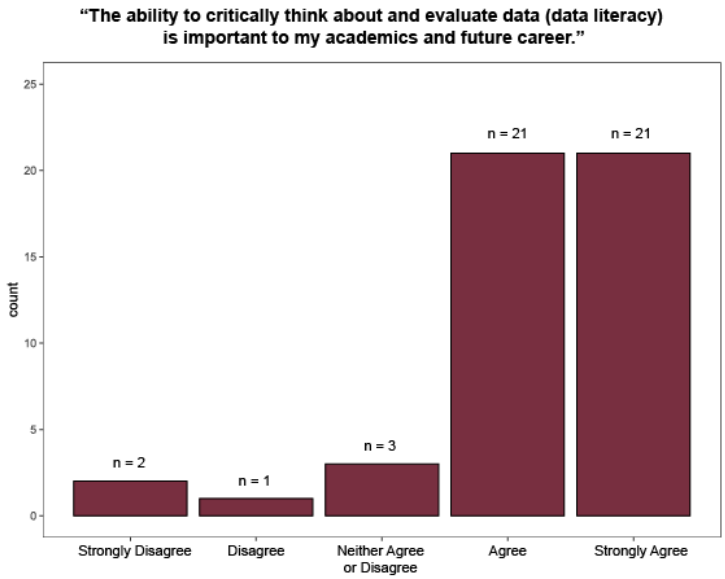
Figure 4: Grouped bar graph of which services respondents used and did not use

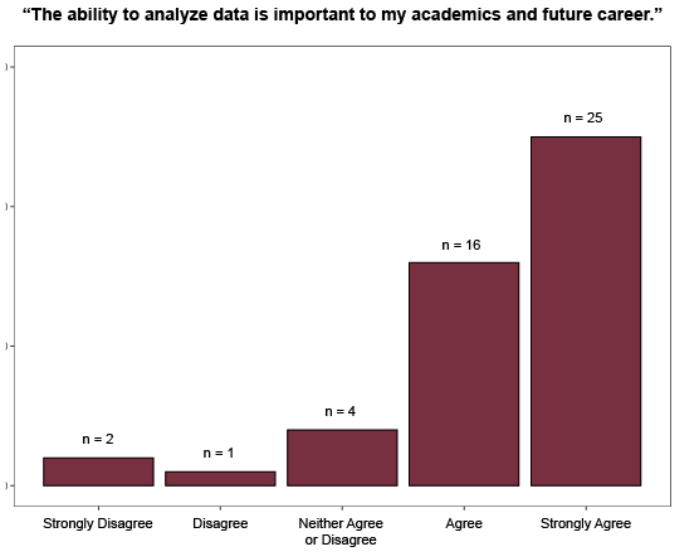
When we started analyzing the data, we had the hypothesis that major type and/or academic level could influence whether a respondent was more likely to use a research data service. The reason for this consideration was twofold. Firstly, data-related skills are traditionally seen as the purview of STEM fields. Additionally, traditional undergraduate students tend to focus more on developing career-related skills and/or graduate school readiness as they get closer to graduating from undergraduate programs.

In the context of whether major type or academic standing would impact whether someone would use a research data service, we were only able to find one significant difference with a specific demographic. First-year respondents were the least likely to use a research data service. It can be noted that this was shown both by the data itself (i.e., no first-year respondents reported using a research data service), as well as with analytic measures (i.e., first-year students were statistically different from all other academic level groups with 75% confidence, even if the other academic level groups were not significantly different from each other in terms of research data service use.) As first-year students were not significantly different from other academic level groups in the context of hearing about research data services, there may be an assumption that they do not need to use these services until later in their college careers (if at all).

Respondent Sentiments Toward Data Literacy, Data Analysis, and Data Visualization

In general, we found that survey respondents tended to answer with “Strongly Agree” (5) to sentiment questions related to data literacy and data analysis and “Agree” (4) for data visualization, with only slight variation/difference between majors. Some of this skew toward “Strongly Agree” and “Agree” answers may be explained by the previously mentioned lack of Arts and Humanities students who participated in this research– although we did not find a statistically significant difference in any major types and response patterns, we also have fewer Arts and Humanities majors in our sample. Meanwhile, the reason for less unanimous agreement on data visualization has less obvious reasons for happening. One possible reason could be that the concept of data visualization was a foreign concept to some survey respondents. However, further research would be required in order to not extrapolate from data we do not have (i.e., we did not ask if respondents previously heard about data visualization before.)





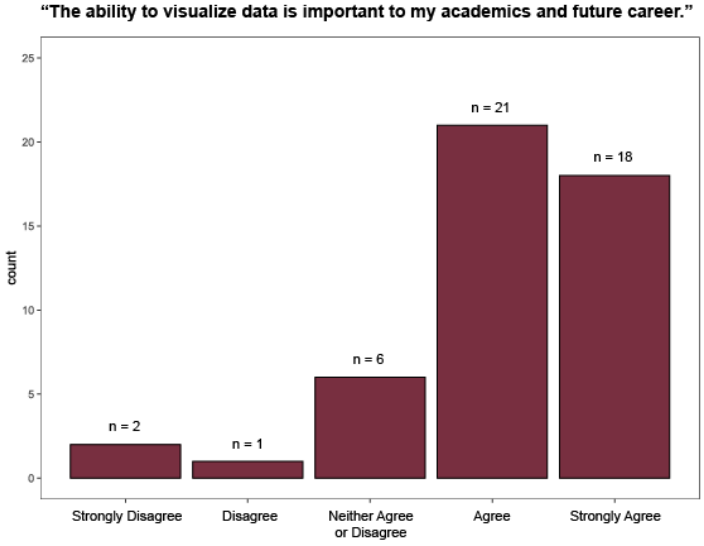


Figure 5.1: Bar graph distribution of respondent answers to Question 7 Statement 1

Figure 5.2: Bar graph distribution of respondent answers to Question 7 Statement 2

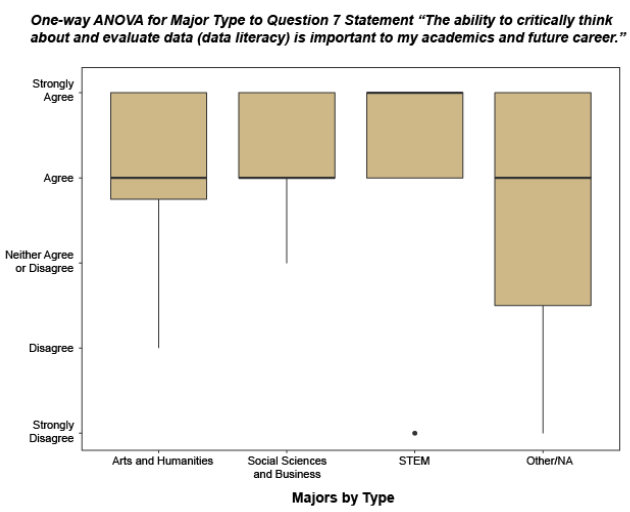
Figure 5.3: Bar graph distribution of respondent answers to Question 7 Statement 3

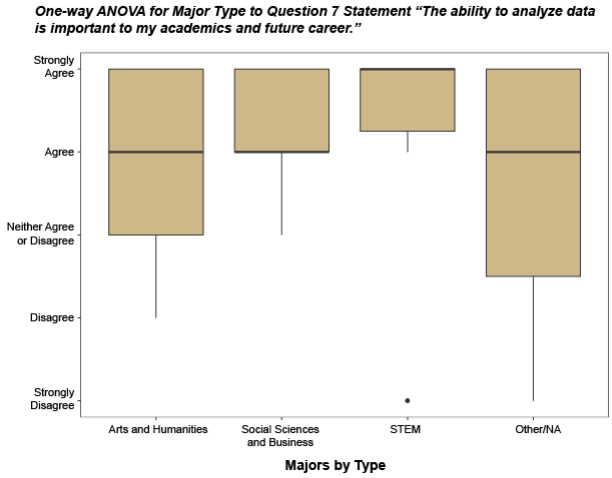
When adding further ANOVA to the three statements to compare major types to our Question 7 statements and the responses we got, we did find some slight differences, even if they did not appear to have statistical significance in the traditional sense.

For the first and second statements (“The ability to critically think about and evaluate data (data literacy) is important to my academics and future career”, and “The ability to analyze data is important to my academics and future career”), we found that all other fields outside of STEM tended to answer “Agree” (4), while STEM majors tended to answer “strongly agree” (5). There was a wider distribution of potential answers between Arts and Humanities majors than for the Social Sciences and Business majors, while our Other/NA category had the widest variety of responses. It could be presumed that STEM majors, Social Science majors, and Business-related majors may traditionally interpret themselves as needing to understand and analyze data more often than Arts and Humanities majors.

Our results for the third statement (“The ability to visualize data is important to my academics and future career”) were slightly different, as we found that all major types unanimously tended to choose “agree” (4). There was a wider distribution of potential answers between Social Sciences and Business majors and STEM majors in comparison to the Arts and Humanities majors with a narrower distribution, while our Other/NA category had the widest distribution of responses. One hypothesis for the distribution for Arts and Humanities majors being narrower in spite of their answer distributions being wider for the other two statements may have to do with the vague definition of data visualization. However, it should be reiterated that we cannot extrapolate what the respondent interpreted data visualization to mean.

We found less consistent patterns in how respondents would answer questions based on their academic standing, with answers between “Strongly Agree” (5) and “Agree” (4), with no obvious pattern between academic standing and the average response provided for each level, as well as no clear reason why the different years would answer they way they did on average.





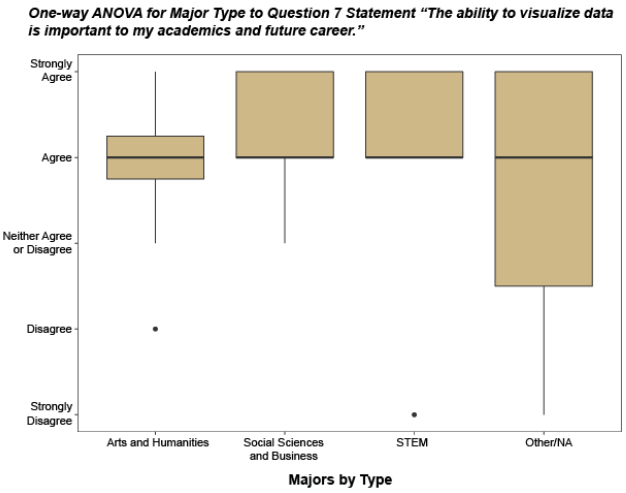
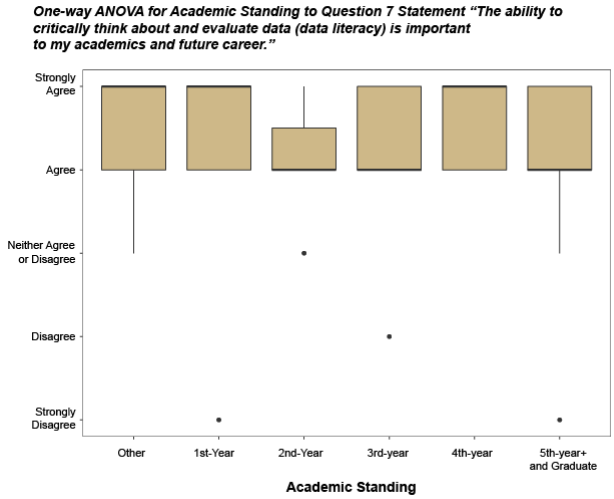
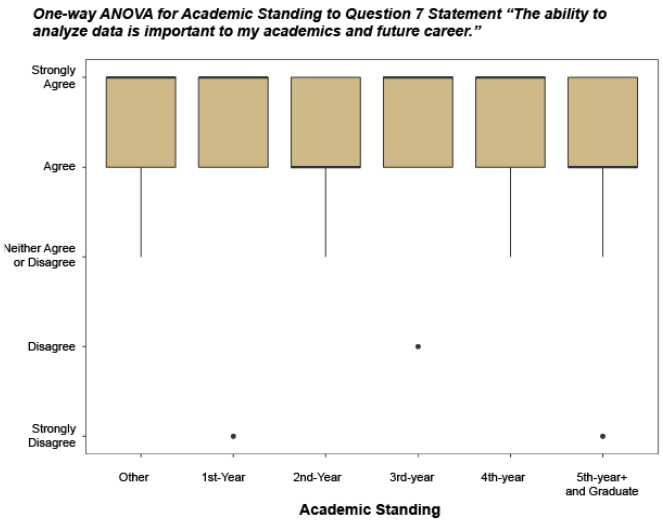


Figure 6.1: One-way ANOVA graph for Question 7.1 and major type

Figure 6.2: One-way ANOVA graph for Question 7.2 and major type

Figure 6.3: One-way ANOVA graph for Question 7.3 and major type





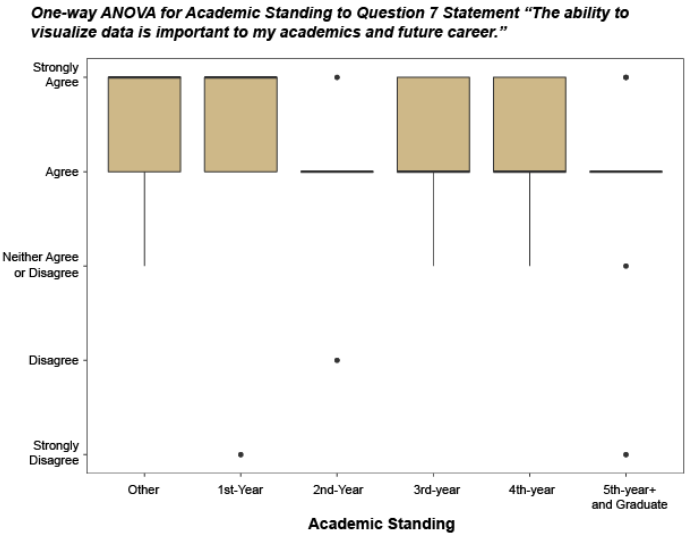


Figure 7.1: One-way ANOVA graph for Question 7.1 and academic level

Figure 7.2: One-way ANOVA graph for Question 7.2 and academic level

Figure 7.3: One-way ANOVA graph for Question 7.3 and academic level

We did some analysis regarding response patterns for people who did hear about at least one research data service, versus people who used no research data services. This generally would measure whether *awareness* about the existence of research data services available at FSU Libraries

For the first and third statements ("The ability to critically think about and evaluate data (data literacy) is important to my academics and future career" and "The ability to visualize data is important to my academics and future career"), we found that both groups generally centered around "agree" (4) as their response. However, the distribution was narrower for students who heard about at least one data service, as they tended to answer "agree" more unanimously than the respondents who did not previously hear of a research data service. Moreover, for the second statement ("The ability to analyze data is important to my academics and future career"), we found that respondents who did not previously hear of at least one research data service would typically answer "agree" (4) as their response, while respondents who heard about *at least one* research data service would typically respond with "strongly agree" (5). For the third statement ("The ability to visualize data is important to my academics and future career"), we found that both groups generally centered around "agree" (4) as their response. All three statements and the groups of respondents who heard about or did not hear about any research data services prior to taking our survey were statistically significant with at least 95% confidence. We can hereby conclude that marketing efforts for research data services have a statistically significant impact on a respondent's opinions on data literacy, analysis, and visualization, even if they do not proceed to use any research data services.

Furthermore, we found that with the raw ANOVA table output from R, using at least one data service was significantly different for how much respondents would agree or disagree with each statement from Question 7, where we found that there was a statistically significant difference at the 90% confidence level for statement 2 ("*The ability to analyze data is important to my academics and future career*".) However, neither statement 1 ("The ability to critically think about and evaluate data (data literacy) is important to my academics and future career") nor statement 3 ("The ability to visualize data is important to my academics and future career") showed any statistically significant difference with respect to whether or not someone would proceed to use a service. Thus, it appears that respondents who would proceed to use at least one research data service would only be inclined to do so if they felt that data analytics alone was related to their academics and future career.

Using two-way ANOVA to find whether or not there were interaction effects for the number of years in college *and* the type of major a respondent, we found that there was no interaction effect between the two. This held regardless of whether or not standard scientific 90%, 95%, and 99% confidence intervals were used, or when using 75% or 80% confidence intervals standard in business applications. This makes sense, considering that neither major type nor academic standing independently impacted the likelihood of using a data service or what they think about data-related concepts.

We were also able to find the likelihood that someone would proceed to use a service based on how they would respond to Question 7's second statement ("The ability to analyze

data is important to my academics and future career"). We found that students were less than 1% likely to use a research data service if they answered "strongly disagree" or "disagree", and only about 3% likely to use a research data service if they answered "neither agree or disagree", 11% likely to use a service if they answered "agree", and 33% likely to use a service if they answered "strongly agree" to the second statement.

IV: Future Steps

From the above analysis, a few potential courses of action could be made.

Firstly, the lack of engagement from first-year students is clear enough to make targeted efforts to that demographic a possible goal. As we had no first-year participants who used a research data service despite some respondents previously hearing about these services, this population does not appear to perceive these services as applicable to their academic careers early on. This may call for early academic career outreach so that earlier students might use services earlier on.

Additionally, data analysis skills are the main driver for students to use research data services. In contrast, data literacy and data visualization have little to no impact on whether or not someone will proceed to use a research data service. Therefore, there are two distinct directions that FSU Libraries can go with service planning. One potential direction would be to expand upon data literacy and data visualization advocacy/awareness among the general student population to address this gap. The other potential direction that can be taken is focusing more on the student population that is incentivized to learn data analysis, as they have the highest propensity to use at least one research data service. With this population, expanding services related to data analysis might be beneficial. Some potential options may include a broader selection of data analysis workshops, expanded times, etc.

V: Appendix

Survey Design

In order to gather data for this research endeavor, we implemented two rounds of a 1-2 minute survey with eight questions.

For our beta version of the survey that was tested with OPS staff within FSU Libraries, the questions that were asked were as follows:

1. "What is your current academic standing?"
    - First-year
    - Second-year
    - Third-year
    - Fourth-year
    - Post-baccalaureate
    - Graduate

    - Other
2. What is your major?
    - Open response text box
3. Which of the following FSU research data services have you heard of? (Research Data Services are defined as "library resources to help students and faculty manage, find, use, analyze, store, and learn about data.")
    - Workshops and training opportunities
    - Consultations with data librarians
    - Data reference hours
    - Research data management
    - Data repositories and statistical databases
    - None of the above
4. Have you used any data services from the FSU Libraries system this semester?
    - Yes
    - No
    - I don't know
5. [If "No" was selected for 4, this question was skipped] Which of the following data services have you used?
    - Workshops and training opportunities
    - Consultations with data librarians
    - Data reference hours
    - Research data management
    - Data repositories and statistical databases
    - None of the above
6. [If "No" was selected for 4, this question was skipped] On average, how often do you use any data services from the FSU Libraries system?
    - More than once a week
    - At least once a week
    - More than once a month
    - At least once a month
    - Less than once a month
7. On a scale of "Strongly Agree" to "Strongly Disagree" how much would you agree with each of the following statements?
    - "Data literacy is important to my career"?
        i. Strongly Disagree
        ii. Disagree
        iii. Neither Agree Nor Disagree
        iv. Agree
        v. Strongly Agree
    - "Data analysis is important to my career"?
        i. Strongly Disagree
        ii. Disagree
        iii. Neither Agree Nor Disagree
        iv. Agree

        v. Strongly Agree
    - “Data visualization is important to my career”?
        i. Strongly Disagree
        ii. Disagree
        iii. Neither Agree Nor Disagree
        iv. Agree
        v. Strongly Agree
8. What additional data services from FSU Libraries would you be interested in seeing?
    - Open response text box

Using additional information given to us from respondents for the beta survey on how to improve the final survey, we were also able to implement a final version of the survey with slightly modified questions and answers.

1. What is your current academic standing?
    - First-year
    - Second-year
    - Third-year
    - Fourth-year
    - Fifth-year or more
    - Post-baccalaureate
    - Graduate
    - Other
2. What is your major?
    - Open-response text box
3. Which of the following FSU research data services have you ***heard of***? Please check all that apply. (Research Data Services are defined as "library resources to help students and faculty manage, find, use, analyze, store, and learn about data".)
    - Workshops and training opportunities (e.g. Intro to Python Bootcamp, R, MATLAB, SPSS, QGIS, etc)
    - Consultations with data librarians (e.g. scheduling an appointment with a subject librarian to ask about specialized questions)
    - Data reference hours (i.e. using the “Ask a Data Librarian” feature on the FSU Libraries website to chat with a data librarian)
    - Research data management (i.e. consulting with research librarians to develop a research data preservation plan for your research)
    - Data repositories and statistical databases (e.g; using DigiNole to store your own research data, using the FSU Libraries website to find data)
    - None of the above
4. Have you used any FSU research data services this semester?
    - Yes
    - No
    - I don’t know
5. Which of the following FSU research data services have you ***used***? Please check all that apply.

    - Workshops and training opportunities (e.g. Intro to Python Bootcamp, R, MATLAB, SPSS, QGIS, etc)
    - Consultations with data librarians (e.g. scheduling an appointment with a subject librarian to ask about specialized questions)
    - Data reference hours (i.e. using the "Ask a Data Librarian" feature on the FSU Libraries website to chat with a data librarian)
    - Research data management (i.e. consulting with research librarians to develop a research data preservation plan for your research)
    - Data repositories and statistical databases (e.g; using DigiNole to store your own research data, using the FSU Libraries website to find data)
    - None of the above
6. On average, how often do you use any FSU research data services?
    - More than once a week
    - At least once a week
    - More than once a month
    - At least once a month
    - Less than once a month
7. On a scale of "Strongly Agree" to "Strongly Disagree" how much would you agree with each of the following statements?
    - "The ability to critically think about and evaluate data (data literacy) is important to my academics and future career."
        i. Strongly Disagree
        ii. Disagree
        iii. Neither Agree Nor Disagree
        iv. Agree
        v. Strongly Agree
    - "The ability to analyze data is important to my academics and future career."
        i. Strongly Disagree
        ii. Disagree
        iii. Neither Agree Nor Disagree
        iv. Agree
        v. Strongly Agree
    - "The ability to visualize data is important to my academics and future career."
        i. Strongly Disagree
        ii. Disagree
        iii. Neither Agree Nor Disagree
        iv. Agree
        v. Strongly Agree
8. What additional research data services from FSU Libraries would you be interested in seeing?
    - Open-response text box

The survey was implemented in Qualtrics and distributed via a QR code on fliers around the FSU campus that respondents could scan, with the fliers having an incentive for potential

rewards written on the fliers. The QR code would lead to the link for the Qualtrics survey and had a disclosure statement for anonymity and the purpose of the survey before a respondent would start the survey.

Variable Coding

Question 1: Description of academic level. Survey respondents were numerically coded as 1 = First Year, 2 = Second Year, 3 = Third Year, 4 = Fourth Year, with 5 = Five+ Year, Graduate students, and Post-Baccalaureate students in one group. 0 = Other.

Question 2: Description of respondent's major. The first version of this variable was coded 0 = Arts and Humanities, 1 = Social Sciences and/or Business, and 2 = STEM, whereas the second version of this variable was coded as 0 = Arts and Humanities, 1 = Social Sciences, 2 = Business, and 3 = STEM.

Question 3: Description of whether a survey respondent heard of a research data service. This question yielded six different variables. The first five were for each individual research data service, where 0=did not hear of that specific service and 1 = heard of that specific service. There was also a variable added for hearing of *at least* one research data service, where 0=did not hear of *any* services and 1 = heard of *at least one* service.

Question 4: Description of whether someone proceeded to use a research data service. The question was coded as a binary variable, where No = 0 = did not use any research data services and Yes = 1 = used *at least one* research data service.

Question 5: Description of whether a survey respondent used research data service. This question yielded six different variables. The first five variables were for each individual research data service, where 0 = did not use that specific service, and 1 = used that specific service.

Question 6: This survey question measured how often a respondent who used services would proceed to use a service.

Question 7: This survey question yielded three separate variable groups for the three statements in the question. For each individual statement, there was an associated value where 1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, and 5 = Strongly Agree.

Question 8: This open-response question allowed for further commentary on the survey itself and/or research data services.

Library Staff: The beta test version of our survey included OPS workers, most of whom were also students. Due to this pre-existing data, we could merge the initial and final survey data into one data file. Statistical testing between the first and second survey groups yielded no statistically significant difference in response patterns, so we elected to use the combined dataset.

VI: Bibliography

Business Higher Education Forum, IBM, and Burning Glass Institute. “The Quant Crunch: How the Demand for Data Science Skills Is Disrupting the Job Market | BHEF.” ..Com. bhef.com. Accessed April 25, 2023. https://www.bhef.com/publications/quant-crunch-how-demand-data-science-skills-disrupting-job-market.

Bureau of Labor Statistics. 2022. “Occupational Outlook Handbook: Data Scientists.” ..Gov. Occupational Outlook Handbook: Data Scientists. September 8, 2022. https://www.bls.gov/ooh/math/data-scientists.htm.

Davenport, Thomas H., and D. J. Patil. 2012. “Data Scientist: The Sexiest Job of the 21st Century.” *Harvard Business Review*, October 1, 2012. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century.

Davenport, Thomas H., and D. J. Patil. 2022. “Is Data Scientist Still the Sexiest Job of the 21st Century?” *Harvard Business Review*, July 15, 2022. https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century.

Holdren, John P. “MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES: Increasing Access to the Results of Federally Funded Scientific Research.” obamawhitehouse.archives.gov, February 22, 2013. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Nelson, Alondra. “MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research.” whitehouse.gov, August 25, 2022. https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf.