Identification of Educational Gaps in Data Science Training Across Agricultural Genomics

Gabriella Roby Dodd[1,2], Cedric Gondro[3], Tasia M.Taxis[3], Margaret Young[4], and Breno Fragomeni[1]

[1]Department of Animal Science, University of Connecticut
[2]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph
[3]Department of Animal Science, Michigan State University
[4]Department of Natural Sciences, Elizabeth City State University

Author Note

Correspondence regarding this article should be addressed to Breno Fragomeni, Department of Animal Science, University of Connecticut, Storrs, 06269. Email: breno.fragomeni@uconn.edu.

Acknowledgements
This research was supported by the intramural research program of the U.S. Department of Agriculture, National Institute of Food and Agriculture, [Agricultural Genome to Phenome Initiative, grant no. 2022-70412-38454, 2021-70412-35233, and 2020-70412-32615].

Disclaimer
The findings and conclusions in this preliminary publication have not been formally disseminated by the U. S. Department of Agriculture and Should not be construed to represent any agency determination or policy.

Abstract

The objectives of this study were to identify gaps in educational training for undergraduate and graduate students in agricultural data science, propose paths for filling these gaps, and provide an annotated list of resources currently available to different training levels. Data in this study was collected through three voluntary surveys catered to undergraduate students, graduate students, and faculty or professionals in fields of agricultural data analytics. Resources were identified through search engines and annotated based on cost, target audience, and topic. Undergraduate students were found to be inexperienced in statistics, data analysis, and coding. Graduate students were better trained than undergraduate students but did not find university curriculum to be the primary source of education. Faculty and professionals indicated that interest in their field is high but the number of qualified applicants for positions is low. Additionally, there was interest by faculty and professionals to fund training programs for employees but low access to resources for these programs. Education resources identified through the search were limited and many had high cost to students. All resources identified were published in an online catalog (https://agdata.cahnr.uconn.edu/).

Keywords: undergraduate education, graduate education, education survey, data science

With recent advances in agricultural development, there have been vast improvements in the quality and quantity of agricultural data available. The ease of obtaining genomic data such as microbiome, gene expression, high-density SNP markers and genome sequencing provides the volume, velocity, variety, and veracity of data to be considered “Big Data” (Coble et al., 2018). Developments in data availability have enabled the American agricultural industry to shift from a management-focused approach to data analytical decision making process (Himesh et al., 2018). Improvements in the quality and quantity of data has added value to agricultural products and improved consumer value. However, to maintain this pace of innovation and adapt to the ever-changing industry, it is necessary to train individuals capable of understanding, analyzing, and utilizing this wealth of data.

As the complexity of data and analytical methods increases, the lack of qualified researchers is becoming more apparent. This issue has already been building up for many years (Misztal, 2007) and while industry has made an effort to incorporate data analytics into its operations, college graduates remain untrained and ill prepared to work on “Big Data” analysis and derive meaning from results. In the field of agricultural data analytics there are two major obstacles leading to a scarcity of hirable individuals for the industry: 1) a shortage of individuals entering the field, and 2) a deficit in the training of those in the field.

In the field of quantitative genetics and animal breeding, there has been a vacuum of upcoming talent at the undergraduate and graduate levels (Eisen, 2008). This gap may be attributed to technology changing faster than the undergraduate and graduate curriculum can adapt. This rapidly evolving field results in professors learning new skills and methods alongside their students. To properly cater to modern industry needs it will be important to involve undergraduates in experiential learning opportunities, such as research projects and internships, while providing guidance on course selection and professional options post-graduation (Chong et al., 2022; Gilbert et al., 2014; Lee, 2008). In addition to the lack of interest from the undergraduate population, there is suspected insufficient recruiting and funding of graduate students in the field. To reverse this deficit there is a need for increased funding through research grants and more support from industry partners to provide data and internship opportunities.

This issue is not unique to quantitative genetics and animal breeding; the fields of plant breeding (Shakoor et al., 2019) and environmental science (Hernandez, 2012) are facing similar issues. Students entering the field of biology with a focus on analytics need to understand the broad range of available research fields, learn how to “define, test, and refine” experiments, and have a strong knowledge of coding and data analysis (Argueso et al., 2019). A lack of training in coding and data analysis leads to graduates without the skillset necessary to meet the needs or industry or a graduate program.

The aims of this study were to identify opportunities for curricula improvements in agricultural data analytics. An online catalog of existing resources for data science in agriculture was compiled (https://agdata.cahnr.uconn.edu/). The items in the catalog were enriched with annotations including information on the target audience, topics covered, cost and availability of the resource. While compiling these resources, the current gaps in education and areas that have potential for improvement were identified. In addition to this evaluation of the current resources, a survey with undergraduates, graduates, and professionals was conducted to assess the current status of education in the field. Current educational gaps were then identified and areas for improvement were proposed as a first step towards addressing the deficit of qualified human resources in agricultural data analytics.

Materials and Methods

This study and data collection protocols were approved by the Institutional Review Board from the University of Connecticut (Exemption # X22-0011). The voluntary survey was carried out by Qualtrics (Provo, UT).

Data

The data for this study was collected through three distinct voluntary surveys for each of the demographics surveyed (Undergraduate Students, Graduate Students, and Faculty/Professionals). All three surveys were targeted at North American universities and distributed to potential participants primarily through email contact. Methods of distribution included listservs for Department Heads in Animal Science, Ecology, Plant Science, and Natural Resources. Additionally, the survey was made available to participants through Twitter. “Snowball sampling,” existing participants recruiting new participants for the study through word of mouth, was then used to distribute the surveys to different groups. The survey was open for 80 days between March 23, 2022 and June 10, 2022; and a total of 220 responses were collected across the three surveys.

All three surveys attempted to evaluate the current status of education in the field. Student questions assessed course availability, education quality in statistics, education quality in coding, and interest in the field. The Faculty and Professional survey questions evaluated the interest of new students, competency of new students, and willingness to fund external training for students. No questions in any of the surveys strictly evaluated knowledge; the questions were designed to evaluate perceptions and experiences with education in the field. A full list of all survey questions is listed in Appendices I, II, and III. Analyses on survey responses were conducted in R version 4.2.2 (R Core Team, 2022).

Before data analysis, the survey results were cleaned and filtered. Responses with duplicated IPs were removed from the poll. Additionally, incomplete responses and responses with a duration of less than two standard deviations from the mean were not used for analysis. Finally, Qualtrics response quality features (Qualtrics, Provo, UT)) were used to remove low-quality survey responses.

Resource Collection

The approach to resource cataloging was to identify the training that would be reasonably accessible to a student looking for further training in their respective fields. The search for resources was first conducted using generic search engine results, as this is the most likely way that students will search for these options. In addition, announcements that were distributed via listservs or conferences were added to the list. Once a resource was found, it was established whether the resource was still active. Then the following questions were answered:

1. What field is the resource targeting? (Animal Genetics, Plant Genetics, Population Genetics, Conservation Genetics, Ecological Data Analytics, Bioinformatics, or General Genetics)
2. What audience is the resource targeting? (Undergraduates, Graduates, Professionals, or Educators)
3. In what mode is the resource delivered? (Online Live, Online Recorded, Text, or In-Person)
4. Is the resource free? If not, what is the cost?

After the resources were found and annotated they were compiled into an accessible list for students. All resources relevant to the search parameters were included in this catalog.

Table 1

Summary of respondent information across all three surveys

Survey	Undergraduate	Graduate	Faculty/ Professionals	Total
Respondents (N)	85	61	74	220
University Affiliation (N)	3	9	NA	9
Field Representation (N)	10	23	16	NA

Results and Discussion

Number of Respondents

A total of 220 responses were received between March 23, 2022 and June 10, 2022. Of those answers, 85 were from undergraduate students, 61 from graduate students, and 74 from faculty and professionals (see Table 1). Despite an uncertain response rate, these results were lower than similar studies in the past (Hernandez et al., 2012; Serão et al., 2021) which received more than double the number of respondents within a narrower demographic scope. Due to the nature of Snowball Sampling, it is not possible to calculate a response rate relative to the number of individuals who were given the opportunity to respond. A conservative estimate of 100 students on average in each department, and 50 departments being contacted, would mean that 5,000 students could be reached by this survey. Because there were undergraduate responses from only three institutions, it is likely that most departments did not share the survey. It is possible that the low engagement seen in this study can be attributed to low student motivation, lack of distribution by administration, survey fatigue, or timing of survey distribution.

The COVID-19 pandemic changed most aspects of academia over the past three years and lead to overall fatigue, anxiety, and stress in the student population (Son et al., 2020; Wang et al., 2020). This fatigue may have decreased the motivation of students to participate in a voluntary survey and therefore may have lowered the number of respondents among this demographic. This problem is unique to the timeframe in which this project was conducted and would likely pose less of an issue in future studies.

In the student surveys, respondents were distributed among three universities for undergraduates and nine universities for graduate students. These numbers are lower than expected based on the distribution methods and the number of universities contacted throughout the course of the study. It is assumed that the surveys were not made available to students at universities which were not represented in the responses to any degree. This lack of response was interpreted as a lack of interest from the universities and departments in the field of agricultural data analytics, especially for undergraduate education.

Finally, the timing of the survey release may have impacted the availability of students and faculty to respond to the survey. Since the survey was made available in late March, it is likely that students and faculty may have been busy with end-of-semester assignments and exams. In future studies, this problem could be reduced by providing the survey earlier in the school year. Monetary incentives are also recommended, as they increase the response rate in online surveys (Ryu et al., 2006) without compromising data quality (Cole et al., 2015).

Undergraduate Student Survey

Respondents of the undergraduate survey consisted of 85 students across 3 universities (see Figure 1). Universities represented in this study were University of Connecticut, Michigan State University, and Mississippi State University. The majority of respondents (87.7%) were current bachelor’s students with no associate’s degree. There were 10 fields among the respondents with the majority (69.4%) being animal science students and the second most common (8.3%) being plant science students. Bias towards animal science students will likely have an impact on results of the undergraduate survey since there is a heavy prevalence of pre-veterinary students in the animal science field. Undergraduate students were asked questions regarding the guidance and education provided by their university in agricultural data analytics, the interest of participants in the field, and their capability in coding and statistical skills.

Results from the undergraduate survey suggest that most students are required to take some form of statistics training prior to graduation (see Figure 2). The majority of students also reported that their department does provide courses in agricultural data analytics and many of their department courses regularly employ data analysis or statistics (see Figures 3 and 4). We believe that this result may be biased due to a skewed sample population and a desirability bias of participants answering what may be a more desirable response. Additionally, 70% of students reported that their department employs professors working on data analytics in agriculture or the environment (see Figure 5). These statistics, if accurate, are promising for upcoming undergraduate students with an interest in the field to get involved with research and receive guidance to further their education or career. However, 63.5% of respondents were partially or fully unaware of what courses to develop their knowledge of agricultural data analytics (see Table 2). This gap suggests that while there are many professors employed in the field they are not supporting the undergraduate population nor encouraging participation in data analytics.

Figure 1

Demographics information for surveyed undergraduate students for universities present (A), degree status of students (B), and field spread (C)

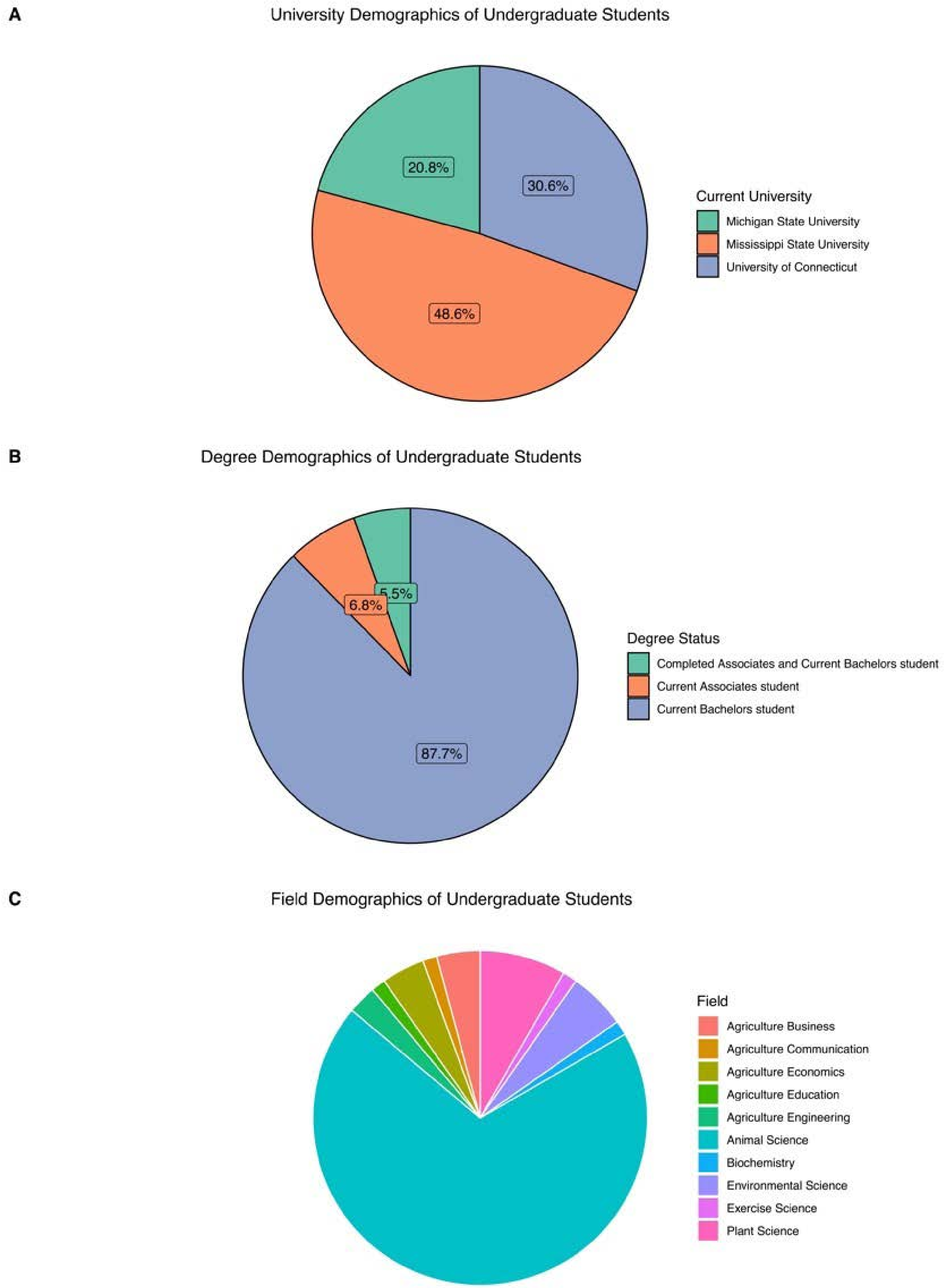


Figure 2

Proportion of surveyed undergraduate students that are required to take a statistics course to graduate

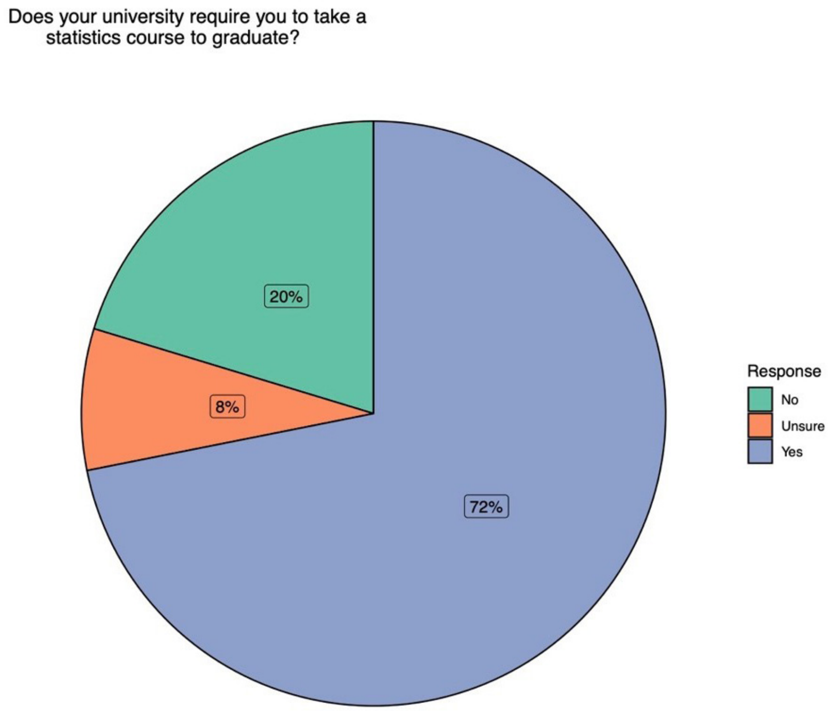


Figure 3

Number of undergraduate students that have courses in agricultural and environmental data analytics at their university

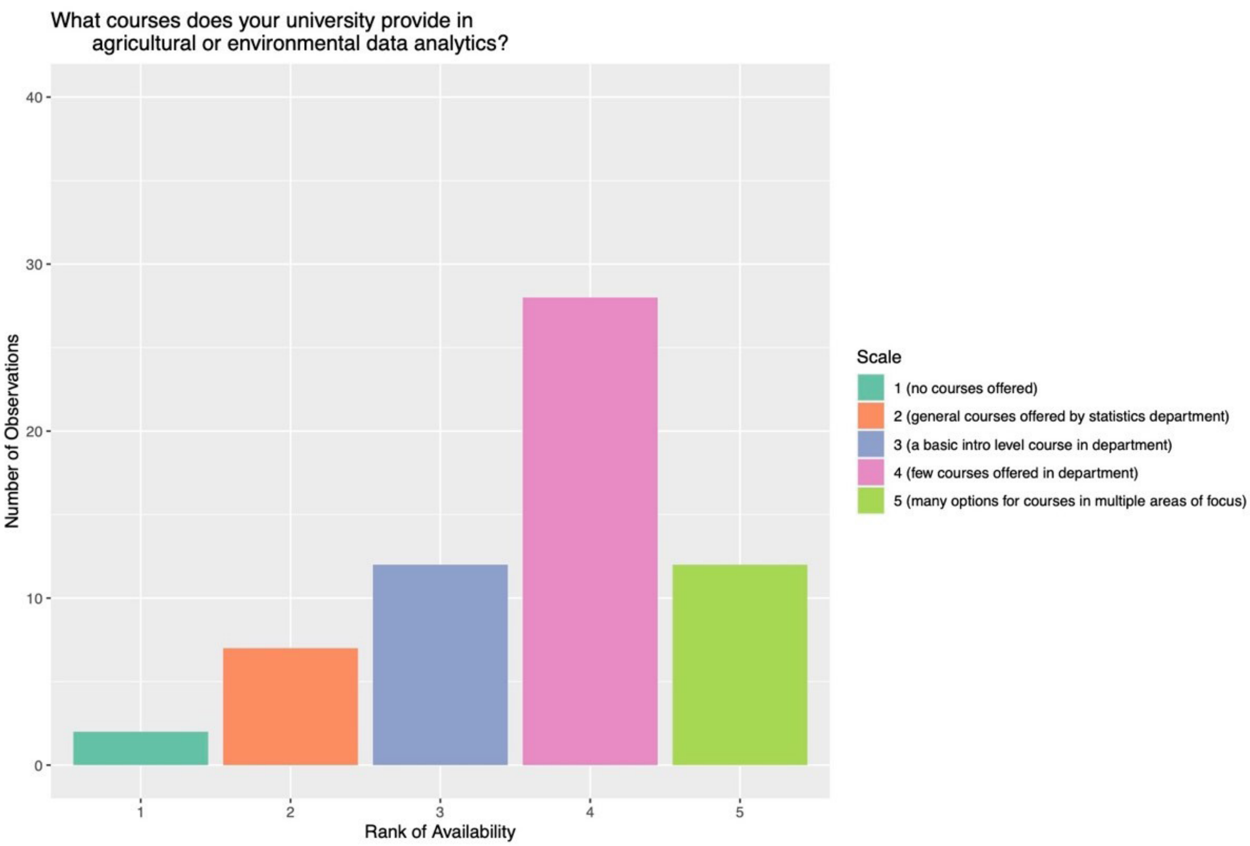


Figure 4

Number of undergraduate students that utilize data analytics, statistics, or mathematics in their courses

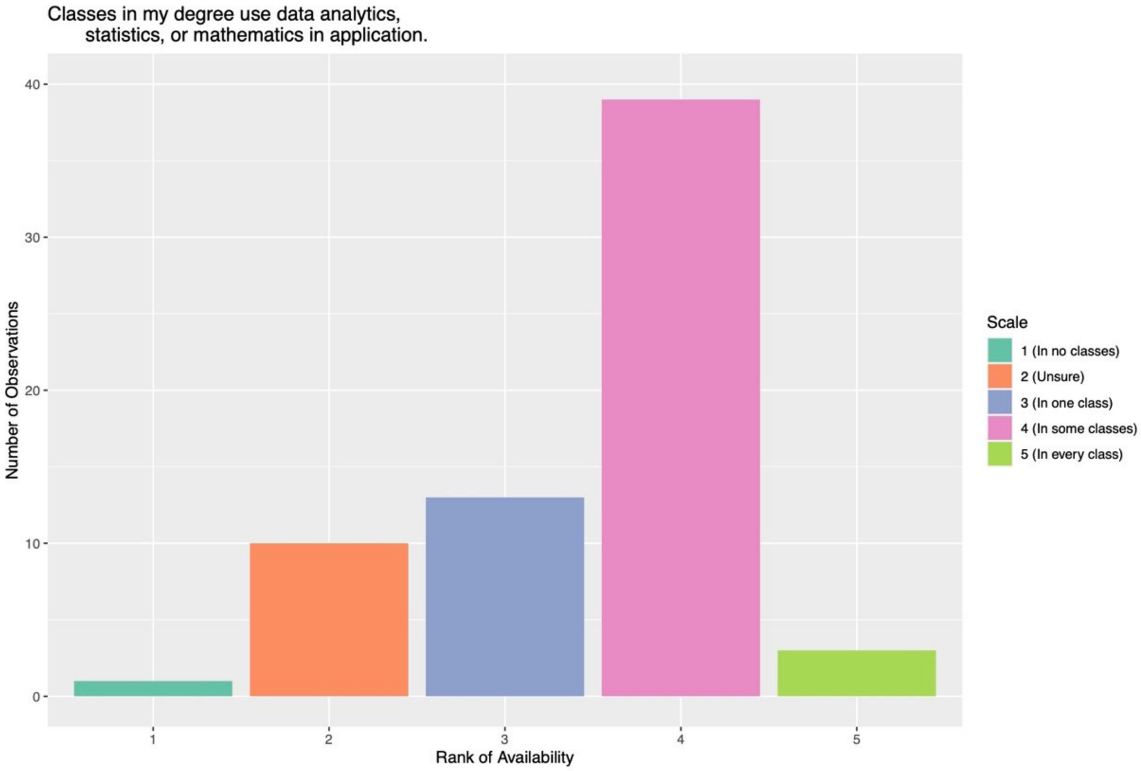


Figure 5

Proportion of surveyed undergraduate students whose university has faculty that utilize data analytics for topics in agriculture or the environment

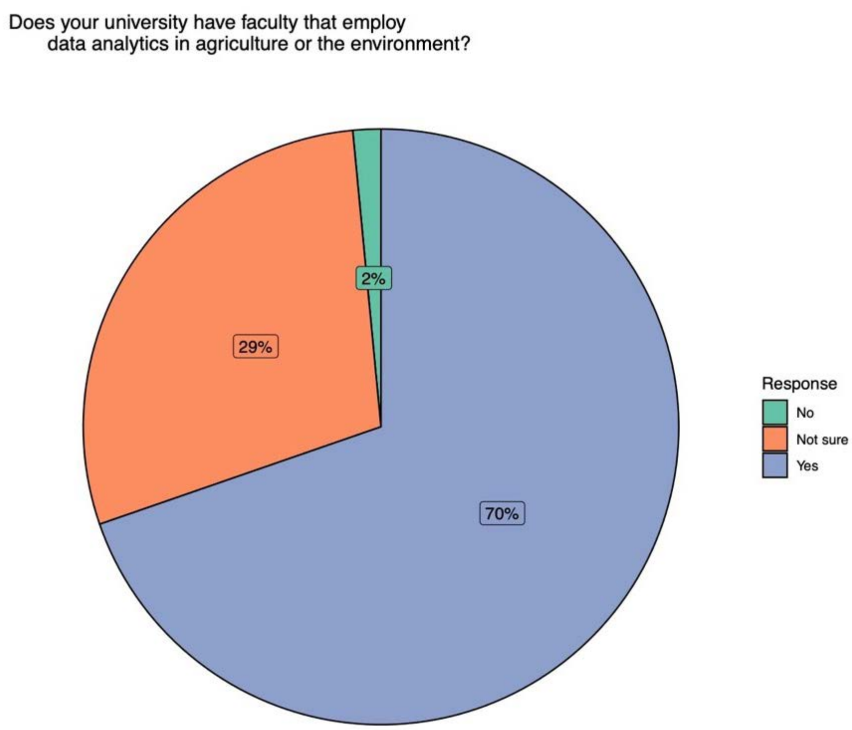


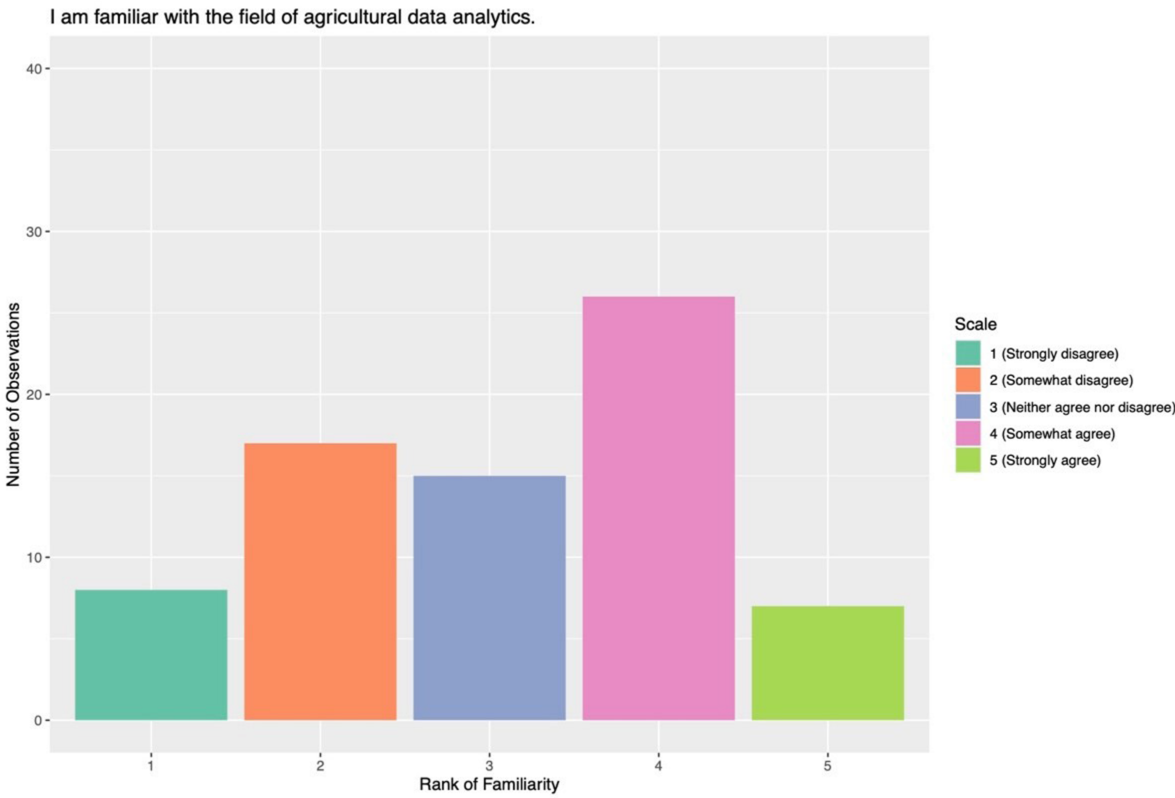
Table 2

Proportion of surveyed undergraduate and graduate students who are aware of courses to take for preparation in agricultural/environmental data analytics

Are you aware of what courses to take for pursuing a career in agricultural/environmental data analytics?	Undergraduate students	Graduate students
No, not aware of any courses	36.5%	13%
Somewhat aware of courses	44.5%	56%
Yes, aware of specific courses	19.0%	31%

Figure 6

Familiarity of surveyed undergraduate students with the field of agricultural data analytics



There was a high degree of interest in agricultural data analytics though there was a wide range of self-described familiarity with field (see Figures 7 and 8). Additionally, nearly half (44%) of the students searched for knowledge in the field outside of university coursework (see Figure 8). This further strengthens the argument that while students are interested in the field and may have resources available, they are not receiving guidance from the university and/or their professors.

There was a stark difference in the coding capabilities of undergraduates (see Figure 9) compared to their capabilities in statistics (see Figure 10). Many students believe they have some knowledge of statistical methods. The greatest understanding seemed to be for simple linear regression and descriptive statistics. However, all skills evaluated had some students with a high degree of understanding. This distribution illustrates that the majority of students surveyed were required to take a statistics class prior to graduation and generally use these skills in their courses. Additionally, the outlier students with a higher degree of understanding may be a reflection of students taking advanced elective courses or involved with undergraduate research, which reflects the proportion of US undergraduate students engaged in research activities (Douglass and Zhao, 2013). This result is encouraging and suggests potential for the undergraduate demographic to succeed in the field if given more guidance and opportunities. There was a stark difference in the coding capabilities of undergraduates (see Figure 9) compared to their capabilities in statistics (see Figure 10). Many students believe they have some knowledge of statistical methods.

Figure 7

Interest of surveyed undergraduate students in applying data analysis skills to problems in agriculture or the environment

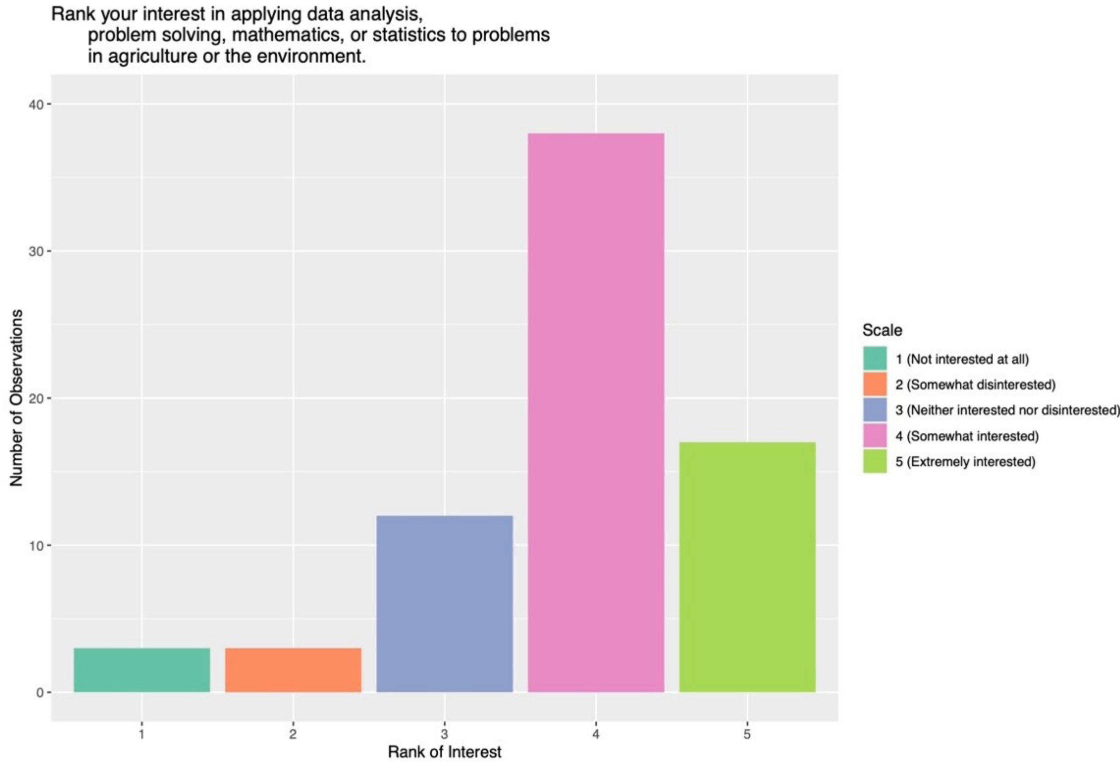


Figure 8

Interest of surveyed undergraduate students in applying data analysis skills to problems in agriculture or the environment

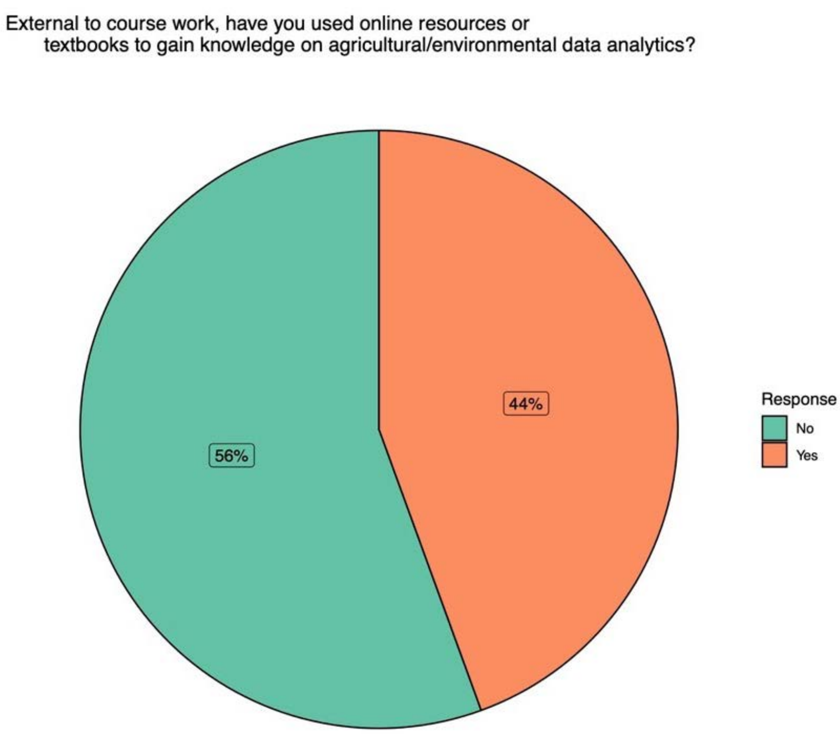


Figure 9

Coding capabilities of surveyed undergraduate students in seven common coding languages

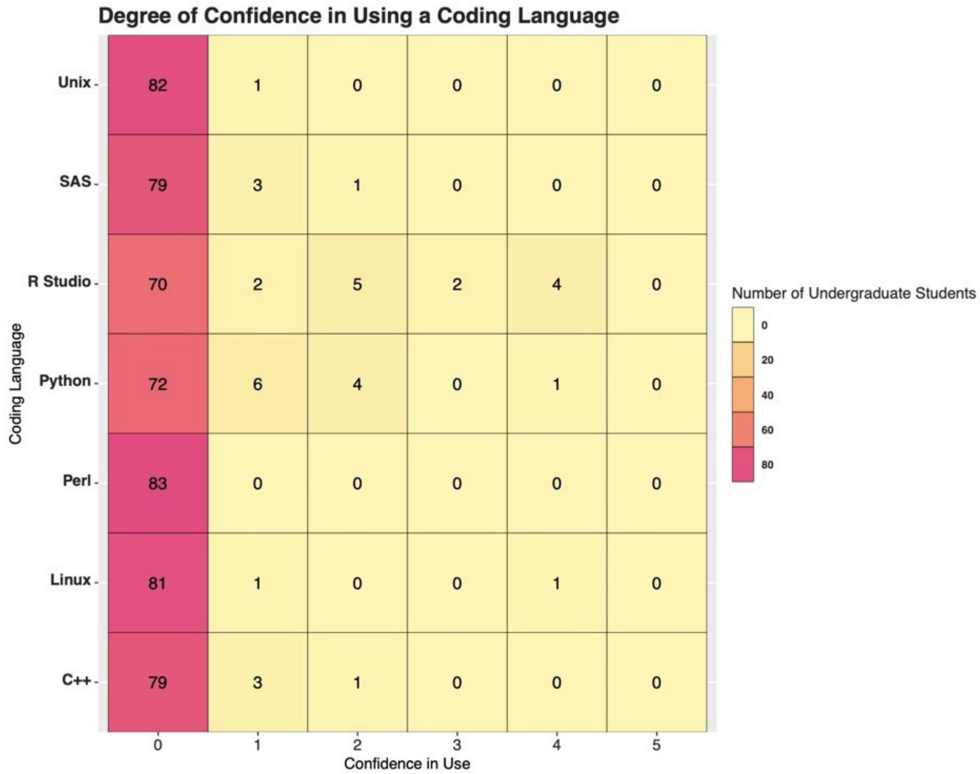
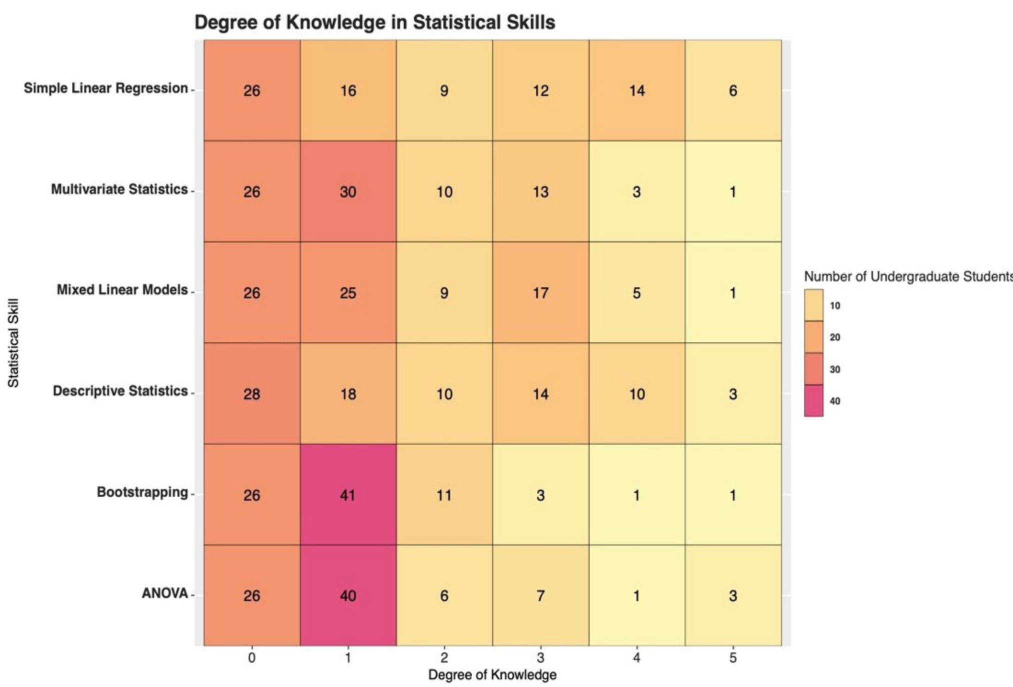


Figure 10

Statistical capabilities of surveyed undergraduate students in six common statistical skills



The greatest understanding seemed to be for simple linear regression and descriptive statistics. However, all skills evaluated had some students with a high degree of understanding. This distribution illustrates that the majority of students surveyed were required to take a statistics class prior to graduation and generally use these skills in their courses. Additionally, the outlier students with a higher degree of understanding may be a reflection of students taking advanced elective courses or involved with undergraduate research, which reflects the proportion of US undergraduate students engaged in research activities (Douglass and Zhao, 2013). This result is encouraging and suggests potential for the undergraduate demographic to succeed in the field if given more guidance and opportunities.

This degree of knowledge was not mirrored when evaluating coding capability. Nearly all students had no experience with any coding language and no student felt they had mastered any coding language (see Figure 9). The two languages that were most common among undergraduates were R/R Studio and Python which had 15.7% of students and 13.3% of students respectively with some degree of understanding in the languages. This is a large deficit when looking at the potential for success of undergraduate students in data analytics, however, this result was not unexpected when looking at previous studies on undergraduate capabilities in data management, specifically in the field of ecology. Strasser and Hampton (2012) surveyed professors on the content taught to undergraduate students in ecology regarding data management and coding and found that students were unprepared for their field. Their conclusions also found that instructors were either minimally educated on these topics or unaware of the available resources for teaching these topics. One of the aims of the Agricultural Genome to Phenome Initiative (AG2PI) is to train students for innovation (Tuggle et al., 2022), and such demand is illustrated by survey results. The next generation of agriculturalists will need to be trained to properly use relevant datasets and tools, and these educational gaps must be filled through curriculum change or additional external resources in order for students to succeed.

Graduate Student Survey

There were 61 respondents for the graduate school survey from 9 universities (see Figure 11). Of the 9 universities represented in the survey, the two most common responses were Mississippi State University and Michigan State University. Doctoral students (without a Master's degree) represented 48% of respondents with 35% of respondents being Master's students. The remaining 17% represented students who had completed a Master's degree and were pursuing a doctorate. There was a broad diversity in field responses from the graduate student survey with the most common responses being Animal Breeding and Animal Genetics. Similar to the undergraduate survey, graduate student questions evaluated the guidance and education provided by their university in agricultural data analytics and individual perceived capability in coding and statistical skills. However, graduate students were additionally asked about availability of external training.

More than half of the students surveyed (57.7%) do not feel they learn most about their field from university courses (see Figure 12). Instead, the students indicated learning about their field from their advisor (25%), online courses (21.2%), and other (11.5%) which included scientific papers and articles. It is likely that due to the lack of specificity in university courses in the field, students are required to learn about the specifics of agricultural data analytics outside of the university curriculum. Previous research supports that graduate students will often look to peers, advisors, or research experience to learn data analytics or programming skills rather than formal coursework (Theobold and Hancock, 2019). This pattern emphasizes the importance of quality online resources to provide training in topics that are not feasible to cover in more general university courses. Additionally, 87% of respondents felt they are partially or fully unaware of the courses to take for success in their field, which was an improvement relative to that of the undergraduate population (see Table 2). However, the predominant answer of "Somewhat aware of courses" is concerning in the graduate student population and indicates there may be insufficient guidance from advisors.

When asked about statistical capability, students had a distribution of confidence with a positive skew in the results. The majority of students felt they were somewhat confident in their ability to run analyses unassisted (see Table 3). This was reflected in the evaluation of specific skills as the majority of students felt they were at least somewhat confident in all skills listed (simple linear regression, multivariate statistics, mixed linear models, descriptive statistics, bootstrapping, and ANOVA) (see Figure 17). This result is an improvement on the undergraduate results and suggests improved knowledge with further education. Nearly all students agree that skills in data analysis are crucial for success in their field which seems reflected in the focus being placed on training in these areas (see Figure 15).

Similar to the results of the statistical skills analysis, there was variation in confidence for applying coding skills, however, the responses for coding confidence were consistent across the languages (see Figure 16). This result was reflected again in an evaluation of confidence using different coding languages. The majority of respondents reported having no understanding of the coding languages presented with the exceptions being R Studio and SAS, in which case there was an even distribution of students across skill levels. Though this language is certainly the most commonly used in data analysis (Oliver et al., 2019; Ozgur et al., 2022), it would be expected that students would utilize more than one language in their studies. Ozgur et al. (2022) compared several languages and concluded that both Python and R have desired characteristics to justify their use in the classroom and industry. This result may be due to a lack of knowledge of other languages available or lack of confidence in trying a new language, or simplified research focus due to disinterest of advisors to train students in more than one language. Additionally, the R language and R Studio are publicly available and free which makes it very accessible for use.

Figure 11

Demographics information for surveyed graduate students for Universities present (A), degree status of students (B), and field spread (C)

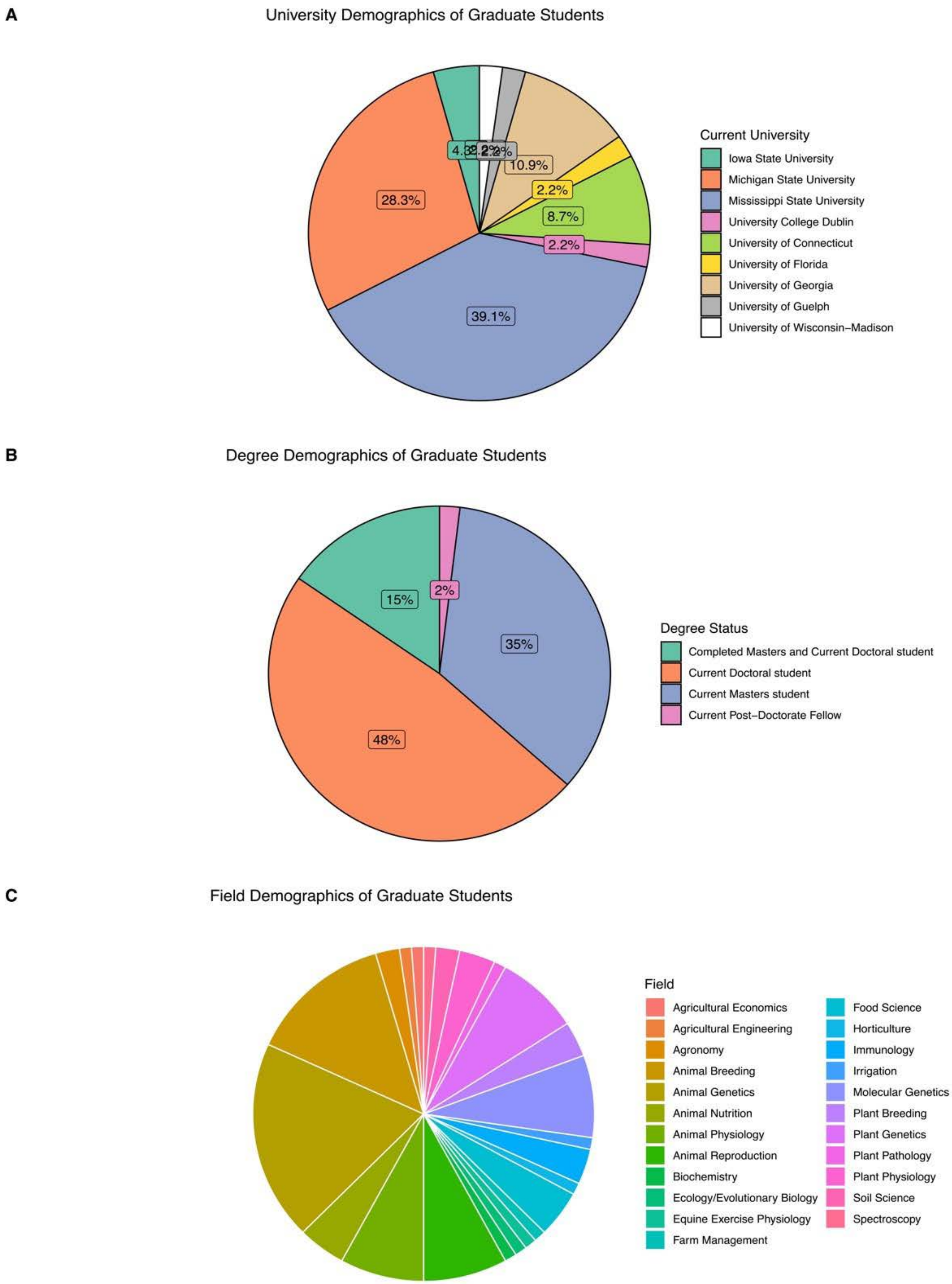


Figure 12

Sources that surveyed graduate students learn the most from in their respective fields

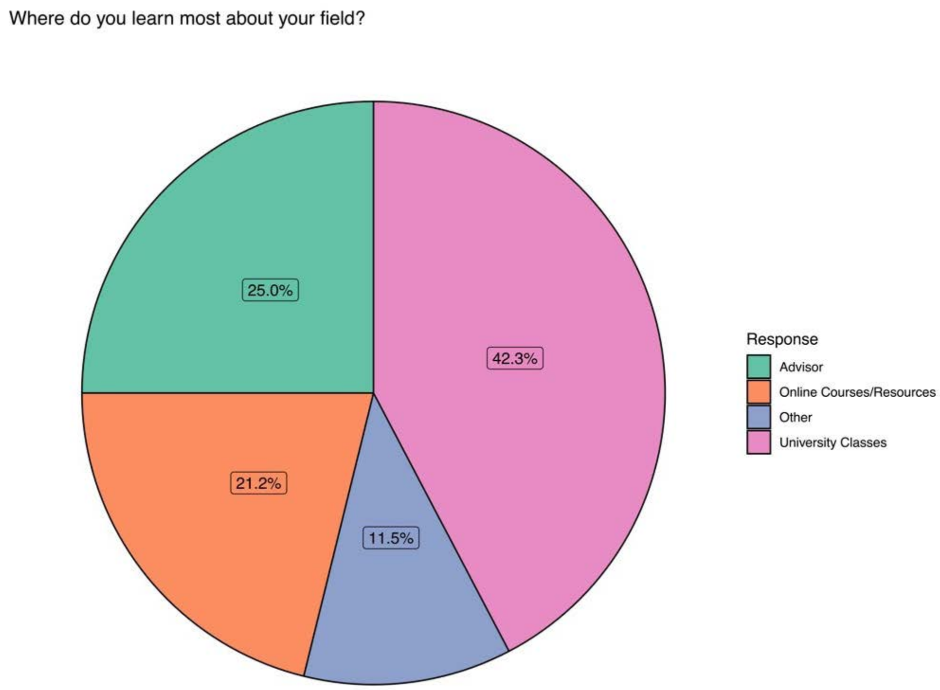


Figure 13

Proportion of surveyed graduate students that have been recommended to take courses external to their university for education in their field

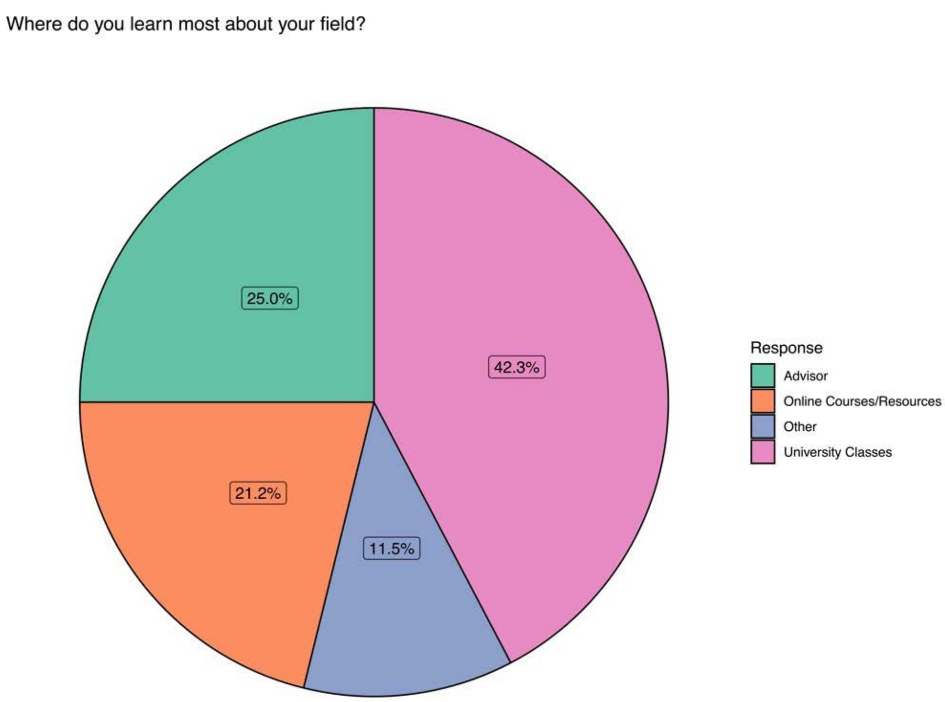


Figure 14

Proportion of surveyed graduate students who have external training funded by their advisor, department, or university

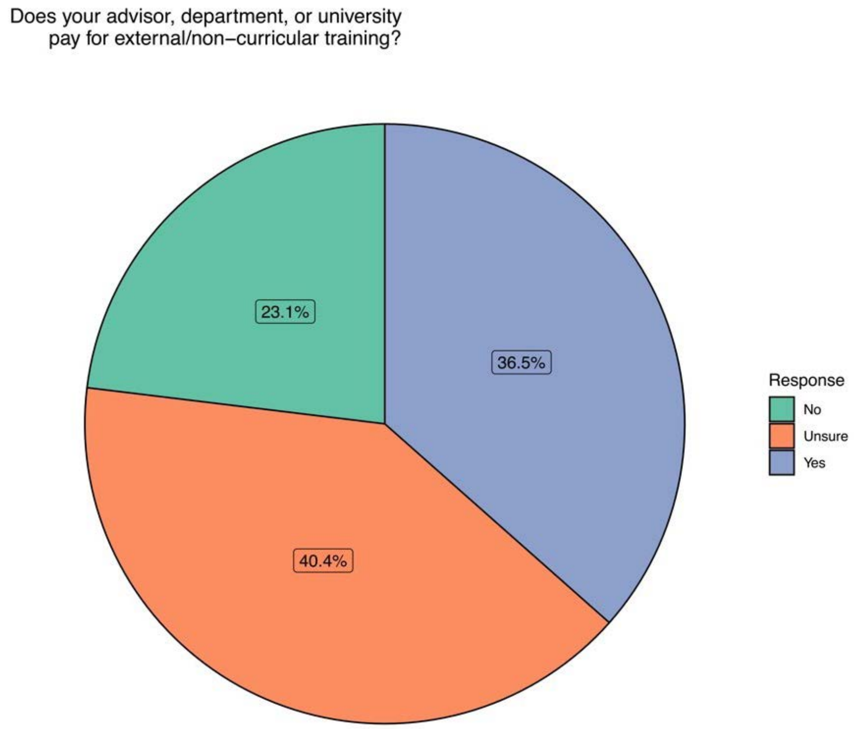


Table 3

Confidence of surveyed graduate students in writing a computational program for analysis and running statistical or data analysis without assistance

Response	How confident are you in your ability to write a program for analysis unassisted?	How confident are you in your ability to run statistical/ data analyses unassisted?
1 (Not at all)	11	3
2 (Just beginning)	7	10
3 (Intermediate)	14	7
4 (Somewhat confident)	13	21
5 (Extremely confident)	6	10

Alternatively, this result may be due to the small sample size observed in the study. Degree of knowledge in coding was similar in this study to existing studies focused on graduate students in environmental science (Hernandez et al., 2012). The lack of change in coding understanding from the undergraduate to the graduate level is concerning for the upcoming generation of scientists in the field. Multiple languages are commonly used in the field of agricultural data analytics (Kamilaris, et al., 2017), and scientists benefit from being versed in a variety of programming languages and software to achieve success in their field. It is important that the upcoming generation of professionals are versed in more than one resource to maintain the performance observed in the founders of the field.

Faculty and Professionals Survey

There were 74 respondents to the survey designed for faculty and professionals with the majority employed in academia (76%) followed by industry appointments (19%) and finally government appointments (5%) (see Figure 18). There were 16 fields of focus that were represented in this survey with the greatest percentage in the focus of Animal Genetics (32%) and Animal Breeding (19%). This survey did not request any information about employer to maintain a level of confidentiality among those surveyed. In their survey, faculty and professionals were asked to evaluate the competency of students, the interest of students, and their experience and opinions on external curricular training.

Results from the faculty and professionals survey indicate that students are more capable in their field upon

Figure 15

Perspective of surveyed graduate students on the value of data analysis in their field

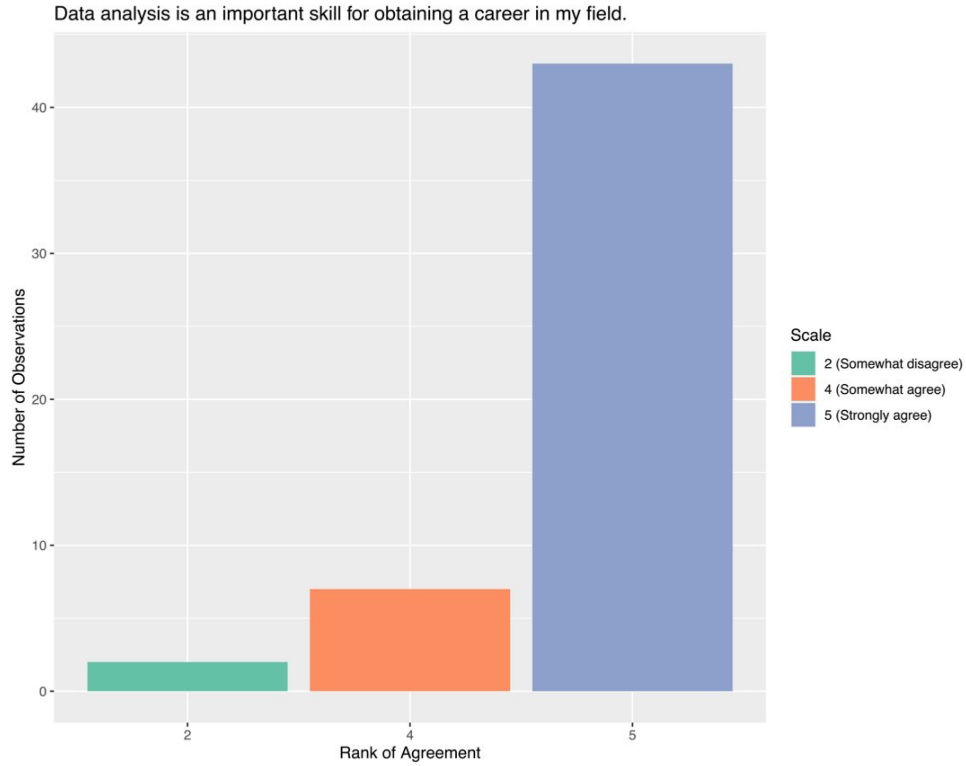
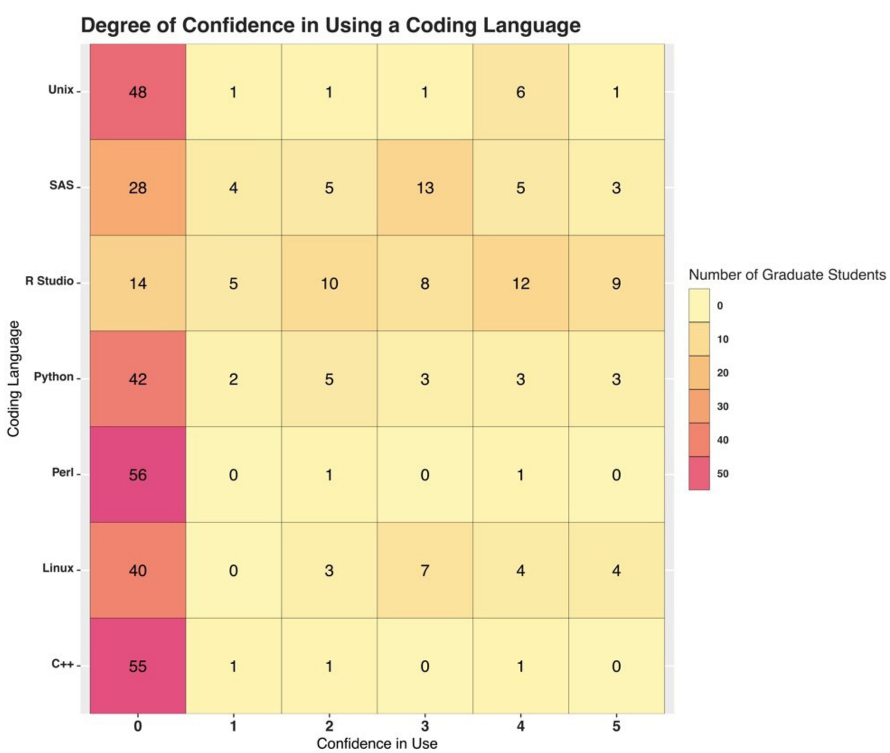


Figure 16

Confidence of surveyed graduate students in utilizing seven common coding languages



Coding Language	0	1	2	3	4	5
Unix	48	1	1	1	6	1
SAS	28	4	5	13	5	3
R Studio	14	5	10	8	12	9
Python	42	2	5	3	3	3
Perl	56	0	1	0	1	0
Linux	40	0	3	7	4	4
C++	55	1	1	0	1	0

exiting their position compared to when they entered (see Table 4). This result suggests that the training received throughout graduate school or a professional position is valuable. However, the majority of results were in the somewhat competent and not in the extremely competent category. This difference implies that there is still improvement to be made in these training programs and/or graduate level curriculum.

The majority of respondents perceived there to be a high degree of interest in their field from undergraduate students (see Table 4). High level of interest is encouraging for the field as this is the foundation for developing more qualified researchers in agricultural data analytics. This data also matches the results of the undergraduate survey which resulted in high levels of interest among the undergraduates surveyed.

Table 5 outlines results regarding the ability and willingness of faculty and professionals to budget for external training for their subordinates. Almost half of the individuals that responded to this question (47.5%) said they are currently able to budget for external curricular training for their employees or graduate students. This value aligns with what was found in the graduate survey where 36.5% of students had advisors that paid for external curricular training. Approximately 23% of surveyors were currently unable to budget for external curricular training for their employees or graduate students. A large proportion of participants (29.5%) were unsure whether they were able to budget for student or employee education. This statistic was particularly concerning because it may indicate that these professors have not considered this option when budgeting for projects. This may cause a gap in training that could easily be filled by encouraging this type of budgeting by employers.

When asked if they would budget for external curricular training for their employees or graduate students, nearly all respondents (84%) said they were likely to do so given the option. This statistic is encouraging for the future of student training as it seems that the most fiscally reasonable option for improving access to field-specific courses will be external curricular training rather than revisions to university graduate student curricula.

External Resources Study

Through an extensive online search, 33 available resources were identified that pertained to agricultural data analytics. These resources have been made publicly available with a forum to submit additional resources not listed (https://agdata.cahnr.uconn.edu/). Out of these, the majority (70%) were offered in Online mode and the second most common mode of instruction was in-person educational environments. There were more resources (60%) identified through this search that were freely available, however, all in-person resources had a cost with prices ranging from $250 USD to $2,000 USD. There were five non-free Online resources with prices ranging from $260 USD to $3,100 USD. The resources identified were predominantly catered towards graduate students (81%) with undergraduate (40%), professionals (24%), and educators (9%) having less available resources. Nearly half (45%) of the resources found here catered to more than one group.

There was a broad distribution of fields represented in the resources found, however, 80% of the resources were related to genetics (general, conservation, animal, and plant genetics). Other fields represented included ecology (12%), plant science (6%), and bioinformatics (3%).

The limited resources found in this study reflect the information gathered through student surveys. More than half of undergraduate students stated they had not used

Figure 17

Confidence of surveyed graduate students in utilizing six common statistical skills

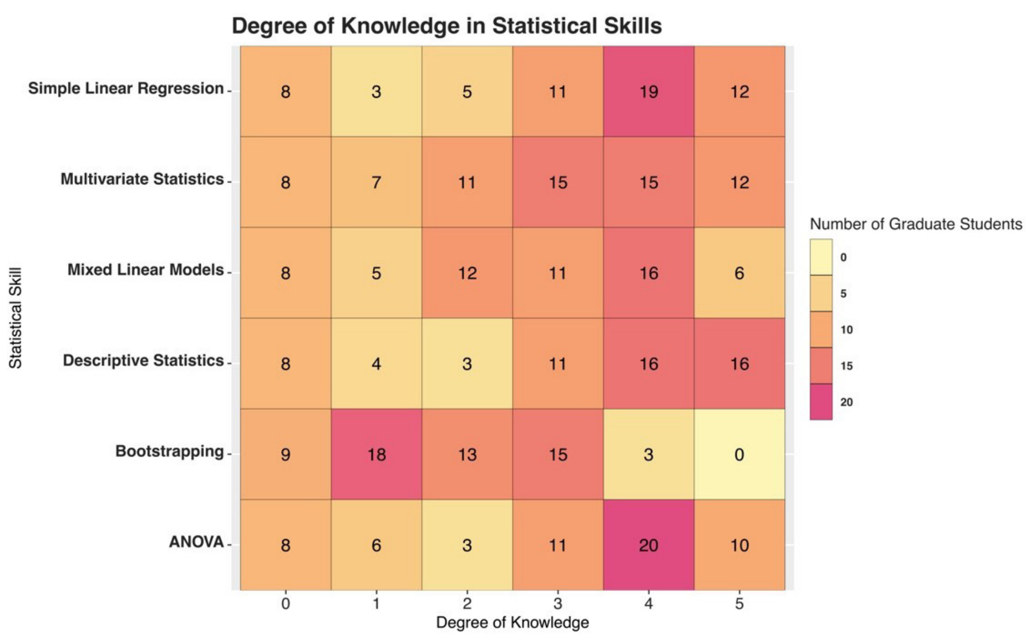
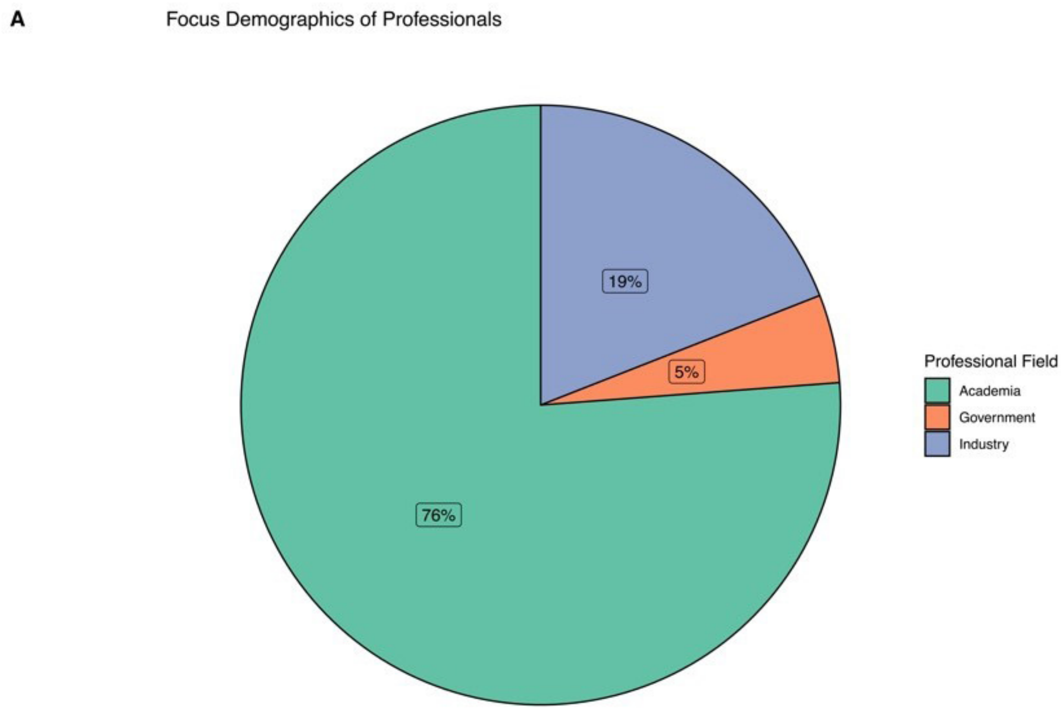


Figure 18

Demographics information for surveyed faculty and professionals including type of position (A) and field of focus (B).



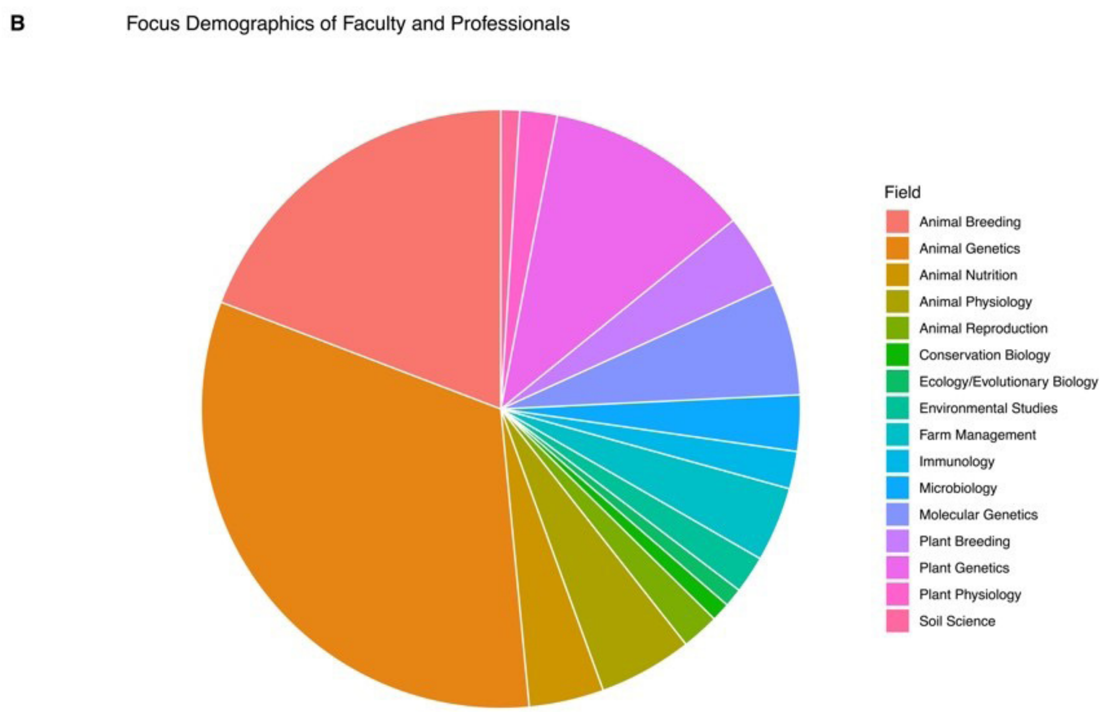


Table 4

Perspective of surveyed faculty and professionals on the degree of interest of undergraduates in their field of work as well as competency of students when entering and exiting an academic or industry position

Response	What degree of interest do you believe the next generation of students (undergraduates) has in your field?	What degree of competency do you believe students, in general, have upon entering graduate school or a professional position?	What degree of competency do you believe students, in general, have upon exiting graduate school or a professional position?
1 (low)	2	2	0
2	8	12	1
3	8	24	6
4	33	23	44
5 (high)	10	1	11

Table 5

Perspective on budgeting for external curricular training

Questions	Answers	Strongly Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree	Total
Are you able to budget for external curricular training for employees/ graduate students?	No, I am not able	2	0	3	4	5	14
	Unsure if I am able	0	1	3	6	8	18
	Yes, I am Able	0	0	1	14	14	29
	Total	2	1	7	24	27	61

resources outside of their courses to learn about their respective fields which aligns with the lack of resources that cater to undergraduate students. Resources primarily catered to graduate students which accounts for the greater percentage of graduate students who claimed to utilize external resources.

Considering the quantity and cost of the priced resources, the low number of graduate students that receive departmental or lab funding to pursue workshops or courses is concerning. Assuming an average stipend salary of approximately $25,000 USD, many courses were beyond the reach of disposable income of a graduate student and would not be feasible to attend (Kirchner and Petzoldt, 2022; Szkody et al., 2023). Workshops and courses outside of a student's home university may act as a solution to limited availability of university courses (Hernandez et al., 2012; Theobold et al., 2021), but support from advisors and departments is essential to make these courses possible for graduate students.

Limitations and Next Steps

The results in this study were primarily limited by the number of responses received. The target number of responses was approximately 500, however the number of respondents fell short of this number by approximately 280 responses. Additionally, there are follow-up questions and clarifying questions that may have improved the results of this study and provided a more well-rounded look at the issues. Due to these limitations, the current results should be used as preliminary data for a broader future study.

This study has revealed training gaps in statistics and coding skills of undergraduate and graduate students of agricultural data analytics. There does not seem to be adequate formal coursework offerings of these subjects in universities. In fact, most students obtain training outside of the university to gain their desired knowledge in agricultural data analytics (Theobold and Hancock, 2019). This lack of coursework can likely be attributed to saturation of the workload in existing courses or a lack of resources by the university to host program-specific courses in this field. There have been a few universities over the past 10 years that have revised their curriculum to include new interdisciplinary courses such as Standford's Bio-X program and the University of Illinois Urbana-Champagne's Animal Sciences x Computer Sciences and Crop Science x Computer Science degree programs. However, these new degree programs are not large enough or distributed to a large enough roster to train enough students to fill the upcoming job placings needed.

Previously, members in the field have suggested that these skills have become fundamental in modern research and thus must be incorporated throughout the educational systems akin to training in writing and mathematics (Nolan and Temple Lang, 2010), however, this change is not a feasible solution in the short term. Simpler solutions have been to use experience as the best form of training or to incorporate workshops into the curriculum for students (Chong et al., 2022; Gilbert et al., 2014; Lee, 2008). Hampton et al. (2017) outlined a roadmap for closing these gaps that included baseline coursework taught by other departments and specific training fostered through repetition in graduate and professional positions. Others in the field have encouraged the use of workshops to close the gap between general training in data analysis and programming and more specific applications in the student's field (Hernandez et al., 2012; Theobold et al., 2021). Finally, Rexroad et al. (2019) encouraged the development of a centralized repository with resources to improve the training quality for students as well as faculty. The repository developed as a result of this study will begin to fill the gaps in training found in this study, but the lack of available resources limits the impact. Further production of quality resources will be required to improve the training level of students in the coming years.

Summary

This study surveyed undergraduate and graduate students, and professionals in the field of agricultural data analytics to identify gaps in educational resources. This study was unable to reach an audience broad enough to understand the entirety of the issue that persists throughout the field of agricultural data analytics, however there are observations that can be made from these results. Undergraduate students are broadly untrained in statistics as well as data analytics and coding methods. While graduate students were observed to have more extensive training compared to undergraduate students, this training was primarily sought out external to the university curriculum or by learning from solely one resource – their advisor. Results from faculty and staff indicate an interest in improving the current education curriculum but there are not enough resources at this stage to make progress. Additionally, industry and graduate positions will require training and knowledge in the field that is not currently seen in the applicant market. An online catalog was curated with currently available educational resources to help students bridge some of the gaps in their training in data analysis.

References

Argueso, C. T., Assmann, S. M., Birnbaum, K. D., Chen, S., Dinneny, J. R., Doherty, C. J., Eveland, A. L., Friesner, J., Greenlee, V. R., Law, J. A., Marshall-Colón, A., Mason, G. A., O'Lexy, R., Peck, S. C., Schmitz, R. J., Song, L., Stern, D., Varagona, M. J., Walley, J. W., & Williams, C. M. (2019). Directions for research and training in plant omics: Big Questions and Big Data. *Plant Direct*, *3*(4), e00133. https://doi.org/10.1002/pld3.133

Chong, S. C. S., Anderson, K., Mackenzie-Stewart, R., Hobbs, C., & Conna, B. (2022). Embedding experiential learning in undergraduate health promotion curriculum. *Health Promotion Journal of Australia: Official Journal of Australian Association of Health Promotion Professionals, 33 Suppl 1*(Suppl 1), 9–16. https://doi.org/10.1002/hpja.629

Coble, K. H., Mishra, A. K., Ferrell, S., & Griffin, T. (2018). Big Data in Agriculture: A Challenge for the Future. *Applied Economic Perspectives and Policy*, *40*(1), 79–96. https://doi.org/10.1093/aepp/ppx056

Cole, J. S., Sarraf, S. A., & Wang, X. (2015). *Does use of survey incentives degrade data quality?* https://scholarworks.iu.edu/dspace/handle/2022/23761

Douglass, J. A., & Zhao, C.-M. (2013). Undergraduate Research Engagement at Major US Research Universities. *Center for Studies in Higher Education*, *14*(13).

Eisen, E. J. (2008). Can we rescue an endangered species? *Journal of Animal Breeding and Genetics*, *125*(1), 1–2. https://doi.org/10.1111/j.1439-0388.2008.00720.x

Gilbert, B. L., Banks, J., Houser, J. H. W., Rhodes, S. J., & Lees, N. D. (2014). Student Development in an Experiential Learning Program. *Journal of College Student Development*, *55*(7), 707–713. https://doi.org/10.1353/csd.2014.0072

Hampton, S. E., Jones, M. B., Wasser, L. A., Schildhauer, M. P., Supp, S. R., Brun, J., Hernandez, R. R., Boettiger, C., Collins, S. L., Gross, L. J., Fernández, D. S., Budden, A., White, E. P., Teal, T. K., Labou, S. G., & Aukema, J. E. (2017). Skills and Knowledge for Data-Intensive Environmental Research. *Bioscience*, *67*(6), 546–557. https://doi.org/10.1093/biosci/bix025

Hernandez, R. R., Mayernik, M. S., Murphy-Mariscal, M. L., & Allen, M. F. (2012). *Advanced Technologies and Data Management Practices in Environmental Science: Lessons from Academia. BioScience*, *62*(12), 1067–1076. https://doi.org/10.1525/bio.2012.62.12.8

Himesh, S., Rao, E. V. S. P., Gouda, K. C., Ramesh, K. V., Rakesh, V., Mohapatra, G. N., Rao, B. K., Sahoo, S. K., & Ajilesh, P. (2018). Digital revolution and Big Data: A new revolution in agriculture. *CAB Reviews*, *13*(021), 1–7. https://www.cabdirect.org/cabdirect/abstract/20183235919

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, *143*, 23–37. https://doi.org/10.1016/j.compag.2017.09.037

Kirchner, M., & Petzoldt, J. (2022). Entomology Graduate Stipends Fail to Cover Necessities. *American Entomologist*, *68*(1), 22–23. https://doi.org/10.1093/ae/tmac018

Lee, S. A. (2008). Increasing Student Learning: A Comparison of Students' Perceptions of Learning in the Classroom Environment and their Industry-Based Experiential Learning Assignments. *Journal of Teaching in Travel & Tourism*, *7*(4), 37–54. https://doi.org/10.1080/15313220802033310

Misztal, I. (2007). Shortage of quantitative geneticists in animal breeding. *Journal of Animal Breeding and Genetics*, *124*(5), 255–256. https://doi.org/10.1111/j.1439-0388.2007.00679.x

Nolan, D., & Temple Lang, D. (2010). Computing in the Statistics Curricula. *The American Statistician*, *64*(2), 97–107. https://doi.org/10.1198/tast.2010.09132

Oliver, J. C., Kollen, C., Hickson, B., & Rios, F. (2019). Data Science Support at the Academic Library. *Journal of Library Administration*, *59*(3), 241–257. https://doi.org/10.1080/01930826.2019.1583015

Ozgur, C., Colliau, T., Rogers, G., & Hughes, Z. (2022). MatLab vs. Python vs. R. *Journal of Data Science*, 15(3), 355–372. https://doi.org/10.6339/JDS.201707_15(3).0001

Qualtrics. Qualtrics XM: The Leading Experience Management Software. (2023, March 21). https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/response-quality/

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rexroad, C., Vallet, J., Matukumalli, L. K., Reecy, J., Bickhart, D., Blackburn, H., Boggess, M., Cheng, H., Clutter, A., Cockett, N., Ernst, C., Fulton, J. E., Liu, J., Lunney, J., Neibergs, H., Purcell, C., Smith, T. P. L., Sonstegard, T., Taylor, J., … Wells, K. (2019). Genome to Phenome: Improving Animal Health, Production, and Well-Being - A New USDA Blueprint for Animal Genome Research 2018-2027. *Frontiers in Genetics*, *10*, 327. https://doi.org/10.3389/fgene.2019.00327

Ryu, E., Couper, M., & Marans, R. (2006). Survey Incentives: Cash vs. In-Kind; Face-to-Face vs. Mail; Response Rate vs. Nonresponse Error. *Int J of Public Opin R*, *18*. https://doi.org/10.1093/ijpor/edh089

Serão, N. V. L., Petry, A. L., Sanglard, L. P., Rossoni-Serão, M. C., & Bundy, J. M. (2021). Assessing the statistical training in animal science graduate programs in the United States: Survey on statistical training. *Journal of Animal Science*, *99*(5), skab086. https://doi.org/10.1093/jas/skab086

Shakoor, N., Northrup, D., Murray, S., & Mockler, T. C. (2019). Big Data Driven Agriculture: Big Data Analytics in Plant Breeding, Genomics, and the Use of Remote Sensing Technologies to Advance Crop Productivity. *The Plant Phenome Journal*, *2*(1), 180009. https://doi.org/10.2135/tppj2018.12.0009

Son, C., Hegde, S., Smith, A., Wang, X., & Sasangohar, F. (2020). Effects of COVID-19 on College Students' Mental Health in the United States: Interview Survey Study. *Journal of Medical Internet Research*, *22*(9), e21279. https://doi.org/10.2196/21279

Strasser, C. A., & Hampton, S. E. (2012). The fractured lab notebook: Undergraduates and ecological data management training in the United States. *Ecosphere*, *3*(12), art116. https://doi.org/10.1890/ES12-00139.1

Szkody, E., Hobaica, S., Owens, S., Boland, J., Washburn, J. J., & Bell, D. (2023). Financial stress and debt in clinical psychology doctoral students. *Journal of Clinical Psychology*, *79*(3), 835–853. https://doi.org/10.1002/jclp.23451

Theobold, A., & Hancock, S. (2019). How Environmental Science Graduate Students Acquire Statistical Computing Skills. *STATISTICS EDUCATION RESEARCH JOURNAL*, *18*(2), 68–85. https://doi.org/10.52041/serj.v18i2.141

Theobold, A. S., Hancock, S. A., & Mannheimer, S. (2021). Designing Data Science Workshops for Data-Intensive Environmental Science Research. *Journal of Statistics and Data Science Education*, *29*(sup1), S83–S94. https://doi.org/10.1080/10691898.2020.1854636

Tuggle, C. K., Clarke, J., Dekkers, J. C. M., Ertl, D., Lawrence-Dill, C. J., Lyons, E., Murdoch, B. M., Scott, N. M., & Schnable, P. S. (2022). The Agricultural Genome to Phenome Initiative (AG2PI): Creating a shared vision across crop and livestock research communities. *Genome Biology*, *23*(1), 3. https://doi.org/10.1186/s13059-021-02570-1

Wang, X., Hegde, S., Son, C., Keller, B., Smith, A., & Sasangohar, F. (2020). Investigating Mental Health of US College Students During the COVID-19 Pandemic: Cross-Sectional Survey Study. *Journal of Medical Internet Research*, *22*(9), e22817. https://doi.org/10.2196/22817