INFORMATION SYSTEMS EDUCATION JOURNAL

Volume 21, No. 3
July 2023
ISSN: 1545-679X

In this issue:

This issue focuses on three areas of cybersecurity (two in teaching case study form), a chatbot for grad advising, and two papers on student competencies as they relate to curriculum. The chatbot leverages AI to provide automated chat responses for grad student advising questions, while the competency papers address integration of the IS2020 Elective competencies into the curriculum, along with a study of digital competencies of the incoming student class in a first-year experience course. Our first cybersecurity paper includes a study of cybersecurity defenses in health care, which might readily be used in cybersecurity or health care informatics courses. The other two papers are case studies – one focusing on risk management in manufacturing, and the other on the risk elements of autonomous vehicles.

iscap
Information Systems & Computing Academic Professionals

The **Information Systems Education Journal** (ISEDJ) is a double-blind peer-reviewed academic journal published by **ISCAP** (Information Systems and Computing Academic Professionals). Publishing frequency is five times per year. The first year of publication was 2003.

ISEDJ is published online (https://isedj.org). Our sister publication, the Proceedings of EDSIGCON (https://proc.iscap.info) features all papers, abstracts, panels, workshops, and presentations from the conference.

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the ISCAP conference. All papers, whether award-winners or not, are invited to resubmit for journal consideration after applying feedback from the Conference presentation. Award winning papers are assured of a publication slot; however, all re-submitted papers including award winners are subjected to a second round of three blind peer reviews to improve quality and make final accept/reject decisions. Those papers that are deemed of sufficient quality are accepted for publication in the ISEDJ journal. Currently the target acceptance rate for the journal is under 36%.

Information Systems Education Journal is pleased to be listed in the Cabell's Directory of Publishing Opportunities in Educational Technology and Library Science, in both the electronic and printed editions. Questions should be addressed to the editor at editor@isedj.org or the publisher at publisher@isedj.org. Special thanks to members of ISCAP who perform the editorial and review processes for ISEDJ.

INFORMATION SYSTEMS EDUCATION JOURNAL

Editors

Exploring the Strategic Cybersecurity Defense Information Technology Managers Should Implement to Reduce Healthcare Data Breaches

Maurice Mawel
mauricemawel@yahoo.com
U.S. Department of State (DoS)
Washington, DC – U.S.

Samuel Sambasivam
Samuel.sambasivam@woodbury.edu
Computer Science Data Analytics
Woodbury University
Burbank, CA 91504

Abstract

The principal investigator (PI) conducted this research study to explore the strategic cybersecurity defense IT Managers should implement to reduce healthcare data breaches. The PI conducted a systematic literature review and selected articles that addressed healthcare data security breaches, information disclosure, cybersecurity in healthcare, and IT Managers' lack of leadership competence. Also, various annotations from contextual, seminal, grey, and recent literature were used to find the research problem: The strategic cybersecurity defense IT Managers should implement to reduce healthcare data breaches has not been established. The PI collected secondary data from the Office of Civil Rights (OCR)/Department of Health and Human Services (HHS). The analysis, results, and findings are provided below in Part 9. Nevertheless, the routine interaction during health information exchange (HIE) on an interoperable network and the behavior of care providers and third parties who use computers and mobile devices to exchange patient data is an opportunity for cybercriminals to install malware or launch a ransomware attack to exploit potential vulnerabilities whether to steal sensitive data or compromise the network systems. Therefore, strategic cyber defense or an innovative security model would mitigate the threat. An exploratory design is used, and an epistemological approach supports the research method and design. The study is significant, and it will contribute to the body of knowledge the PI suggested for future research and provide major recommendations.

Keywords: Health information exchange, Innovative security models, Interoperability, and Ransomware attack.

1. INTRODUCTION

The PI explores the strategic cybersecurity defense IT Managers should implement to reduce healthcare data breaches. Healthcare data and information improve the quality of care and safety. Medical data security is considerable because it becomes a fundamental challenge that necessitates establishing proactive strategic cyber defense aiming to secure healthcare data and minimize the occurrence of data breaches when the data is at rest or in motion (Offner et al., 2020). Saleem and Naveed (2020) concluded that the existing security technologies could not meet many healthcare organizations' requirements or needs in preventing data breaches. One of the requirements is the association of different entities of an organization to create a framework for data breach mitigation strategies.

The healthcare industry collects, and stores highly sensitive and confidential data constantly exchanged between medical staff, patients, and third parties (Offner et al., 2020).
Human error is attackers' primary opportunity to compromise organizations and data breaches. Some human errors that lead to data breaches are social engineering, reusing passwords, storing plain-text passwords, ignoring intrusion warnings, failing to update software, and uploading sensitive data to the cloud (Saleem & Naveed, 2020).

However, the PI conducted a systematic literature review involving contextual, seminal, grey, and recent literature coupled with primary and secondary data related to prior studies—the study started by locating scholarly journal articles related to previous research on healthcare data breaches. Healthcare providers include doctors, clinicians, psychologists, dentists, pharmacists, surgeons, and nurses. Business associates or third parties are a health plan, including insurance companies, billing companies, electronic medical companies, and government programs like the Centers for Medicare and Medicaid Services (CMS). Therefore, cybersecurity and other security tools must be efficient or helpful in detecting, deterring, and identifying cyber threats and addressing security breaches quickly and systematically (Arockiasamy, 2021). The qualitative method and exploratory design were selected to help the PI examine the phenomenon, and the method is supported by the conceptual framework and its three sub-theories defined in the paper. The study will contribute to the body of knowledge and will benefit other researchers, practitioners, IT professionals, postgraduate students, many healthcare organizations, and non-healthcare organizations not only in the United States (U.S.) but in other countries.

2. RESEARCH PROBLEM

The problem to be studied is that the strategic cybersecurity defense IT Managers should implement to reduce healthcare data breaches have not been implemented (Sultana et al., 2020; Offner et al., 2020). Spanos and Angela found in 2016 that 37 related journal articles from 45 studies were related to healthcare data breaches.

The two researchers analyzed that 75.6 percent of the event studies indicated data breaches have significant and adverse effects on organizations' strategic goals (Juma'h & Alnsour, 2020). Bourquard and Berler (2021) suggest that innovative solutions are needed in healthcare for secure interoperability and digital services. The Health Information Portability and Accountability (HIPAA) reported 642 data breaches as the most significant breach in 2020 by healthcare providers, healthcare clearinghouses, and other third parties associated with 25% record-breaking, which is more than 2019 (HIPAA Journal, 2021).

Security Managers in most healthcare organizations lack modern security tools to mitigate cyber risks that lead to data breaches. However, security awareness and training are provided to hospital employees (Legaspi, 2019). In 2014, a hacker group known as Deep Panda or Black Vine targeted Anthem (a healthcare insurance company) network system with malware and masqueraded on their Virtual Private Network (VPN). Attackers had privilege escalation, which led them to cause a healthcare data breach of 78.8 million patients' records (Saleem & Naveed, 2020).

3. THE RESEARCH QUESTION

The central research question is outlined here: What the strategic cybersecurity defense IT Managers implement to reduce healthcare data breaches should be? In addition, the PI used twelve probe questions related to the central guiding question:

1) What is the frequency of healthcare data breaches in a healthcare organization in the U.S.?
2) What are the most prevalent data security breach challenges of the last three years in a healthcare organization?
3) How many years of experience have IT managers implemented cybersecurity defense to secure healthcare data or prevent data breaches?
4) What strategies were not successful for securing healthcare data from security breaches?
5) How do IT leaders plan to develop an interprofessional collaboration in establishing a strategic cybersecurity defense across healthcare organizations in the U.S?
6) What mode of security training, formal or informal, to improve the security of HER was provided during the past 12 months in a healthcare organization?
7) What model of cryptography algorithm is used to encrypt health data and enhance data security in a healthcare

organization?
8) What successful security strategies do IT Managers use to protect healthcare data from security breaches?
9) What strategies were not successful for securing data from breaches?
10) What cloud service is used in healthcare organizations (SaaS, PaaS, IaaS, CLAaaS, DaaS, and more)?
11) What are the most significant risks to storing healthcare data in the cloud?
12) Is there a data security breach while using any of the cloud services?

The PI used the above questions listed to deduce the answers from the collected secondary data and to analyze the cleaned data.

4. LITERATURE SEARCH STRATEGIES

The research study identified relevant search engines and databases, such as Computer Science Database (ProQuest), ProQuest dissertations and theses, Google Scholar, ACM Digital Library, and Science Direct, from Colorado Technical University (CTU) Doctoral Library database and other online databases. Besides, the PI examined peer-reviews, journal articles, publications, books or electronic books, periodic magazines, and some dissertations during the literature review. Contextual, seminal, grey, and recent literature dating from 2016 to 2022 were examined and selected to raise the relevance and impact of the management and organization of the research study (Adams et al., 2016).

However, numerous safeguards are needed because many incidents of unauthorized data security breaches have been reported regularly in the Indian healthcare organization, which yields the urgency of establishing effective security mechanisms (Aljuaid & Parah, 2021). Keywords search focus on healthcare data breaches, cause of health data security breaches, cloud data security management, and solutions to a security breach were used with a Boolean expression. Furthermore, the PI expanded the search to healthcare organizations in the Washington Metropolitan area, including the Department of HHS with its various entities: HIPAA Journal, Office of Civil Rights (OCR), Office of the National Coordinator (ONC) for Health Information, which publish periodic Magazines or journals, HHS Blog, and News Release on their website (https://www.hhs.gov/news). The department publishes reliable and valuable information related to healthcare and government-related issues, precisely healthcare information such as data security and privacy, COVID-19, and annual financial reports, and shares their archival data with its subscribers.

5. CONCEPTUAL FRAMEWORK

The research study aimed to explore the strategic cybersecurity defense IT Managers should implement to reduce data breaches in healthcare organizations in Washington metropolitan area. The research gap listed earlier was depicted from recent, contextual, grey, and seminal literature. The relationship between the three theories from the conceptual framework elicits interrelated behaviors. It demonstrates how IT Managers, covered entities, and third parties interact with patients' data, whereas a cybercriminal may exploit any potential vulnerability. The study problem was that the strategic cybersecurity defense IT Managers should implement to reduce healthcare data breaches has not been established. The conceptual framework supported the research study and defined the following three theories:

Routine Activity Theory (RAT)
These relationships reside in users' attitudes or routines when using computers or mobile devices to perform activities, access cloud healthcare data, and share EHRs. Some researchers have used RAT to study six cybercrimes or criminal behaviors and determined that the RAT elements showed varying results, although it presented clear indications for policy and crime prevention strategy (Cohen & Felson, 1979; Leukfeldt & Yar, 2016). Moreover, human error is associated with the routine and behavior of not creating a strong password, scanning their devices, or using anti-virus software. A lack of compliance with the security requirements might become an issue that may allow a cybercriminal to cause data breaches. The second cause of data breaches is employees' error, negligence, and voluntary disclosure of data (Swede et al., 2019).

Criminological Theory (CT)
Criminology theory explains the mystery of why crime occurs, and it allows us to consider the social context in which theories are formulated, published, and accepted as feasible (Lilly et al., 2018). The Identity Theft Resource Center (ITRC) report indicates data breaches happen the most through hacking, including ransomware, phishing, and other malware attacks (Swede et al., 2019). From 2010 to 2018, over 2,529 data breaches affected 194,74 million patient's EHRs due to hacking (Hossain et al., 2019).

Game Theory for Security (GTS)
The theory deals with cybersecurity or data breach security issues and describes the attack landscape and decision-making process. Further, it designs an optimal controller for cyber-physical systems, including relevant countermeasures based on hardware and software security on a network with various dynamic defense mechanisms and opportunities in a healthcare organization (Zarreh et al., 2018). Cybercriminals exploit vulnerabilities in cloud computing or computer networking where the devices and communication channels may lack proper strategic security systems and security awareness training. Thus, this can enable threats to health information (Ondiege et al., 2017). Increased connectivity to existing computer networks has exposed medical devices to new cybersecurity vulnerabilities (Swede et al., 2019).

6. POPULATION AND DATA COLLECTION PROCEDURE

The study sample is a subset of the selected population; the estimated population comprises three different healthcare organizations within the Department of HHS. This organization and its entities store and publish reports on healthcare data breaches, cybersecurity incidents, and various information concerning the U.S. Government and public and private sectors. These entities work concomitantly with the Department of HHS or closely with hospitals, clinics, the Centre of Medicare and Medicaid Services (CMS), and the Centers for Disease Control and Prevention (CDC), located in the Washington Metropolitan area. According to Arockiasamy (2021), the population listed above contains a data breach portal that is the publicly available database that subscribers can access and download, although the IRB requires evidence of data collection permission to enhance transparency.

However, a pilot study was conducted to determine the feasibility of the study when the PI tried to proceed with semi-structured interviews. However, the result of this study persuaded the PI to immediately switch to secondary data collection rather than meeting participants to collect data. Ln (2017) explained that a pilot study is the first step of the entire research protocol and is often a smaller-sized study assisting in planning and modifying the main research study. The PI made these changes after sending emails to potential participants or making phone calls to explain the objective and goal of the research study; however, the PI received no feedback from ten IT leaders. In other words, their non-response or lack of communication obligated the PI to switch to a plan B, which was Not Human Subject Research (NHSR). Koczkodaj et al. (2019) explained that security rules established by HITECH and HIPAA stipulate the unsecured protected health information breach is the official Act term that requires, in some cases, care providers and third parties or covered entities to provide breach notification to the media.

The IRB approval authorized the PI to start data collection after obtaining site permission, email, and evidence of data use permission letter. The IRB committee regulates the research study processes by providing guidelines and assurance that no physical harm or psychological, social, or economic risk should alter the study before, during, and after the study endeavor (Legaspi, 2019). After contacting the IT Manager at the OCR for NHSR, the PI reiterated the data collection purpose and explained that there would be no risk of harm or exposure. EHRs, PHI, PII, or EMRs are targeted the most by hackers because they contain patient data and information, such as credit cards, financial information, name and address, telephone number, and more that can be sold in a dark market (Rivers, 2020). Still, safety, privacy, and confidentiality were ensured before, during, and after the study.

7. INSTRUMENTATION, TRANSCRIPTION, AND RETENTION

The PI used instruments during data collection, such as a Dell laptop computer with 1TB disk storage capacity and 8GB RAM and a Segregate External Hard drive to download and store secondary data. There was also an Apple iPhone 11 series to take images or photographs of relevant data or pages (that the computer could not download) that contained valuable information related to healthcare data breaches. During data collection, the environment and site were exempted from noise and distraction. At the end of data collection, the PI kept the data and instruments inside a locked bag on the car's trunk from the site to his residence to ensure security, confidentiality, and accuracy of the data collected. The length of collecting secondary data was about 60 minutes before noon. Data transcription took place at home—extracting relevant data or information from computer storage media and using handwriting to collect data from iPhone cameras. However, as the HIPAA data retention policy recommends, the collected data has been saved inside one of my

external hard drives for at least six years (archive).

8. STUDY SAMPLE AND ANALAYSIS OF SECONDARY DATA

The PI kept 85 kilobytes (KB) of healthcare data for analysis in a word document and excel spreadsheets that addressed data breaches and hacking incidents. Among the 85 KB, the PI extracted 845 records from the 65536 excel spreadsheet of data breaches and hacking IT incidents published by the OCR from 2019 to 2021. To be more explicit, the PI compiled excel spreadsheet data but extracted only 845 records of data breaches and hacking IT incidents that happened precisely from January 15, 2021, to December 21, 2021, in the Washington Metropolitan area: District of Columbia (DC), Virginia (VA), and Maryland (MD), and the rest of the word document data that was over 15KB, only 2KB of that data were analyzed or coded. A proposed data analysis model starts with transcription, data coding, categorizing, theming, and interpretation. During the data analysis and interpretation, bias and error were eliminated to ensure accuracy or reliability or to determine trustworthiness.

The following figure (Figure 1) describes the above coding process or data analysis procedure:

Interpretation-Focused Coding and Theming Strategies

The data analysis process necessitated using the interpretation-focused strategy highlighted above. After labeling the research question, searching for relevant information in the data, and assigning an anchor code to the relevant information, the PI moved on by grouping or sorting related codes. Furthermore, the PI used the same strategy for theming by looking at the high frequency of each code to finally interpret the associated themes that constitute the data and information obtained. These twelve probe questions or sub-questions listed in point 3, the research question was assigned to a definite anchor code to look for relevant data that answers the research question—seemingly a participant's response to an interview question.

9. RESULTS AND FINDINGS

After analyzing each anchor code and grouping themes based on their relationships, the PI combined and compiled the patterns to generate three main or top themes:

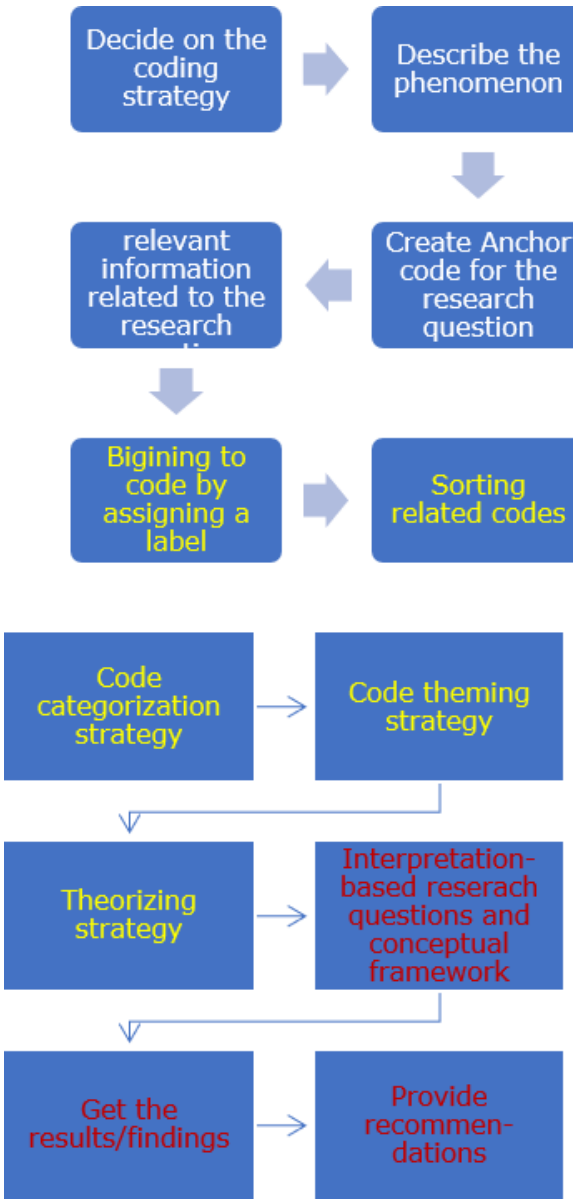


Figure 1: Understanding Data Coding, Analysis, and Interpretation

Top Theme 1

Corresponds to the lack of IT Managers' expertise or experience, and it appears 35 times.

Top Theme 2

Corresponds to third parties' involvement in HIE and appears 20 times.

Top Theme 3

Corresponds to major hacking/IT incidents and breaches appears 14 times.

Nevertheless, the implication of covered entities and business associates in accessing and sharing

EMRs or ePHI over the cloud or network is a root cause of hacking/IT incidents and the rise of data breaches. The PI found a lack of management expertise in establishing the strategic cybersecurity defense, and such weakness lures cyber-attackers to exploit vulnerabilities found during HEI. The PI used the triangulation approach by compiling the results generated from this analysis and OCR's annual report of data breaches and hacking IT incidents.

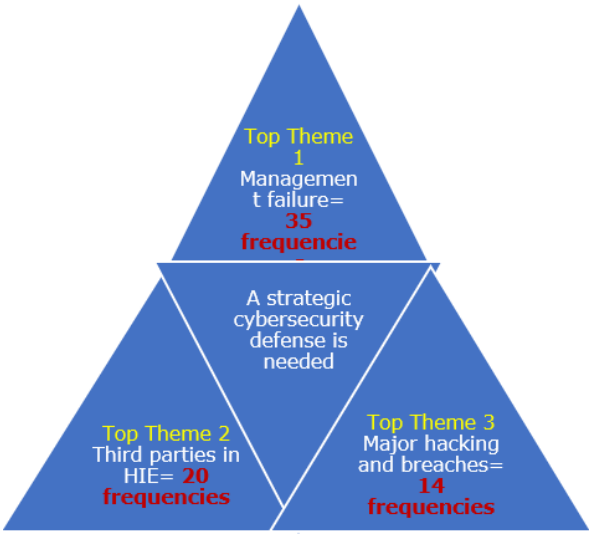


Figure 2: The Code frequency rates and three top themes

Therefore, the PI assumed that the results and findings were related to the conceptual framework and answered the research question based on the analysis. The analysis presented the following results: 73 codes and 69 frequencies, and IT Managers have failed to establish cybersecurity defense in an interoperable network. Furthermore, there is a relationship between the rise of healthcare data breach incidents and the involvement of covered entities and business associates in accessing and sharing EMRs or ePHI over the cloud or network. The frequency depends on each healthcare organization's security platform.

The PI coded the data manually, although doing so was time-consuming. Findings confirmed a lack of establishment of the strategic cybersecurity defense IT Managers should implement to reduce healthcare data breaches. The PI's coding strategy, known as interpretative-focused coding, sustained the interpretation of these codes and data. Still, qualitative data analysis, whether manual or with software tools, should support the researcher's efforts in presenting the results and findings that enhance trustworthiness or transparency (O'Kane et al., 2019).

10. FUTURE RESEARCH AND RECOMMENDATIONS

Further Research

The PI suggests a quantitative research method to investigate IT Managers' perception of interprofessional collaboration and continuous learning. Kidd et al. (2020) discussed that IT Security Engineering needs interprofessional skills to work collaboratively and negotiate the contributions of various disciplines while working on cybersecurity problems that require a multidisciplinary approach. Integrating knowledge and practices from multiple IT subject matter experts will enhance their perspective and ability to develop innovative solutions to fight against cyber criminals or threats. A second quantitative research method is suggested to explore third parties' level of education as they use computers to access health information.

Recommendations

According to Chigada and Madzinga (2021), the WHO research analysis indicates that cyber-attacks have increased quantitatively because research studies on cyber risks are still insufficient. Therefore, innovative security management practices are needed to solve cyber threats in the healthcare industry. As such, the PI recommends strategic cybersecurity defense to IT Managers and other practitioners; that is, IT Managers should strive to:

- Use modern or sophisticated tools to recognize threat actors.
- Categorize users into groups based on risk level and know the most attacked user.
- Ensure the security platform addresses use cases and compliance concerns.
- Enhance security awareness training and education on email phishing techniques, ransomware attacks, and social engineering because users generally represent a risk to any organization. Moreover, covered entities and third parties shall take periodic security awareness training-based information systems security courses.

11. CONCLUSION

The PI conducted the qualitative research study to explore the strategic cybersecurity defense IT Managers should implement to reduce healthcare data breaches. The question that was addressed

in the study was: What strategic cybersecurity defense IT Managers implement minimize healthcare data breaches should be? The PI collected and analyzed secondary data to answer the above central research question and twelve other probe questions. The findings provided a theoretical explanatory approach to the research problem and the contextual aspects associated with IT Managers' lack of expertise in implementing the strategic cybersecurity defense that should reduce healthcare data breaches. In addition, the PI analyzed that users' routine and behavior in accessing and sharing patients' healthcare data was found to be associated with cyber-attackers' exploitation of vulnerabilities in an interoperable network. The PI provided valuables recommendations and suggested further research studies in quantitative methods to investigate IT Managers' perception of interprofessional collaboration and continuous learning, and secondly to conduct another quantitative research method to explore third parties' level of education as they use computers to access health information.

12. ACKNOWLEDGMENTS

I dedicate this research study and doctorate to three people— God with his Son Jesus-Christ, my late mother Francisca Ngo Bikai, and my nuclear family. My mother is the inspiration for my education; she took care of me after my dad's passing in 1965; sacrificed herself to enroll me in school in a small city of Cameroon known as Eseka. The Almighty God, through Jesus Christ, opened a door of opportunities by taking me from my original country Cameroon to the United States. Also, they supported me by answering my prayers during the entire doctoral journey and achieving this goal. Lastly, I will not forget my wife Therese-Desiree, my daughter Elodie, and my three sons: Rene, James, and Nathaniel. They motivated me both intrinsically and extrinsically, although they sometimes worried about my long-standing and intermittent confinement in the study room. However, they did not stop showing me love, a positive mood, or a successful mindset. I will also give a big thanks to all my instructors from CTU with a special bonus to Dr. Sambasivam for making me a great man through this doctoral journey.

13. REFERENCES

Adams, R. J., Smart, P., & Huff, A. S. (2016). Shades of grey: Guidelines for working with the grey literature in systematic reviews for management and organizational studies. *International Journal Management Reviews*. https://doi.org/10.1111/ijmr.12102

Aljuaid, H., & Parah, S. (2021). Secure patient data transfer using information embedding and hyperchaos. *Sensor Networks and IoT for E-health Applications 21(1*), 282. https://doi.org/10.3390/s21010282

Arockiasamy, A. A. (2021). *Impact of information breaches on health care records* (Publication No 28318026) [Doctoral dissertation, Walden University]. ProQuest Dissertations and Theses Global.

Bourquard, K., & Berler, A. (2021). Health information exchange: The overarching role of integrating the health enterprise (IHE). *Introduction to Nursing Informatics*. https://doi.org/10.1007/978-03-030-58740-6_5

Chigada, J., & Madzinga, R. (2021). Cyberattacks and threats during COVID-19: A systematic literature review. *South African Journal of Information Management 23*(1). https://doi.org/10.4102/sajim.v23i1.1277

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review 44(*4), 588-608. https://www.jstor.org/stable/2094589

HIPAA Journal. (2021, January 19). 2020 Healthcare data breach report: 25% Increase in breaches in 2020. *HIPAA Journal*. https://www.hipaajournal.com/2020-healthcare-data-breach-report-us/

Hossain, M. M., Hong, A. Y., & Symp, A. (2019). Trends and characteristics of protected health information breaches in the United States. *AMIA Annual Symposium Proceedings Archives*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7153056/

Juma'h, A. H., & Alnsour, Y. (2020). The effect of data breaches on company performance. *International Journal of Accounting and Information Management 28(*2), 275-301. https://doi.org/10.1108/IJAIM-01-2019-0006

Kidd, J., Kaipa, K., Sacks, S., Ringleb, S., Pazos, P., Gutierrez, K., Ayala, M. O., & de Souza Almeida, M. L. (2020). What do undergraduate engineering students and pre-service teachers learn by collaborating and teaching engineers and coding through robotics? *Asee's Virtual Conference at Home*

with Engineering Education. https://par.nsf.gov/servlets/purl/10189269

Koczkodaj, W. W., Masiak, J., Mazurek, M., Strzalka, D., & Zabrodskii, P. F. (2019). Massive health record breaches evidenced by the office for civil rights data. *Iranian Journal of Public Health* 48(2), 278-288. https://pubmed.ncbi.nlm.nih.gov/31205882/

Legaspi, J. (2019). *Exploring the cybersecurity measures healthcare managers use to reduce patient endangerment resulting from backdoor intrusions into medical devices* (Publication No 13813275) [Doctorate dissertations, Colorado Technical University]. ProQuest Dissertations and Theses Global.

Leukfeldt, E. R., & Yar, M. (2016) Applying routine activity theory to cybercrimes: A theoretical and empirical analysis. *Deviant Behavior 37(*3), 263-280. https://doi.org/10.1080/01639625.2015.1012409

Lilly, J. R., Cullen, F. T., & Ball, R. A. (2018). *Criminological theory: Context and consequences*. (7th, Ed.). Sages Publishing.

Ln, Y. (2017). Introduction of a pilot study. *Korean Journal of Anesthesiology 70*(6), 601-605. https://doi.org/10.4097/kjae.2017.70.6.601

Offner, K. L., Sitnikova, E., Joiner, K., & MacIntyre, C. R. (2020). Towards understanding cybersecurity capability in Australian healthcare organizations: A systematic review of recent trends, threats, and mitigation. *Intelligence and National Security 35*(4), 556-585. https://doi.org/10.1080/02684527.2020.1752459

O'Kane, P., Smith, A., & Lerman, P. M. (2019). Building transparency and trustworthiness in inductive research through computer-aided qualitative data analysis software. *RMD Academy of Management.* https://doi.org/10.1177/1094428119865016

Ondiege, B., Clarke, M., & Mapp, G. (2017). Exploring a new security framework for remote patient monitoring devices. *Computers MDPI, 6(1*), 1-13. https://doi.org/10.3390/computers6010011

Rivers, L. (2020). *Strategies for reducing the risk of data breach within the Internet cloud* (Publication No 28264644) [Doctorate dissertations, Walden University]. ProQuest Dissertations and Theses Global

Saleem, H., & Naveed, M. (2020). Sok: Anatomy of data breaches. *Proceedings on Privacy Enhancing Technologies 2020(*4), 153-174. https://doi.org/10.2478/popets-2020-0067

Sultana, M., Hossain, A., Laila, F., Taher, A. K., & Islam, N. M. (2020). Towards developing a secure medical image sharing system based on zero trust principles and blockchain technology. *BMC Medical Informatics & Decision Making*. https://doi.org/10.1186/s12911-020-01275-y

Swede, M. I., Scovetta, V., & Eugene-Colin, M. (2019). Protecting patient data is the new scope of practice: A recommended cybersecurity curricula for healthcare students to prepare for this challenge. *Journal of Allied Health, 48(*2), 148-156. https://www.ingentaconnect.com/content/asahp/jah;jsessionid=f6mw6pmganap.x-ic-live-01

Zarreh, A., Saygin, C., Wan, H., Lee, Y., & Bracho, A. (2018). A game theory-based cybersecurity assessment model for advanced manufacturing systems. *Procedia Manufacturing 26(*1), 1255-1264. https://doi.org/10.1016/j.promfg.2018.07.162

Managing Graduate Student Advisement Questions during a Season of Explosive Growth: Development and Testing of an Advising Chatbot

Reshmi Mitra
Department of Computer Science
rmitra@semo.edu

Dana Schwieger
Department of Management
dschwieger@semo.edu

Robert Lowe
Department of Computer Science
rlowe@semo.edu

Southeast Missouri State University
Cape Girardeau, MO 63701 USA

Abstract

Many universities have, or are facing, the task of providing high quality essential customer services with fewer financial and human resources. The growing diversity of students, their needs and proficiencies, along with the increasing variety of university program offerings, make providing customized, on-demand, automated solutions crucial to delivering high quality customer service. In this paper, the authors describe the development of an artificial intelligence-backed chatbot to aid in answering student advisement questions. The main objective of the chatbot is to provide 24/7 supplemental program specific advising assistance to graduate computer science majors to lessen the advising load of university faculty. In this paper, the authors describe the development of a chatbot prototype using IBM Watson Assistant running on a specially designed website and Slack platform to address frequently asked questions of computer science graduate students. Results of a pilot study conducted on 99 graduate students indicate that the prototype is a positive step forward in making student advising accessible, usable, and scalable to a broader audience.

Keywords: Conversational Recommender Systems, Machine Learning, Natural Language Processing, Artificial Intelligence, Learning Analytics

1. INTRODUCTION

Over the past few years, higher education has faced significant changes. The tentacles of COVID-19 stretched across multiple areas of higher education from changes in teaching, meeting, and conference modalities, to classroom size limits and gathering restrictions, to prompting a significant decline in national and international new student applications (Redden, 2021). Enrollment numbers were hit hard with restrictions placed on international travel. The

OpenDoors report on International Educational Exchange (IEE) - funded by the U.S. Department of State - indicated a 46% drop in new international students and a 15% drop in the total international student population for U.S. higher educational institutions from the 2019/20 to 2020/21 academic years (Redden, 2021).

However, according to a snapshot survey conducted by IEE and nine other higher education associations, the fall 2021 enrollment of new international students surged 68% over the previous fall 2020 numbers (Redden, 2021). This surge represented a 4% growth in the total number of international students across 860 U.S. colleges and universities. For university programs whose majority population consists of international students, the sudden onslaught of students following the reduction of U.S. visa restrictions caught many programs unprepared and under-staffed to provide standard services (e.g., advisement, small class sizes, and common services requiring individual interactions.)

In addition, university COVID-19-related protocols encouraging electronic communication, added to the number of individual interactions. The combination of sudden expansions in enrollment numbers, especially among international students, and multiple individualized repetitive responses resulted in higher education advisors being overwhelmed with inquiries from current and prospective students. In this paper, the authors describe a pilot study in which a chatbot was developed to assist in the advising workload of a graduate computer science program. The goal of the authors in the development of the chatbot is to provide high-quality, program-specific, immediate, individualized responses to student questions as a means of reducing the advising workload on already overtaxed faculty.

2. USE OF CHATBOTS IN HIGHER EDUCATION

The use of conversational recommendation systems, commonly referred to as chatbots, in higher education is not uncommon (Ho et al., 2018; Kumar & Rose, 2010; Nittaya et al. 2020; Nwanko, 2018; Sjöström et al., 2018; Wang et al, 2020). Chatbots have been used in education to boost student engagement (Wang et al., 2020), provide individualized tutoring (Goel, 2020), answer and respond to general needs and questions (Dibitonoto et al., 2018; Goel, 2020; Rana, 2019), and provide assistance when personnel resources are insufficient (de Lange et al., 2021).

Heller et al. (2005) introduced Freudbot to psychology students to determine their reaction toward chatbots. Although the students had basically neutral opinions, they recognized the potential chatbots had in the future. Kumar and Rose (2010) described the use of the Basilica architecture to develop chatbots that could provide engineering students with individualized tutoring. The Basilica framework demonstrated that chatbots could be developed that could give students individualized advising sessions. Further expanded research on Basilica demonstrated how chatbots could be used to assist students with productive feedback as well as grade student essay answers (Dyke et al., 2013). The chatbot used synonyms and example sentences to grade and compare student answers with the correct answer key. Jill Watson SA (Social Agent) was a virtual assistant (or chatbot) introduced in two online Master of Science classes in the computer science program at Georgia Tech. The chatbot encouraged students to engage with the other students in their courses and build communities (Wang et al., 2020). Jill Watson SA, an integration of three other AI technologies (VERA, Jill Watson Q&A, and Agent Smith), was used to help students get answers to common course questions saving instructors hundreds of hours of time (Goel, 2020).

Students have a variety of common general needs and questions that they would like answered quickly (Dibitonoto et al., 2018; Rana, 2019). Previous research has shown that chatbots offer a better experience for users with general inquiries than traditional menu-based interfaces (Adamopoulou & Moussiades, 2020). Chatbots have been used in connection with web sites to direct students to appropriate web pages (Ghose & Barua, 2013). They have also been used in a study examining automated consent management of learning management system logs (de Lange, et al., 2021). Thus, chatbots have been proven an effective resource to students as well as others in many different cases (Goel, 2020; Yilmaz & Yilmaz, 2020).

2.1 Chatbots and advisement

Jiang, Pardos, & Wei (2019) noted that students needed help with course advisement and selection and that chatbots could be used to provide that assistance (Suschevskiy & Khalil 2021). Wagner et al. (2021) found that students do not always know what courses to take for their programs nor how to get recommendations. In addition, some courses have prerequisites requiring students to take courses in sequence with some courses offered

in certain semesters (Laghari, 2014). Some students do not know the requirements of their program or courses while others just want to take courses that are interesting to them (Maphosa, Doorsamy, & Paul, 2020).

Chatbots are already being used in many ways to help students with the advising process. The MOOCBuddy chatbot was created to help students find the best courses for their area of interest and needs (Holotescu, 2016). Even in the development stages with limited capabilities, MOOCBuddy received mostly positive feedback implying that students could use help in identifying the best courses to take. Course Recommender, another e-advising tool, was developed using state-of-the-art data mining techniques and conversational recommenders (Guruge, Kadel, & Halder, 2021).

Chatbots offer students a means of providing personalized advice on selecting the best courses for their degrees and possibly suggestions of courses that might be of greater interest (Suschevskiy & Khalil, 2021). Artificial intelligence (AI) such as machine learning (ML) and natural language process (NLP) are part of the innerworkings behind chatbots helping to provide personalized course and information support (Yu et al., 2017).

3. BACKGROUND IN AI, ML, AND NLP

Artificial Intelligence (AI) machines mimic the human brain's functionality as to how they think and perceive. The ideal AI machine may have its own knowledge base and may be able to operate on its own with very little, if any, human support. A powerful chatbot needs to utilize AI with as little human support as possible since the fundamental purpose of a chatbot is to automate tasks and queries.

Machine Learning (ML) is a technique of having a program look for, find, and determine what patterns exist in a dataset and then learn from the patterns. To do this, a machine is given training samples from which to extract data and learn. For example, a camera learning to detect whether subjects are smiling would analyze many samples of photos of subjects and find patterns in their features. After this, the machine would be given more training samples to test what the "machine" had learned. ML is critical in the development, improvement, and usability of a chatbot (Hiremath et al., 2018).

Natural Language Processing (NLP) is used to make machines "understand" how humans naturally communicate. This is critical for a chatbot as the users are free to type in content using any arbitrary language style. This means that a user can use their own grammar/dialect and still be understood by the chatbot. Thus, a user could ask the same question in multiple ways and get the same results. For instance, a user trying to learn about an advising form could ask a chatbot having NLP "Where's the advising form?" or "How do I find the advising form?"

In 2018, Hiremath et al. found that the use of ML and NLP was critical for improving the usability of chatbots. The authors also concluded that users were more satisfied with chatbots' responses when they were detailed, yet concise and avoided unwanted information. Al Muid et. al (2020) found that an unsupervised approach to ML in chatbots in which the chatbot recognized patterns on its own (without interference or limitations) was more successful, especially in the field of education. In addition, Debnath & Agarwal (2019) emphasized the importance of having a mobile-accessible interface to be critical in user acceptance. The next section describes the impetus behind, and development of, an AI driven advising chatbot developed at the authors' institution for their graduate program.

4. AUTHORS' INSTITUTION

The Master of Science in Computer Science (MSCS) program at the authors' institution was started in 2018 with 30 students. Since that time, the number of students in the program has exploded to 400 and continues to receive substantial active interest from both domestic and international students. This rapid growth created multiple pressure points, especially since computer science faculty are being asked to advise upwards of 150 students while carrying teaching loads of four and five courses each semester.

4.1 Advising Process

The advising process for the MSCS program began by replicating the standard advising practices implemented across campus. General program and course information was posted to the University's web site. Program specific advising information was shared by the program coordinator via individual emails to each prospective and current student. The process gradually morphed into intermittent email broadcasts to regularly updated group mailing lists.

The university's website, the "starting point" of the academic journey for many students, provides a broad spectrum-view of the various university program offerings without addressing the special needs of high-growth programs like the MSCS. Hence, it is difficult, especially for first-time users, to find appropriate program details from the existing pages. The site lacks navigational features (e.g., actionable buttons/widgets, etc.) that can systematically guide the students through the various aspects of their academic journey (e.g., selecting a concentration, developing a degree path, choosing classes, etc.).

Although existing tools such as degree audits are available, providing customized on-demand advising help, specific to each student, has been difficult with limited resources. Thus, MSCS students generate an overwhelming number of emails requiring additional faculty time and effort for personally crafted responses. Because the advisors' time is limited and they are not always available to address advising needs, a backlog of emails often accrues waiting future attention.

4.2. Advising Issues

Common student advising questions usually focus on the courses students need to take for their program of study, the prerequisite requirements for their courses, frequency of course offerings, the modalities in which the courses are offered, graduation requirements, thesis expectations, and internship requirements. In addition, international students often have questions related to maintaining their I-9 visa status.

The advising process, in an ideal world, would be simple (Figure 1). The student would be able to attend an advising session, fill out their advising form correctly, have the form approved by the advisor, and then sign up for their courses without any errors or issues. Unfortunately, the actual advising process usually has additional steps, as many students struggle with the advising form and other areas of advisement.

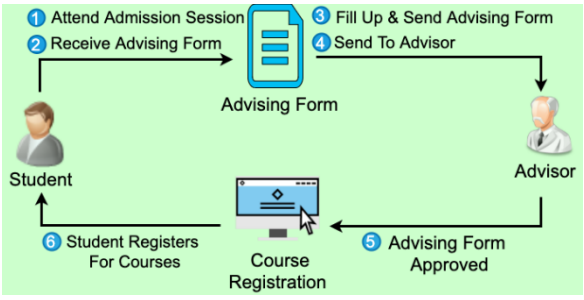


Figure 1: Ideal course registration process

Figure 2 illustrates the difficulties faced by students and advisors during the course registration process. Each exchange of the advising form between student and advisor adds additional layers of communication wasting time for both. The two major periods of this hyper-active advising process can be categorized into: (a) a 2-month window around the start of the semester for enrolling new students, and (b) a 1-month window in the middle of the semester when current students need advising for the next semester's courses. Hence, this activity essentially takes up most of the 4-month semester with brief periods of dormancy.

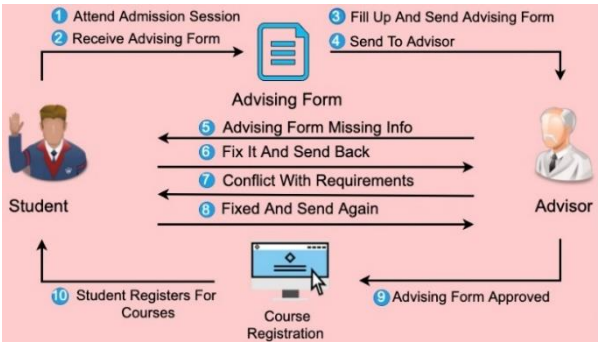


Figure 2: Actual course registration process

4.3 Steps Taken to Address Advising Process Issues

The authors used a multi-step approach to lower the cost of student advisement support while helping the students assimilate into the academic program. *First*, the authors approached the university's undergraduate advising center to learn about their best advising practices. *Second*, several bite-sized (8-10 minutes) carefully crafted videos were recorded describing the various steps of the advising process. A content-rich curated playlist was created by combining MSCS advisor created videos with other videos from the university's collection. Advisors shared the playlist with all incoming and current graduate students. This playlist has proven to be a popular student resource with, in less than a year's time, each of the videos garnering between 600-1000 views on the MSCS YouTube channel.

Third, students' questions and comments from advisor emails, phone calls, and office hour meetings were compiled. Learning Analytics (LA) was used to determine the most frequently asked advising questions. The questions and requests were reorganized into meaningful categories to provide a planned approach capturing the complexities of student-advisor interactions. Considerable effort was made to specify goals and

define potential next steps in the advisement process.

The table of questions, as well as the insights gained from advising center discussions, were then used to design a highly structured interactive web environment. (See Table 2 in Appendix.) A Jekyll-based content management system, hosted free of charge on GitHub, was used to publish bite-sized informational posts on advisement. In addition, a chatbot, named after the University's mascot, was developed using IBM Watson Assist to assist in answering frequently asked questions and encourage a rapport with the students.

5. DEVELOPMENT OF AN ADVISING CHABOT

The chatbot gives students an efficient and informed advising experience during course registration at the beginning of each semester. Using IBM Watson Assistant as a backend, the authors were able to integrate a chatbot into both a website and Slack. The chatbot is currently designed for a limited student population of Computer Science and Cybersecurity (CS/CY) masters students but can be expanded to include all university departments in the future.

5.1. Tree Design

The landing page for the web-based advising tool begins with a "topic tree" that visually represents the interrelationship of critical categories for the main advising tasks. Each category in the topic tree is associated with branches that can be expanded into sub-categories essentially representing FAQ items or pathways for user queries. By pruning or adding new branches, the tree structure systematically customizes to the complex knowledge landscape of student advisement. For example, the advisors found that many recently admitted students had queries about pre-requisite waivers. The sub-tree was modified to include the list of pre-req courses, waiver options, and general messaging explaining the reasoning behind prerequisite requirements.

The topic tree morphed into a "conversation tree," the backbone for conventional *digital conversations* in the automated chatbot. It is essentially a collection of narrative pathways in which the user creates a personalized dialog flow through the choices to meet the desired advising outcomes. Complex conversational flows can be addressed by placing strategic links. Well-designed chatbot dialog is delivered through planned recommendations that quickly resolve user queries in a matter of few seconds. In contrast, bad dialog design can lead to unnecessary user aggravation and ultimately, poor tool adoption.

5.2. Website Layout

The website caters to three distinct student roles: prospective, newly admitted, and current students. A vertical navigation bar and an accordion view limits the views to only those posts associated with the selected role. For example, online-only domestic students are concerned about exam modality, whereas international students are curious about visa issues related to internships. (See Table 2 in the Appendix.) This planned web layout (Figure 3) is a space for implementing navigational tools to guide students through various advisement goals and associated recommendations. The student can independently peruse through the webpages, or the advisor can recommend a specific page during the advising appointment.

The website is built using Jekyll, a static website generator with built-in support for GitHub. GitHub Render hosts the pages free-of-charge. Assumedly, the site should remain secure provided the users' credentials remain confidential. In addition, GitHub offers their own SSL certificates that require minimal effort in providing website security. The site uses markdown and HTML files with customized layouts for MSCS program-specific content.

The vertical accordion view of stacked header list items was chosen for housing the content, as it is the established workhorse for responsive web design. This pattern facilitates progressive disclosure by displaying the critical broad categories from a carefully designed topic tree while the rest of information is just a click away. The focused approach of the accordion view allows users to choose between skimming the topics or revealing details only when necessary. This approach makes digesting the web page content less daunting. This is especially helpful for Gen-Z prospective students with limited attention spans. This population is shifting towards independently exploring education content with minimal human contact, especially in light of the aftermath of the pandemic.

5.3. IBM Watson Assistant

IBM Watson Assistant uses ML and NLP in order to provide real-time responses to text-based questions, essentially FAQ questions containing custom-crafted content appropriate to the students' level of academic maturity. This

programmable tool is extremely effective in handling hundreds of simultaneous student conversations in diverse topics (IBM, 2022). For programming the Watson chatbot, the developer needs to create *entities, intents, and dialogs*. (See Table 1 in the Appendix for a summary of terms). Relevant examples of multiple possible user questions are listed beneath each intent. The chatbot learns from these examples and infers potential user queries which may not be explicitly specified in the intent.

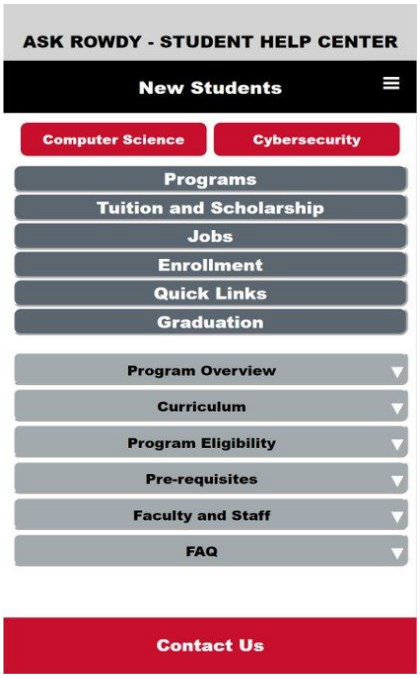


Figure 3: Website Layout

An entity is specifically tied closely to a time, item, or name. Intents and entities are linked to dialogs which are the responses that will be given back to the user. Once the assistant recognizes an intent or entity, a dialog associated with that intent/entity is returned. Sometimes there is a hierarchy of dialogs. Under such cases, the chatbot will send a message to the user seeking more information to clarify the user request. For example, when a user asks for the core courses for their degree, the chatbot must discover which specialization the user is pursuing (e.g., CS/CY). The program names and their corresponding courses are under one main dialog (e.g., 'core course') with each having a separate dialog for their own specialization. Both the Slack and webpage integrations use the same hierarchy of intents, entities, and dialogs in Watson Assistant to obtain their responses. The infrastructure of the advising chatbot is shown in Figure 4.

We used the Slack app to create separate virtual group chats, called a "channel," which are dedicated to individual query. For example, there might be a channel for students to network with each other, a channel for users to message the administrators about an error, or even a channel for just the administrators to converse about an issue. Individuals would have to be invited to join the channel before they would be able to see or contribute content.

This allows for fostering student conversations and reduces information overflow. The MSCS student workspace containing several such channels was created during the pilot study. (Figures 5 and 6 show some of these features.) The next section describes the results of the pilot study survey.

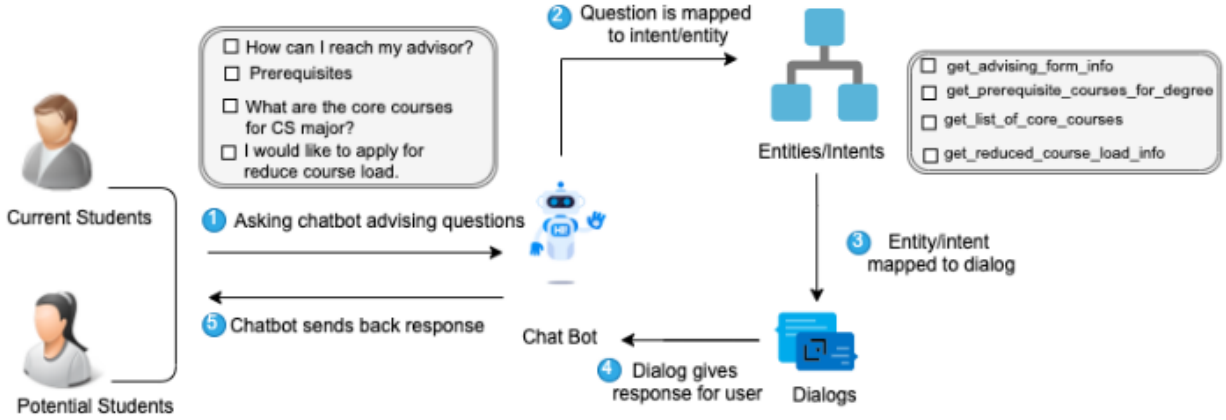


Figure 4: Infrastructure of Ask Rowdy Chatbot

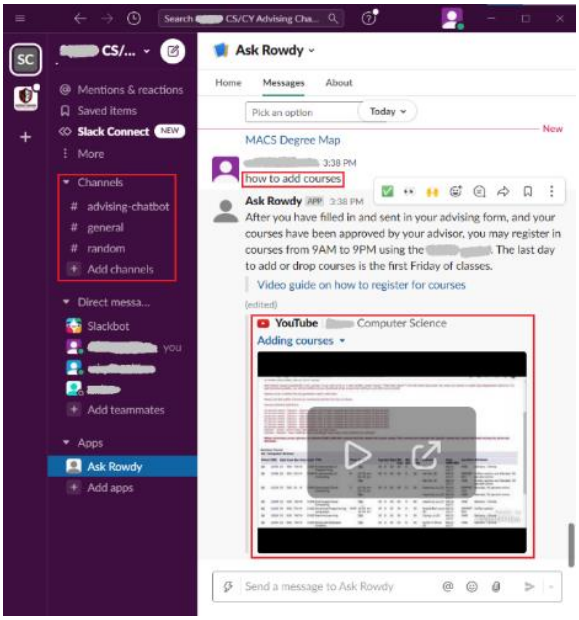


Figure 5: Slack integration

6. RESULTS AND ANALYSIS

A pilot study was conducted to gather student opinions on user experience, chatbot features, and chatbot responses for standard advising concerns. Most students found the chatbot to provide an overall positive experience and was useful for meeting advisement needs. However, further work will need to be conducted to continue to improve on the current functionality.

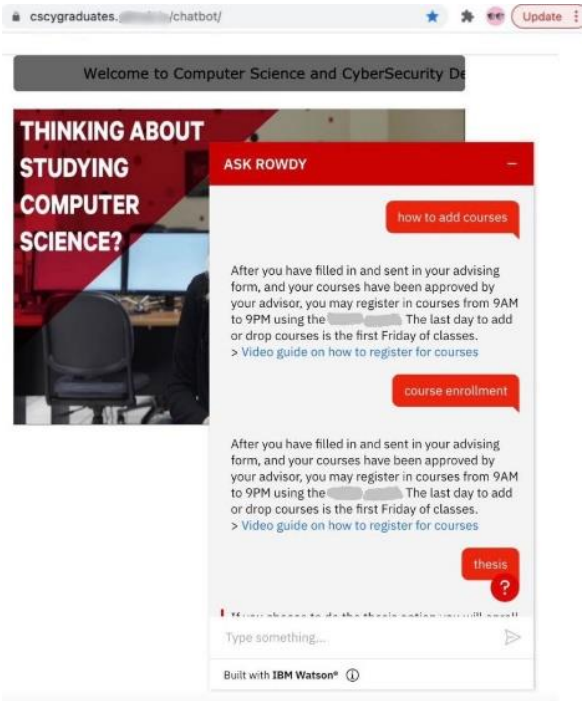


Figure 6: Webchat integration

6.1 Pilot Study

During the pilot study, students were asked to complete a survey form containing questions about usability, quality, functionality, speed, and design of both the website and Slack platforms. The study was targeted toward current and prospective CS/CY masters' students to limit the study to those who would benefit from the project. Out of a total population of 500 students, 99 students responded to the survey representing both newly admitted and current students coming from both domestic and international backgrounds. Students were asked to try out both chatbot interfaces and to answer a Microsoft Forms survey upon completion. Data was recorded anonymously to alleviate privacy concerns. The survey took less than three minutes for the students to complete. The sample questions are discussed in Section 6.2 and the feedback results are presented in Section 6.3.

6.2 Questionnaire for Chatbot Feedback

Questions 1 through 3 in the survey form were demographic questions such as length of study, part or full-time student, and domestic or international student. These questions were aimed at better understanding the student profile and reflected on the level of their experience with the university's advising procedures. The next set of questions asked the users about the chatbot's performance, usability, and responses. We used a 3-level Likert scale of satisfaction. Users were asked about the chatbot's response to see if they believed it was engaging, informative, and friendly.

The next question asked the users if the chatbot was able to accurately answer their questions about a variety of advising questions. The last question allowed the users to rank the chatbot out of 5 stars. The responses could be used to determine where the chatbot was working well and where potential holes in the program needed to be fixed. The rankings indicated whether the users would want to use the chatbot in the future (Figure 7).

6.3 Discussion on Pilot Study Feedback

Respondent feedback from the pilot study was mostly positive (Figure 8). The IBM Watson chatbot back-end supported both the web and Slack app versions. Students' preference for chatbot access was primarily based upon convenience. However, the web interface was used more often by international students.

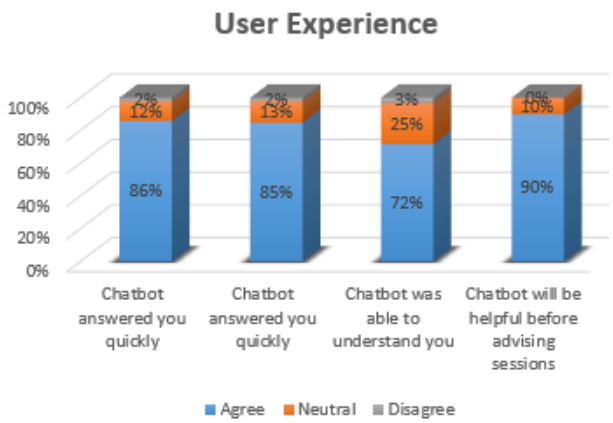


Figure 7: User Experience with Ask Rowdy

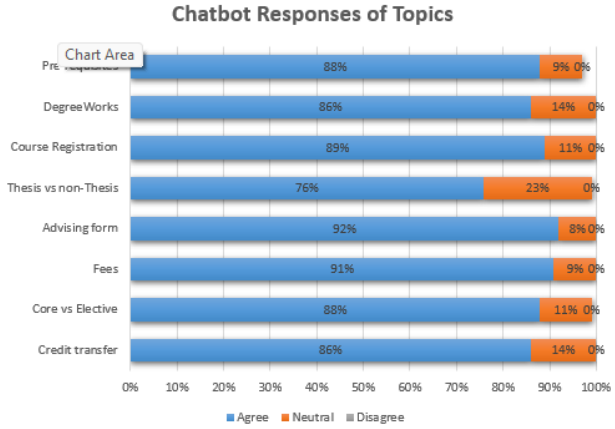


Figure 8: Chatbot Response Rating

Respondents were asked to rank the chatbot out of five stars. The overall rating for the chatbot was a 4.5 out of 5 stars. Most users were able to use the chatbot effortlessly and found its answers to be helpful in their academic advising experience. The respondents each had different academic maturity levels leading the researchers to conclude that all levels of students could potentially find the chatbot a reliable tool for their academic advising.

The students' academic maturity levels were also used in the chatbot to customize the question flow and provide suggestions to students. Overall, the chatbot was able to automatically broadcast solutions to main questions and concerns. Each question asked by the students improved the NLP in the chatbot and was stored for record. This can be later analyzed to identify gaps in terms of question content and information flow. Learning analytics data will be very important in the continued development of the chatbot. As an ongoing project, student preferences and priorities are continuously being incorporated into the workflow. This pilot study is a precursor to a large-scale usability study with a few hundred students and eventually, a department-wide project rollout so that this tool can be used actively by faculty advisors and students. More importantly, it is becoming a platform to present program-specific advisement content to a large diverse student population who can peruse the material at their own pace in a non-linear manner from remote/domestic locations.

7. CONCLUSION

Academic advising communication modes are undergoing tremendous change due to the diversity of student profiles as well as the variety of department-level program offerings. As high-profile programs continue to grow without corresponding resource allocations, alternative means for addressing student advisement questions must be found.

In this paper, the authors described the development of an AI-backed advising chatbot created using IBM Watson Assistant to help alleviate stress caused by explosive growth in an already popular academic program. This web-based tool provides compelling goals and clear academic pathways for advising related tasks. Advising information was provided on a readily available and easily accessible platform that easily conforms to the demanding constraints of cognition, time, and money. Two prototypes of the chatbot were tested on a small pilot study of graduate students, one created as an app and the other accessible via the web. Results from the pilot study found that overall, respondents found the chatbot to be helpful and engaging. Both the web page and Slack app interfaces performed equally well with choice of access relying more on convenience than performance.

Now that the chatbot has been piloted, the authors plan to use the data they have gathered to implement additional functionality and broaden the scope of MSCS advising questions. In the future, the authors plan to expand the reach of the chatbot to assist with advising needs in other university departments and programs.

8. REFERENCES

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2, https://doi.org/10.1016/j.mlwa.2020.100006.

Al Muid, M. A., Reza, M. M., Kalim, R. B., Ahmed, N., Habib, M. T., & Rahman, M. S. (2020).

EduBot: An unsupervised domain-specific chatbot for educational institutions. In Masrour, T., El Hassani, I., Cherrafi, A. (eds) *Artificial Intelligence and Industrial Applications. A2IA 2020. Lecture Notes in Networks and Systems*, 144, 166–174 Springer, Cham. https://doi.org/10.1007/978-3-030-53970-2_16 .

de Lange, P., Bengtson, L., Neumann, A.T., & Klamma, R. (2021). Blockchain-based data verification and consent management for trusted learning analytics using mentoring chatbots. In *Proceedings 11th International Conference on Learning Analytics & Knowledge,* 88–90.

Debnath, B. & Agarwal, A. (2019). A framework to implement AI-integrated chatbot in educational institutes. *Journal of Student* Research. https://doi.org/10.47611/jsr.vi.1063

Dibitonoto, M., Leszczynska, K., Tazzi, F., & Medaglia, C. M. (2018). Chatbot in a campus environment: Design of LiSA, a virtual assistant to help students in their university life. In *International Conference on Human-Computer Interaction*. Springer, 103–116.

Dyke, G., Adamson, D., Howley, I., & Rose, C. P. (2013). Enhancing scientific reasoning and discussion with conversational agents. *IEEE Transactions on Learning Technologies*, *6*(3), 240–247.

Ghose, S., & Barua, J. J. (2013). Toward the implementation of a topic specific dialogue based natural language chatbot as an undergraduate advisor. 2013 *International Conference on Informatics, Electronics & Vision (ICIEV)*, Dhaka, 17-18 May 2013. https://doi.org/10.1109/ICIEV.2013.6572650 .

Goel, A. (2020). AI-powered learning: Making education accessible, affordable, and achievable. Retrieved June 5, 2022 from https://arxiv.org/abs/2006.01908

Guruge, D. B., Kadel, R., & Halder, S. J. (2021). The state of the art in methodologies of course recommender systems—a review of recent research. *Data*, *6*(2), 18.

Heller, B., Proctor, M., Mah, D., Jewell, L., & Cheung, B. (2005). Freudbot: An investigation of chatbot technology in distance education. In *EdMedia+ Innovate Learning. Association for the Advancement of Computing in Education (AACE)*, 3913–3918.

Hiremath, G., Hajare, A., Bhosale, P., Nanaware, R. & Wagh. K.S. (2018). Chatbot for education system. *International Journal of Advance Research, Ideas and Innovations in Technology, 4*(3), 37–43.

Ho, C. C., Lee, H. L., Lo, W. K., & Lui., K. F. A. (2018) Developing a chatbot for college student programme advisement, *2018 International Symposium on Educational Technology (ISET),* 52-56.

Holotescu, C. (2016). MOOCBuddy: A chatbot for personalized learning with MOOCs. In *Proceedings of the International Conference on Human-Computer Interaction – RoCHI* 2016[C], 91–94.

Ibm.com. (2022). How to build a chatbot - IBM Watson Assistant - Docs resources. [online] Retrieved June 13, 2022, from https://www.ibm.com/products/watson-assistant/docs-resources

Jiang, W., Pardos, Z. A. & Wei, Q. (2019). Goal-based course recommendation. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge,* 36–45.

Kumar, R. & Rose, C.P. (2010). Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies*, 4(1), 21–34.

Laghari, M. S. (2014). Automated course advising system. *International Journal of Machine Learning and Computing, 4*(1), 47.

Maphosa, M., Doorsamy, W., & Paul, B. (2020). A review of recommender systems for choosing elective courses. *International Journal of Advanced Computer Science and Applications*, 11(9), 287 - 295.

Nittaya, M., Natakorn, T., Thapani, H., Jakkarin, Y. (2020). The neural network conversation model enables the commonly asked student query agents. *International Journal of Advanced Computer Science and Applications, 11*(4), 155-164.

Nwanko, W., (2018). Interactive advising with bots: Improving academic excellence in educational establishments. *American Journal of Operations Management and Information Systems, 3*(1), 6-21.

Rana, M. (2019). Eaglebot: A chatbot based multi-tier question answering system for retrieving answers from heterogeneous sources using BERT. *Electronic Theses and Dissertations,* 1994. Retrieved June 10, 2022 from

https://digitalcommons.georgiasouthern.edu/cgi/viewcontent.cgi?article=3134&context=etd

Redden, E. (2021). International Enrollments Begin to Recover. *Inside Higher Ed,* Retrieved June 3, 2022 from https://www.insidehighered.com/admissions/article/2021/11/15/international-students-increase-following-pandemic-declines

Sjöström, J., Aghaee, N., Dahlin, M., & Ågerfalk, P. J. (2018). Designing chatbots for higher education practice. In *International Conference on Information Systems Education and Research*. 1–10.

Suschevskiy, V., & Khalil, M. (2021). Creating a course recommendation system for exchange students. In *Proceedings 11th International Conference on Learning Analytics & Knowledge*. 76,78.

Wagner, K., Hilliger, I., Merceron, A., & Sauer, P. (2021). Eliciting students' needs and concerns about a novel course enrollment support system. In *Proceedings 11th International Conference on Learning Analytics & Knowledge*. 294, 304.

Wang, O., Jing, S., Camacho, I., Joyner, D., & Goel, A. (2020). Jill Watson SA: Design and evaluation of a virtual agent to build communities among online learners. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

Yilmaz, F. K., G., & Yilmaz, R. (2020). Student opinions about personalized recommendation and feedback based on learning analytics. *Technology, Knowledge and Learning*, *25*(4), 753–768.

Yu, H., Miao, C., Leung, C., & White, T. J. (2017). Towards AI-powered personalization in MOOC learning. npj *Science of Learning 2*(1), 1–5. https://doi.org/10.1038/s41539-017-0016-3

.

Editor's Note:

This paper was selected for inclusion in the journal as an EDSIGCON 2022 Distinguished Paper. The acceptance rate is typically 7% for this category of paper based on blind reviews from six or more peers including three or more former best papers authors who did not submit a paper in 2022.

.

Appendices and Annexures

	Function	Explanation	Examples
Intent	Takes example user questions and extrapolates it to figure out what user is needing	• Used for majority of questions • Each new example shows the chatbot a different way to ask the same question	#thesis_info #get_course_info #get_advisor_info
Entity	• Specific time/item/thing user query	• Used for a limited number of user interactions • Can be used for semesters and courses • There are also built-in entities like day and time	@is_doing_thesis @queried_GA @queried_RA @queried_semester
Dialog	• Response given to user • Can have multiple responses that go sequentially or together	• Text options used to give users answers and links to find more information on topics • Buttons to suggest other topics under the chatbot the user might find helpful	• Specific Course Information • Advising Form Information • Thesis Questions

Table 1: Explaining Intent, Entity and Dialog

	Simple	Complex
Enrollment/ Registration Questions	1.How do I enroll in classes for the upcoming semester? 2.What is the "Portal"? How do I register for my courses using Portal? 3.Identify course modality: face-to-face or online. 4.How can I register for the course that is full? 5.How many face-to-face courses do I need to take to maintain my visa status?	1. What are the special requirements for international students? 2. When can I withdraw from a course without failing it? 3. If I fail a course and then re-register in the next semester for that course, am I required to attend that course's classes? 4. If I want to register for courses in the summer, can the department offer the courses I have not passed?
Program Questions	1. Which are the core courses for my program? 2. How many 600-level courses do I need to register for my program?	Where can I find a course description before registering for it? (e.g., Assignments, Projects & Presentations).
Advising Questions	1. I cannot attend the advising meeting. How can I register for the upcoming semester? 2. I am unsure about some of my courses. Can I request any changes later? 3. Who is my advisor? 4. What is the best way to communicate with my academic advisor?	1. How should I choose the courses I enroll in? 2. How do I complete my advising form?
Degree Prerequisites	1. What is the purpose of prerequisites? 2. What are the prerequisite courses for my degree?	1. I don't have a Computer Science background in Bachelor's degree, but I have enough work experience. Will the prerequisites be waived for me?
Graduation	1. What are the requirements for International Students to graduate? 2. What is the minimum GPA required to graduate?	1. I am an international student, and this is my final semester. I only have 1 or 2 courses left. Can I register for a reduced course load? 2. What should I do if I only have one course left for my last semester?
Thesis and Non-Thesis	1. How does the thesis process work? 2. What is the difference between thesis and non-thesis for graduation?	1. How many elective courses do I need if I choose the thesis option for graduation?
Other Questions	1. How many online courses can I take as an international student?	1. How many online courses can I take for my last semester as an international student?

Table 2: Sample advising questions based on categories

Aligning Course Assignments to Fulfill IS2020 Competencies

Jonathan P. Leidig
jonathan.leidig@gvsu.edu
School of Computing
Grand Valley State University
Allendale, MI 49401

Abstract

Educators are tasked with continually updating course objectives, content, assignments, and assessment to meet model curriculum guidelines. IS2020 proposes program level outcomes for required and elective areas. Two elective areas in IS2020 are Data and Business Analytics and Data and Information Visualization. IS2020 details 14 program level competencies (organized within knowledge elements and skills) that are then integrated into individual course-level design. This work presents a set of laboratory exercises to fulfill the competencies of both elective areas. The set of exercises have been taught in the classroom over several years and have been refined to evaluate coverage of the 14 program competencies. The exercises begin with step-by-step tutorials that build student capabilities with software. Advanced exercises propose open challenges to solve. These resources provide IS programs with a draft of potential exercises to include in courses and a framework for covering program-level objectives.

Keywords: learning objectives, assignments, data analytics, visualization

1. INTRODUCTION

The IS2020 model curriculum articulates requisite and optional program competencies for Information Systems (IS) programs (Leidig, Salmela, Anderson, Babb, de Villiers, Gardner, Nunamaker, Scholtz, Shankararaman, Sooriamurthi, & Thouin, 2021). Meeting the report's knowledge and skills recommendations is an open challenge for IS programs. Covering a comprehensive set of knowledge elements and skills requires programs to design learning objectives and materials across multiple courses.

Data and information management is one of the six high-level competency realms in IS2020. In this realm, the required data management area is supplemented by two elective areas (data / business analytics and data / information visualization), see Table 1. To fulfill the IS2020 guidelines, a program must fulfill the data management area but is not required to offer the data / business analytics or data / information visualization areas. The two optional areas build upon the requisite data management area's knowledge elements, e.g., data manipulation, database joins, and non-relational models.

Area	Example courses
Data / Info. Management (Required)	Databases, information search and retrieval, and knowledge management
Data / Bus. Analytics (Elective)	Artificial intelligence, business intelligence, data mining, and machine learning
Data / Info. Visualization (Elective)	Information visualization and visual analytics

Table 1: The three areas within the IS2020 data and information management realm.

The data / business analytics area covers a broad set of knowledge elements from a variety of domains including business intelligence, data mining, data science, etc. The knowledge elements and skills from multiple courses must be

coordinated in order for students to recognize and internalize connections between the full set of program competencies and course competencies.

A regional school in the Midwest offers an IS degree that includes an optional track in data analysis. The curriculum for the degree with this track requires four courses in database management, data analytics tools, information visualization, and either AI or data mining. Informal student feedback suggests that the logical connections between topics taught in separate courses is not always obvious to students. As these courses are also utilized by several academic programs and taught by multiple faculty members, additional considerations are required for students to integrate the set of related competencies. Courses in the data analysis track are partly coordinated on several of the required knowledge elements, materials, case studies, software laboratories, and projects. A coordinated series of lab assignments and projects were designed to cover technical competencies while focusing on problem solving in the memorable domain of marine ecology. This domain was found to provide engaging datasets, familiar case studies for students from varied backgrounds, intuitive patterns to recognize, and insights to uncover.

The contribution of this work is an organized set of knowledge elements, skills, datasets, and assignments (concentrated in the data management realm) that has been refined in light of IS2020. Appendices A-C contain proposed exercises and are summarized in Table 2. Appendices A and C have been tested in the classroom over the last four years.

Appendix	Marine Ecology-Based Resources
A	A case study, datasets, and tasks for data analytics of tabular data with Python
B	A case study, datasets, and tasks for machine learning of unstructured multimedia data
C	A step-by-step tutorial for interactive visual analytics for quantitative data with Tableau

Table 2: Overview of the Appendices.

The remaining sections of the paper present an approach to fulfilling competencies in the two elective areas of the IS2020 model. Section 2 covers the seven competencies that comprise the IS2020 elective area of Data / Business Analytics. In the first part, the first four of the seven competencies are discussed in light of tabular data exercises. In the second part, the last three of the seven competencies are discussed in light of semi-structured and unstructured data exercises. Similarly, Section 3 covers the seven competencies that comprise the IS2020 elective area of Data / Information Visualization. In the first part, the first five of the seven competencies are discussed in light of descriptive visualizations. In the second part, the remaining two of seven competencies are discussed in light of interaction and discovery.

2. DATA ANALYTICS

IS2020 details the elective area of Data / Business Analytics within section A3.2.2. Programs that offer content in this optional area build student competencies in handling and analyzing large, diverse datasets. Students must be able to inform business problems with actionable solutions by leveraging large underlying data. Analytics tasks involve techniques covered across several business, computing, math, and statistics courses. In particular, the fields of data mining and machine learning cover topics related to classification, clustering, modeling, prediction, optimization, and recommendation. IS2020 recommends seven competencies to be covered via analytics courses.

In the author's IS program, the seven competencies for the data analytics area are intentionally covered within the four-course track. In these courses, much of the course content is aligned with common textbooks and traditional assignments that teach students to perform common analytical tasks on tabular datasets. As an example, competency #7 requires the use of big data tools on real-world case studies via Hadoop, Spark, and the map-reduce framework. In an information management course, students use map-reduce queries to filter, aggregate, and analyze collections of semi-structured JSON files. Students also use HBase (Hadoop) or Google BigTable to store and retrieve social media data in scalable cloud-based columnar databases. Instead of focusing on course materials that have already been widely adopted, this section focuses on the less traditional datasets and exercises that have been incorporated into the data analytics track to specifically meet the IS2020 Bloom's cognitive level recommendation for the seven data analytics competencies.

Tabular and geo-temporal data exercises

Tabular data is used to cover the first four out of the seven Data / Business Analytics competencies, see Table 3. Competencies #1-4 require students to formulate and perform data-driven analytical tasks. The remaining three

competencies (#5-7) will be covered with unstructured data in the next subsection.

Competency	Description
#1	apply the principles of computational thinking, abstraction, pattern recognition, etc.
#2	analyze data science problems
#3	express business problems as data problems
#4	perform data analysis using descriptive statistics and visualization

Table 3: IS2020 competencies in the elective field of Data / Business Analytics.

These competencies can be evaluated in a pipeline of skills that requires students to perform data analytics tasks with quantitative datasets. Tasks in the pipeline include managing data; performing extract, transform, and load (ETL) tasks; analyzing data; uncovering insights; presenting findings in descriptive charts; and suggesting solutions to problems. These tasks provide students with experiential learning in analyses, interpretations, and recommendations. After exposure to diverse analytical techniques in a variety of pre-requisite courses, a graded activity in an upper-level course is used to challenge students to solve several problems related to marine ecology. Students are given a brief background primer on toxic harmful algal blooms (HABs) and hypoxia (a stratified water column with insufficient oxygen to support fish and other aquatic life). The scientific observations and series of analytical tasks are used to challenge students to recognize HABs in field datasets and identify hypoxic events in the Great Lakes region and Gulf of Mexico.

For these competencies, students are given background information on the marine ecology problem of hypoxia as detailed in Appendix A (Louisiana Universities Marine Consortium, 2018). They are then given scientific observations that inform problem solving efforts. Specific problems are paired with appropriate data sources including governmental warning systems (GLERL, 2022), national buoy sensor networks (NDBC, 2021), merchant ships that report real-time water readings (NDBC, 2022), and academic datasets (GLERL, n.d.). In reviewing the metadata and field descriptions for the datasets, students learn about sensors, standard units of measurement (e.g., microSiemens/cm, mg/L, and NTUs), valid data ranges, calibration challenges for the hardware sensors, maintenance service intervals, biofouling and algae effects on sensors, prevalent dirty data, the constant degradation of oxygen sensor membranes (which affects accuracy), sensor failures, prior data cleaning steps, data provenance, etc. The case study requires students to review data descriptions and footnotes in order to self-evaluate their understanding of the data and its context before performing analysis tasks. Students have to take these factors into consideration as they develop Python scripts to clean observations of nonsensical measurements and perform imputation on missing values. Students then solve analytics tasks by developing Python code to manipulate and integrate cleaned data, produce descriptive statistics, and create charts that focus on a specific problem. Specific tasks are updated each semester and might include:

- Visualize hypoxia events in Lake Erie based on a high-dimension dataset. Create a histogram or similar visualization that details when the dissolved oxygen falls below 2 mg/L. Generate at least one chart for each buoy location, covering one year of data. Consider using matplotlib, ggplot2, or Tableau charts.
- Visualize events in Lake Erie by integrating multiple datasets for buoys and shore-based research stations. Create a geo-spatial map that either uses color or size to display the observations for multiple POI locations. Select one attribute (temperature, DO, turbidity, etc.) and generate charts that display all of the readings from multiple buoys at a similar timestamp. Note: the timestamps might not be aligned for the entire set of locations. Consider using a Python mapping library or Tableau maps.
- Visualize the locations of ships that are reporting scientific observations on Lake Erie. You may use either the current set of active ships or else a historic dataset. Create a geo-spatial map that plots the coordinates of each ship as well as field attributes (air temperature, water temp, air visibility, wind direction, wave height, etc.). Generate a dashboard or map showing all active ships for a given timestamp (e.g., over a one-hour window). Consider using a Python mapping library or Tableau maps.

The results of these tasks are then explored, interpreted, and written up in documentation. Students interpret their results to determine the potential for hypoxia (dissolved oxygen readings under 2 mg/L) or HABs (algal blooms with high readings from chlorophyll and phycocyanin sensors). Students reflect on whether HABs or

hypoxia conditions were identified, identify geo-temporal patterns of detected events, and identify any potential need for governmental responses. Students finish the laboratory exercises by modifying their scripts to best present their findings to decision makers on the respective problem. See Appendix A for further details regarding the publicly available dataset and proposed laboratory exercises.

Competencies #1-4 are successfully evaluated by these guided laboratory assignments that require students to reflect on a new domain, investigate a range of available datasets, perform analytics, and formulate their own data-driven solutions.

Multimedia data exercises

Data / Business Analytics competencies #5-7 require exercises for semi-structured and unstructured datasets. Multimedia data is used to cover these last three of the seven Data / Business Analytics competencies, see Table 4.

Competency	Description
#5	explain the principles of classification, clustering, optimization, and recommendation
#6	articulate the potential of big data given volume, velocity, and variety
#7	demonstrate the use of big data tools on real world case studies

Table 4: IS2020 competencies in the elective field of Data / Business Analytics.

Convolutional neural networks (CNNs) and other machine learning models are now widely used to classify unstructured data such as multimedia. See Appendix B for background information regarding two marine ecology problems that require machine learning algorithms to calculate biodiversity and coral cover (Reef Renewal Bonaire, 2021). A new oceanic dataset comprised of coral images was developed to support these machine learning tasks and education (Leidig, 2022). A sabbatical leave period was utilized to travel to The Bahamas, Bonaire, and Florida to capture and organize coral image collections. The dataset contains millions of underwater images with known classes (coral genus and species labels). These new collections provide large training datasets for 40+ species of coral. See Figure 1 for coral images from the collections.

With these new collections, exercises can now be developed for competencies #5-7. Specifically, laboratory tasks cover techniques for managing massive datasets, manipulating unstructured data, executing parallel code, and training machine learning classification models. The Python with PyTorch framework (Paszke, Gross, Massa, Lerer, Bradbury, Chanan, & Chintala, 2019) is used by students to train CNNs using these datasets. See Appendix B for additional details regarding the publicly available collections and exercises. Competencies #5-7 are met by guided laboratory assignments that require students to access large datasets, classify two or more species using common CNN tools, and evaluate the classification performance.

Figure 1 Four coral classes (boulder brain, great star, smooth flower, and staghorn).

3. DATA VISUALIZATION

IS2020 details the elective area of Data / Information Visualization within section A3.2.3. Programs that offer content in this area build student competencies in manipulating datasets, exploring, and presenting insights. Students must be able to inform business decisions by storytelling with clear narratives built upon underlying data. Visualization tasks involve the techniques covered in computing and statistical domains as well as the fields of art and media, cogitative sciences, psychology, etc. In particular, important topics are related to data integration, manipulation, visual encoding, interaction, human perception, and software tools. IS2020 recommends seven competencies to be covered via visualization courses. This section details datasets and exercises that have been incorporated into the data analytics track in light of IS2020's visualization competencies.

Descriptive visualization exercises

In Data / Information Visualization competencies #1-5, students design descriptive visualizations that guide analysts to insights, see Table 5. Competency #1 requires additional emphasis on human cognition. Students are tasked with intentionally considering their intended insight given a message to convey, data to encode, potential charts, and anticipated audience. In addition to the actual analytical tasks performed, assignments related to competency #1 require students to provide written reflections. For each chart, students perform three additional tasks. 1) Write in a sentence: what is the overarching theme of the chart, title, caption, and insight that

should stick out to the user? E.g., what are the temporal patterns of HAB events that occur in Lake Erie? 2) Design multiple charts with different plot types and ranges of context. Each chart must encode data with a single, clear takeaway. Respond: for consumer-facing charts, were pre-attentive cognition and the Gestalt Laws of Psychology utilized to make the takeaway apparent to the chart consumer? Did the chart's title and captions reinforce the insight to guide the consumer in interpreting the chart? 3) Write in a paragraph: Reflect how your visualization follows information visualization theory and principles. Consider discussing: accurate mapping from data to the generated graph; how static infographics show scale, comparison, or patterns; ease for the user to 'unmap' the graphic and understand the insight; ease for users to forage and find information; high ratios of data per pixel; presence of chart junk; adherence to Cleveland's rules; use of the Gestalt Laws of Grouping; interaction techniques; scalability; and use of overviews and navigation. In addition to generating accurate visualizations, reflection forces students to focus their attention on human perception and competency #1.

The scientific datasets and related set of Python-based analytical exercises (described above in the data / business analytics competencies #1-4 subsection) are also used to evaluate data / information visualization competencies #2-5.

Competency	Description
#1	understand human perception
#2	effectively display quantitative datasets
#3	present analyzed datasets and insights using visualization theory
#4	develop scripts to manipulate datasets
#5	design and generate visualizations

Table 5: IS2020 competencies in the elective field of data visualization.

Discovery visualization exercises

Within the seven data / information visualization competencies, #6 and 7 require students to design interactive visualizations that allow consumers to uncover insights and form compelling, data-driven narratives, see Table 6.

Due to the high learning curve of web-development using D3.js, laboratory exercises leverage the intuitive Tableau software package, see Appendix C. With hundreds of thousands of fish surveys dating back several decades, the Reef Environmental Education Foundation (REEF) provides quantitative datasets that track hundreds of fish species on specific reefs across the globe (REEF, 2022). The geo-temporal organization of multi-dimension survey reports allows for hierarchical overviews that group reefs into larger combined reef tracks, countries, and world regions as well as hierarchical time components (month, season, year, and decade). The fish species themselves are also organized by hierarchical taxa (e.g., genus and species). The datasets lend themselves to exploration of the relations between fish species and geo-temporal distribution patterns.

Competency	Description
#6	explore datasets using interactive visual analytics
#7	express compelling discoveries and insights

Table 6: IS2020 competencies in the elective field of data visualization.

To evaluate competencies #6 and 7, students complete a series of step-by-step laboratory exercises that show students how to conduct visual analytics using Tableau. Students are given REEF data monitoring 547 fish species through 26,737 survey reports (dating back to 1993) for the 168 coral reef locations surrounding the country of Bonaire. Tutorials guide the students to learn how to import data, join and integrate datasets, filter observations, design charts, and compose interactive dashboards. The dashboards then serve as a basis for reasoning with descriptive statistics, probability, apparent paradoxes, and pattern interpretation. Students are prompted to answer questions by exploring the dataset using the interactive charts they design. Relevant views of the dataset are identified and interpreted in narratives. As examples, students identify differences in the survey reports to comment on data validation issues when comparing expert surveyors to novice surveyors. Fish distributions inform questions on probabilistic encounters with various wildlife. At least two laboratory exercises are needed for students to become familiar with the software as well as engage in visual analytics.

See Appendix C for further details regarding the publicly available REEF survey datasets and related laboratory exercises. Pairing this high-dimensional dataset with Tableau supports competencies #6 and 7. Specifically, several laboratory reports evaluate skills, hypothesis

generation, interactive visualizations, narratives, visual exploration, and visual explanation.

4. SUMMARY

The IS2020 guidelines led IS programs to consider the extent to which they cover the 21 competencies in the Data / Information Realm. This paper contributes a coordinated set of datasets and exercises used to teach and evaluate knowledge elements and skills. Adopting these materials (or curating alternative collections) would benefit other IS programs aiming to cover the 14 elective Data / Business Analytics and Data / Information Visualization competencies. These data-driven exercises require students to synthetize knowledge elements, skills, and analytical techniques taught across multiple courses in order to demonstrate program competencies in their upper-level courses.

The datasets and exercises presented here have been leveraged over the last four years to provide and evaluate coverage of IS2020's 14 elective competencies. The exercises in Appendix A have been utilized over five semesters to introduce students to data analytics in Python. The exercises in Appendix B have only been tested so far by five graduate students – who individually desired to learn about deep learning with CNNs and Python. These graduate students were able to produce classification models with ~96% accuracy, depending on the coral species being recognized. As the coral classification training dataset was produced in 2022, another year is needed to finalize and evaluate step-by-step lab exercises at a level appropriate for undergraduate students. Appendix C exercises have been utilized in 10 sections of visualization courses over the last four years. Students commonly express their appreciation for the step-by-step tutorials that introduce them to Tableau and performing interactive visual analytics.

This set of resources supports all 14 competencies in the IS2020 elective areas. The exercises were designed to provide IS programs with memorable case studies, provide intuitive problems, and expose students to problem solving within a new domain. Marine ecology was selected as the targeted domain as existing quantitative datasets were well described (with sufficient metadata) and minimal background knowledge was required to begin problem solving (e.g., brief overviews of biodiversity, HABs, hypoxia, and the use of reef photomosaics were sufficient coverage of the domain). The challenges of identifying HABs, hypoxia, coral species, and coral reef growth are familiar, engaging, attention grabbing, and memorable for students regardless of personal or academic backgrounds. The real-world coral case study is an active, open challenge for coral restoration groups that monitor the degradation of coral reefs. Concentrating analyses in a single applied domain (across multiple IS courses) provides a structure for students to recall, reflect, and integrate previous laboratory experiences.

5. REFERENCES

GLERL. (2022). *Experimental Lake Erie Hypoxia Forecast*. NOAA - Great Lakes Environmental Research Laboratory. https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/hypoxiaWarningSystem.html

GLERL. (n.d.). *International Field Years on Lake Erie (IFYLE)*. NOAA - Great Lakes Environmental Research Laboratory. https://www.glerl.noaa.gov/res/projects/ifyle

Leidig, J. (2022). *Coral Reef Image Collections for Machine Learning, Mapping, and Monitoring*. In Proceedings of the 33rd IEEE OCEANS Conference 2022.

Leidig, P., Salmela H., Anderson G., Babb J., de Villiers C., Gardner L., Nunamaker J.F., Scholtz B., Shankararaman V., Sooriamurthi R., & Thouin M. (2021). *IS2020 A Competency Model for Undergraduate Programs in Information Systems*. Association for Computing Machinery and Association of Information Systems. DOI: 10.1145/3460863.

Louisiana Universities Marine Consortium. (2018). *About Hypoxia*. Gulf of Mexico Hypoxia. https://gulfhypoxia.net/about-hypoxia

NDBC. (2021, August). *National Data Buoy Center*. NOAA. https://www.ndbc.noaa.gov

NDBC. (2022, March). *Ship Observations Report*. NOAA. www.ndbc.noaa.gov/ship_obs.php

Paszke, A., Gross S., Massa F., Lerer A., Bradbury J., Chanan G., and Chintala S. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library.* In Advances in Neural Information Processing Systems 32 (pp. 8024–8035). Curran Associates, Inc. Retrieved from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

REEF. (2022). *Reef Environmental Education Foundation Volunteer Fish Survey Project Database*. World Wide Web electronic publication. Retrieved July 15, 2022, from www.REEF.org

Reef Renewal Bonaire. (2021, June 8). *Measuring our Coral Reef Restoration Success*. Reef Renewal Bonaire. Retrieved July 15, 2022, from https://www.reefrenewalbonaire.org/news/measuring-our-restoration-success

APPENDIX A
Tabular and geo-temporal data exercises

Scientific sensors produce large, longitudinal, and semi-structured datasets. Due to the hardware involved, water buoys provide students with challenges to mitigate (dirty data, erroneous readings, missing values, unit conversions, etc.). Students can gain experiences in checking assumptions, skew, outliers, anomalies, descriptive statistics, and clustering. The data also contains geo-temporal patterns to mine.

Content:

- Buoy and ship water sample data. Latent insights include hypoxic events and HABs.

Requirements:

- Less than 5MB of storage space

Recommended tech stack and platform:

- Python environment with common libraries, Tableau (free academic licenses for personal devices and/or cloud access)

Sample Laboratory Exercise

Part 1: Data manipulation (Pandas), analysis (SciPy), and visualization (Matplotlib, NumPy, and Tableau)

Description

SciPy is a set of widely used packages for managing, analyzing, and visualization large-scale content (see https://scipy.org). It consists of libraries for data structures such as DataFrames and analysis (pandas), N-D arrays (NumPy), 2D plotting (Matplotlib), scientific analysis (SciPy), etc. In addition to matplotlib, other popular libraries include *ggplot (any R fans in the class?), plotly, seaborn, pygal, bokeh, geoplotlib, etc.* These packages differ by their customization, expected data input structures, charts available, export formats (e.g., svg), interactivity, dependencies (web integrated, Pyglet OOP interface), etc.

Tableau is a leading software tool for visual analytics and rapidly creating interactive visualizations (see https://www.tableau.com). Download the tool from tableau.com or else login to the GVSU Blade server for Bioinformatics. Students (worldwide) get renewable year-long free licenses (with a valid university email address). If you require brief tutorials (outside of the in-class training) regarding the GUI and its functionality, see https://www.tableau.com/learn/training. Tableau connects with a variety of underlying dataset formats: offline dataset file, an online data server, or a default source (e.g., try out the World Indicators dataset). Note how the columns from the dataset are available on the left pane, you can drag columns to the middle pane to set the x-position y-position color size (which automatically updates the chart graphic), dragging columns to the filter pane generates dynamic filters within the visualization, and the right pane allows you to change the chart type depending on the columns already selected. Users explore their dataset by creating a sheet and pairing various combinations of dimensions and measures. For each combination of columns that you explore, switch the visualization to several options (including Tableau's recommended visualization which is highlighted on the list of possible charts on the right pane). One goal of this lab is to become familiar with Tableau, how to create graphs (sheets), and how to create interactive dashboards.

Learning objectives

- Become familiar with data manipulation and analysis using commonly used programming libraries
- Open and parse a (potentially) large input file

- Apply some transformation/manipulation to the content as needed
- Integrate existing databases and datasets with analytical processes
- Load cleaned data into an appropriate database, staged data, or tool
- Integrate quantitative analysis/modeling with visualization
- Develop visualizations using common libraries and interactive visualizations for exploration and decision making

Step #1 Case Study - Visual Analytics

With a partner (optional), review the following background material on hypoxic environments. Note: this is currently an open, real-world challenge, not a solved problem. You can expect any existing datasets to be less than ideal, dirty, missing values, misaligned with current real-world situation, etc.

Gather background information:

- https://gulfhypoxia.net/about-hypoxia
- http://coastal.ohiodnr.gov/portals/coastal/pdfs/owc/tech/owc_techbull3_Hypoxia.pdf
  Question 1: Define hypoxia. What is the dissolved oxygen concentration threshold that identifies hypoxia? _____
  Question 2: Which species are affected? How are these species affected? _____

Review the animation of hypoxia forecast data from Lake Erie's 2018-2020 seasons:

- https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/hypoxiaWarningSystem.html
- https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/archive/dissolvedoxygen2018.gif
- https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/archive/dissolvedoxygen2019.gif
- https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/archive/dissolvedoxygen2020.gif
  Question 3: Where and when is hypoxia likely to occur in Lake Erie? Either describe the general locations or provide an annotated map. _____

Remote monitoring:

- https://www.ndbc.noaa.gov
  Question 4: Are current NOAA buoys and sensors located in appropriate locations to monitoring and surveillance of hypoxia? If so, which buoy/station IDs are in pertinent locations? _____
- https://www.ndbc.noaa.gov/ship_obs.php (click show observations for the last 12 hours)
  Question 5: Are there shipping lanes through the area that might provide sporadically-sampled monitoring data? Skip this question for now (not graded).

Historic data 2004-2007:

- https://www.glerl.noaa.gov/res/projects/ifyle
- https://www.glerl.noaa.gov/res/projects/ifyle/data.html
- https://www.glerl.noaa.gov/res/projects/ifyle/data/data.mooring.html
- https://www.glerl.noaa.gov/res/projects/ifyle/data/Mooring/data.ysi.html
- https://www.glerl.noaa.gov/res/projects/ifyle/data/Mooring/ysi/eriemap.html
  Review these details on the GLERL collaboration which has attempted to capture and monitor datasets related to hypoxia for the last 15 years in Lake Erie. Select a buoy location that appears to experience hypoxia events throughout the year (e.g., station Y18). Review the metadata and challenges of collecting this type of data.
- https://www.glerl.noaa.gov/res/projects/ifyle/data/Mooring/ysi/ysi_metadata.txt
  Question 6: These datasets contain very useful information (sensor depth, dissolved oxygen concentrations, temperature, turbidity, and chlorophyll) but were not always accurate (e.g., 2007

data). Describe the errors that are known to exist within the dataset. _____

Step #2 Setup a Python, R, or Tableau visualization environment

Connect to the Stratus' Bioinformatics server (Python and Tableau already installed), EOS (already installed), or else install one of R Studio, Python IDE/Interpreter, or Tableau on your laptop. Note: you likely will not have full permission on EOS or the Stratus Blade server to install new software.

Installation Resources

Stratus - https://www.gvsu.edu/hpc/stratus-2.htm
EOS - http://www.cis.gvsu.edu/eosarchitecture-labs
https://www.python.org/about/gettingstarted
https://docs.python.org/3/tutorial
https://www.tableau.com
There is a year-long free license of Tableau Desktop for students:
https://www.tableau.com/community/academic
There is a 14-day free trial of Tableau Desktop: https://www.tableau.com/products/trial

Step #3 Visualization Challenges

Select and complete one of the following challenges using Tableau, Python, or R. Each option will require a dataset to be extracted/scrapped from a website, manipulated, cleaned, and visualized. The exact approach, software, libraries, and solution is left up to the students, in keeping with the spirit of the assignment. For one of the following options, *turn in* a one paragraph writeup containing a description of your methodology/workflow (e.g., data source, how downloaded, how cleaned, and how visualized), a final visualization, and any code/workbooks developed.

Option A

Visualize hypoxia events in Lake Erie based on a high-dimension dataset. Create a histogram or similar visualization that details when the dissolved oxygen falls below 2 mg/L. Generate at least one chart for one location, covering data from the whole year at that location. Consider using matplotlib, ggplot2, or Tableau charts.

Example data:
https://www.glerl.noaa.gov/res/projects/ifyle/data/Mooring/ysi/2007/Y18.txt

where day/time `206.0208` = 30 minutes, between day 205/365 to day 288/365

Option B

Visualize events in Lake Erie by integrating datasets for multiple buoys/stations. Create a geo-spatial map that either uses color or size to display the observations for multiple POI locations. Select one attribute (temperature, DO, turbidity, etc.) and generate a chart that displays all of the readings from multiple buoys at a similar timestamp. Note: the timestamps might not be exactly identical and need to be rounded. Generate at least one chart (for the entire set of locations) at one timestamp. Consider using a python/R mapping library or Tableau maps.

Example data:
https://www.glerl.noaa.gov/res/projects/ifyle/data/Mooring/ysi/eriemap.html

Option C

Visualize the locations of ships that are reporting scientific observations on Lake Erie. You may use either the current, active ships or else a historic dataset. Create a geo-spatial map that plots the coordinates of each ship as well as one attribute (air temperature, water temp, air visibility, wind direction, wave height etc.). Generate at least one chart/map showing all active ships for a given

timestamp (e.g., a one-hour window). Consider using a python/R mapping library or Tableau maps. Example data: https://www.ndbc.noaa.gov/ship_obs.php

Deliverables:

Turn in a Word/text file containing your answers to Step 1 questions and Step 3 text + visualization. Also, include ANY computing code (e.g., R/Python scripts), Tableau workbooks (.twbx file), intermediate datasets (e.g., cleaned CSV files if small in size), etc. Upload a softcopy of these documents.

APPENDIX B
Coral Multimedia Datasets and Exercises

In marine ecology, two common tasks are in estimating species biodiversity and the percent of the seabed that is covered in coral. Thousands of images are captured (looking down at the seabed) and then stitched together to form a single large photomosaic image that covers several square kilometers of the ocean floor. Researchers manually analyze the image to calculate biodiversity and coral cover (Reef Renewal Bonaire, 2021). These two features are calculated by randomly picking pixels from the image and labeling which coral species (if any) was at that location. These human-generated labels are then aggregated to determine biodiversity and coral cover. However, advancements in machine learning, specifically convolutional neural networks (CNNs), promise to semi-automate these human-intensive tasks. Large training datasets of underwater images could be used to develop CNNs that classify the genus and species of a coral found in a top-down view image. With CNNs to classify multiple coral species, the entire photomosaic could be classified at each pixel location - without efforts by human experts.

Additional details on the coral collections, field data collection methodology, data processing, and provenance can be found in the full conference paper documenting the collections release (Leidig, 2022). 5.28 million images of coral organized into 44 separate species collections (see Table 7). Each JPG image is 256x256 pixels with RGB color channels.

Requirements:

- Binary classification exercise: 20GB of storage is required for two species with 1,000 training images each, see example images in Figure 2.
- Multiclass classification exercise: 95GB of storage is required for ten species with 1,000 training images each.
- Full-dataset: 1TB storage is required for all training data over 44 coral species.
- Exercises require access to free software installed on a cloud or local HPC resources (ideally with GPUs). At the time of writing, Google Collaborate, Juypter Notebooks, Python, and PyTorch are available (at no cost) to students and faculty for educational purposes.

Recommended tech stack and platform:

- Python, PyTorch, and computing resources (Google Collab, personal student computers, or lab).

Sequential tasks:

- Practical computing laboratory steps demonstrate how to access and manage the organized multimedia datasets.
- Laboratory steps expose students to CNNs model building u6ing template code, optimizing a small number of hyperparameters.
- Laboratory steps provide background information and libraries for transfer learning with ResNet and similar image-based neural networks.
- Students perform binary classification (distinguishing between two species) with model training and performance evaluation.
- Students are challenged with multi-species classification.

Classes of coral:

	Common Name	Scientific Name	Number of Training Images
1	Staghorn	Acropora cervicornis	1125040
2	Lobed Star	Orbicella annularis	571920
3	Mountainous Star	Orbicella faveolata	461600
4	Boulder Brain	Colpophyllia natans	367720
5	Yellow Pencil	Madracis auretenra	286000
6	Great Star	Montastraea cavernosa	249000
7	Symmetrical Brain	Pseudodiploria strigosa	190200

8	Massive Starlet	Siderastrea siderea	172800
9	Blade Fire	Millepora complanata	168360
10	Whitestar Sheet	Agaricia lamarcki	163800
11	Elliptical Star	Dichocoenia stokesi	136800
12	Smooth Flower	Eusmilia fastigiata	127800
13	Bushy Black	Antipathes caribbeana	123000
14	ClubtipFinger	Porites porites	116760
15	Boulder Star	Orbicella franksi	114600
16	Pillar	Dendrogyra cylindrus	112200
17	Grooved Brain	Diploria labyrinthiformis	107280
18	Maze	Meandrina meandrites	99000
19	Elkhorn	Acropora palmata	90720
20	Mustard Hill	Porites astreoides	63000
21	Spiny Flower	Mussa angulosa	57360
22	Lettuce	Agaricia agaricites	51600
23	Rough Cactus	Mycetophyllia ferox	41400
24	Ridged Fire	Millepora striata	39600
25	Blushing Star	Stephanocoenia intersepta	33600
26	Branching Fire	Millepora alcicornis	31800
27	Thin Leaf Lettuce	Agaricia tenuifolia	27600
28	Six Ray Star	Madracis senaria	20400
29	Wire	Stichopathes luetkeni	19200
30	Ridged Cactus	Mycetophyllia lamarckiana	18000
31	Ten Ray Finger	Madracis carmabi	16200
32	Ten Ray Star	Madracis decactis	13800
33	Rough Star	Isophyllia rigida	9000
34	Lesser Starlet	Siderastrea radians	8640
35	Dimpled Sheet	Agaricia grahamae	7800
36	Knobby Cactus	Mycetophyllia aliciae	7800
37	Butter Print Rose	Meandrina danae	7000
38	Branching Finger	Porites furcata	6600
39	Golfball	Favia fragum	5400
40	Rose Lace	Millepora complanata	4200
41	Knobby Brain	Pseudodiploria clivosa	3480
42	Thin Finger	Porites divaricata	2240

Table 7: Field data collection sizes of ML training datasets (sub-images).

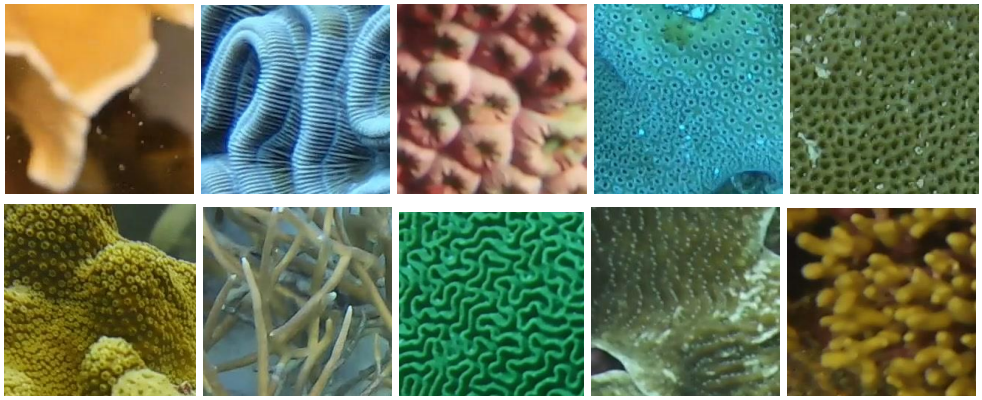


Figure 2 Top 10 classes (top-row) blade fire, boulder brain, great star, lobed star, massive starlet, (bottom-row) mountainous star, staghorn, symmetrical brain, whitestar sheet, and yellow pencil.

APPENDIX C
Fish Survey Datasets and Exercises

Reef Environmental Education Foundation (REEF) collects and aggregates hundreds of thousands of fish surveys from SCUBA-diving volunteers. Additional details on the REEF collections, roving field data collection methodology, data processing, and provenance can be found in (*REEF, 2022)*. These datasets are high-dimensional (100s of columns), with geospatial and temporal components. The many quantitative features, locations, and time components lend themselves to visual exploration in a student-friendly domain.

Requirements:
- Less than 5MB of data, extracted from reef.org

Recommended tech stack and platform:
- Excel and Tableau (Desktop or cloud-based)

Tasks:
- Load data, inner joins, chart generation, filtering, and exploration

Sample Laboratory Exercise

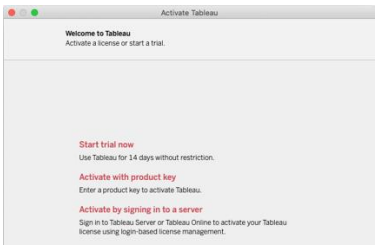
Part 1: An introduction to visual analytics, interactivity, dashboards

Tableau is a leading software tool for visual analytics and rapidly creating interactive visualizations. Students (worldwide) get renewable year-long free licenses (with a valid university email address). If you require brief tutorials (outside of the in-class training) regarding the GUI and its functionality, see https://www.tableau.com/learn/training. The goal of this lab series is to become familiar with Tableau, how to create sheets, exploring data, and how to create interactive dashboards. Lets download the tool **Tableau Desktop** from tableau.com, not the online or public tools.

1. Free License: Visit https://www.tableau.com/academic/students, select the “Free Student License” button, and apply for a free year license. It may take a few hours to a day to receive your key. However, you could likely complete our Tableau labs just using the 14-day free trial and skip the free license step.

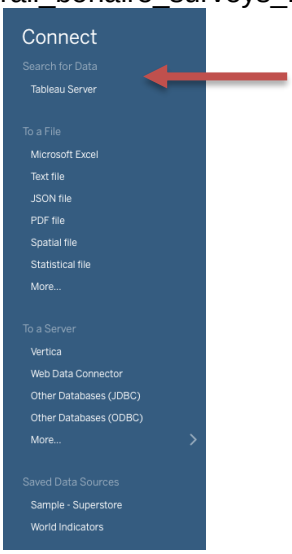
2. Install a 14-day free Tableau trial: Visit https://www.tableau.com/products/trial. You can later enter your registration key later if you haven’t received it yet. Click the “start a free trial” icon and setup an acount or sign into your existing account. Install the software according to the instructions for your Mac or Windows.

3. Start your Tableau software.You can simply use the *start trial now* option until you receive your activation key. Enter in the email you used to register online, if prompted.

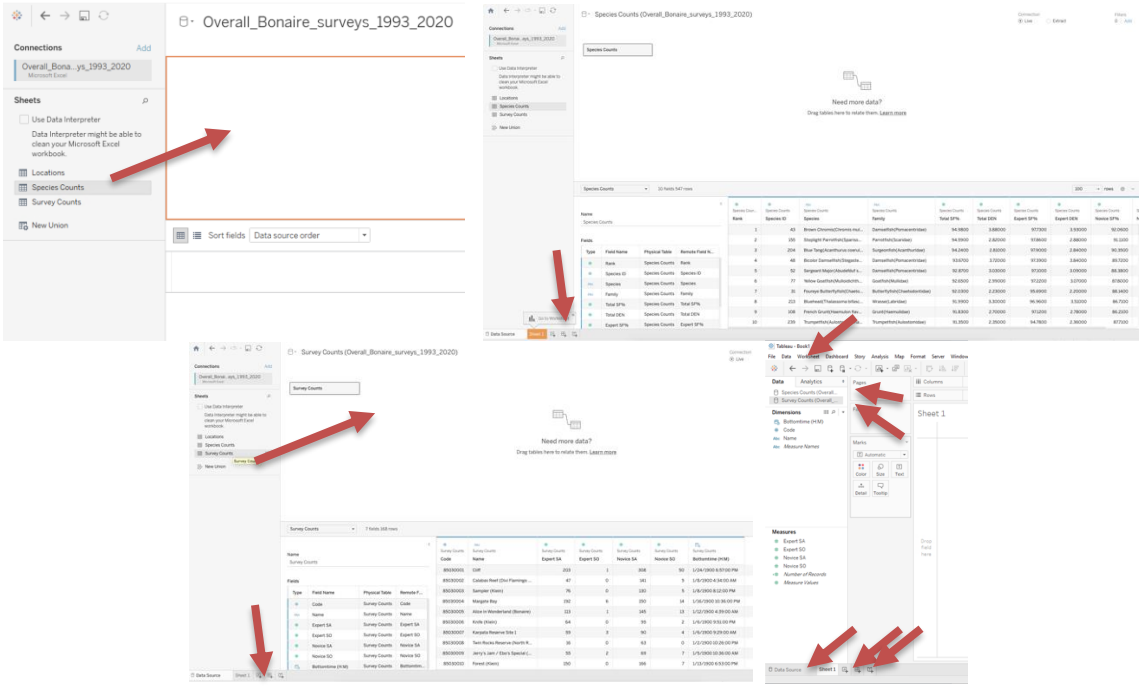


Bonaire is a Carribean island forming part of the Dutch Antilles (with Aruba and Curacao) located a few miles off the coast of Venezula. It is generally well known for its healthy coral reefs, pink salt industry, and tourism based on SCUBA diving and water sports. Let's use Tableau to explore Bonaire's REEF data (www.reef.org).

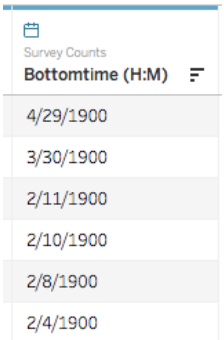
4. Download the files on Blackboard to your computer. You can open the file in Excel to check the data.
5. Import the file in Tableau - a collection of fish surveys by experts and novices. Use the Connect panel on the left to connect to a Microsoft Excel file, find and open your downloaded file: Overall_bonaire_surveys_1993_2020.xlsx. There are two sheets of data for us to import.

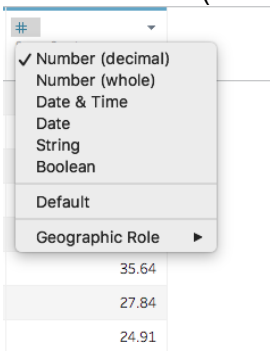


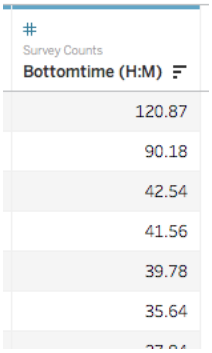
6. We will add the two sheets of data, one at a time. First, select the *Species Counts* Excel worksheet and drag it to the pane on the right. Then click on Sheet 1 on the bottom menu. A *sheet* is the Tableau word for a single chart. Under the Data menu at the top of your screen, select a New Data Source, find your excel file again, and then add the *Survey Counts* worksheet by dragging it to the right pane. Go back to the Sheet 1 chart. After this step, both sheets (*Species Counts* and *Survey Counts*) should be visible in the top left Data menu. Ignore the *Locations* datasheet from the Excel file. The data can always be reviewed in the Data Source tab on the bottom-left corner. Each chart will be created in a new Sheet.



a. How many rows and columns does the Species Counts Data Source contain? Hint: use the Data Source tab.
b. How many rows and columns does the Survey Counts Data Source contain? Hint: use the Data Source tab.
c. When importing the Survey Counts sheet from Excel, note that the variable Bottomtime is a number of hours and minutes (which Tableau mistakes for a date). In the data source pane, let's switch that back to Number(decimal).







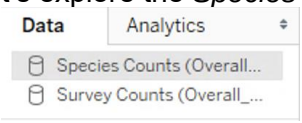
d. For more details, checkout: https://www.reef.org/db/reports/geo?end_date=2020-02-

10&format_type=chart&group_type=species&language=common&region_code=TWA&start_date=1993-01-01&zone_map=0&zones=8503

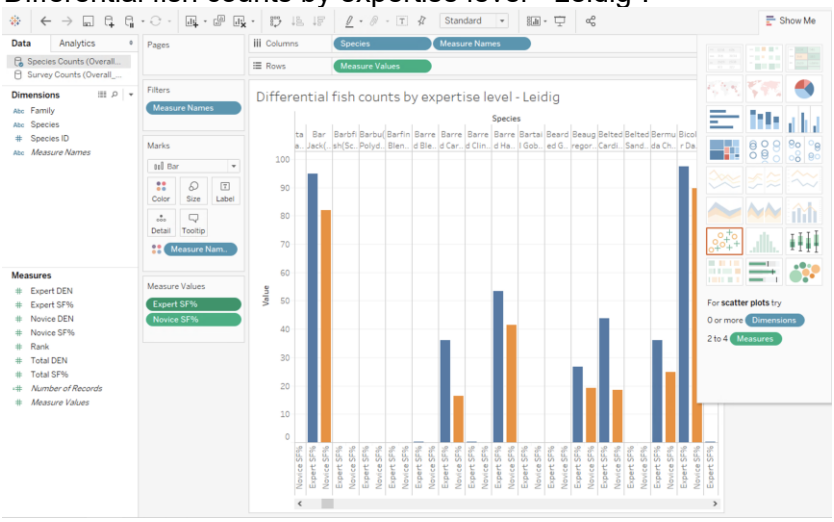
7. Use *Sheet 1*. Note how the columns from the dataset are available on the left pane, you can drag columns to the middle pane to set the x-position/y-position/color/size (which automatically updates the chart graphic), dragging columns to the filter pane generates dynamic filters within the visualization, and the right pane allows you to change the chart type depending on the columns already selected.
8. First, explore your dataset by creating a **sheet** and pairing various combinations of dimensions and measures. For each combination of columns that you explore, switch the visualization to several options (including Tableau's recommended visualization which is highlighted on the list of possible charts on the right pane).

Next use Tableau to explore and answer the following questions. For each question, create a sheet (visualization) that answers the question. Turn in a word/pdf file with a **written answer to the question** and **a screenshot of your sheet** that determined the answer. It is ok if your one screenshot does not fully answer the question; I realize that clicking and scrolling is probably needed within your sheet answer some questions.

First, let's explore the *Species Counts* data.

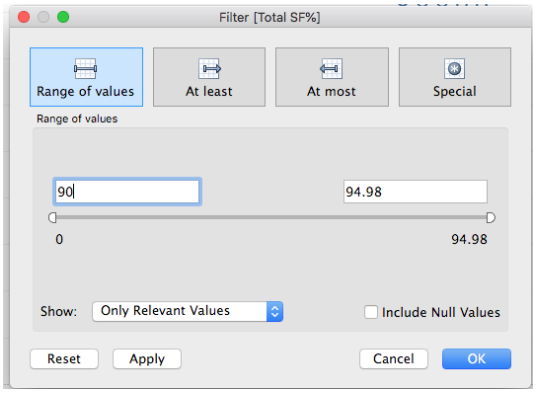


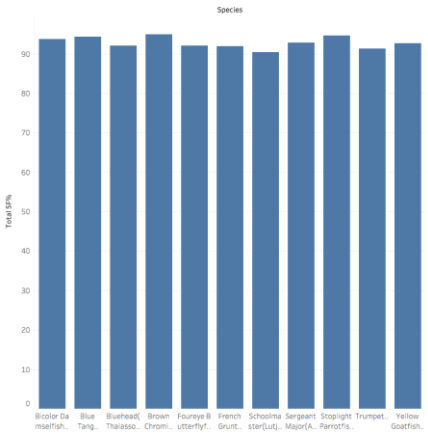
9. Compare *novice sighting frequency* to *expert sighting frequency* for each *species*. **List 5-10 species** that experts seem to identify more frequently than novices. Hint: try a side-by-side bar chart.
    e. Represent the insight in the simplest visualization that gets the story (key takeaway) across. Write a short title that clarifies what the key takeaway is for the reader. **Add your name to the title to get credit for YOUR work**, e.g., "Differential fish counts by expertise level - Leidig".



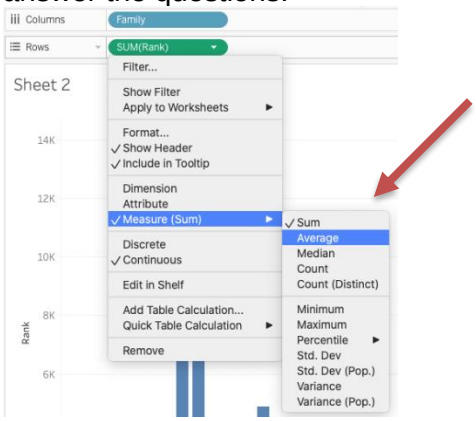
10. Using the same chart, are there any species that *novices* seem to identify more *frequently* than *experts*? **List any**.

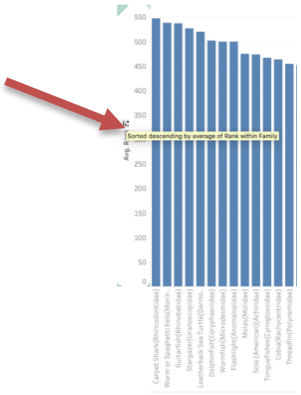
11. Some species are very prevalent and are observed on almost every fish survey. Which species are seen on at least 90% (or more) of total surveys? **Show all of the ~11 species**. Hint: drag the field *Total SF%* to the Filters pane and then only show species found at least 90% of the time.



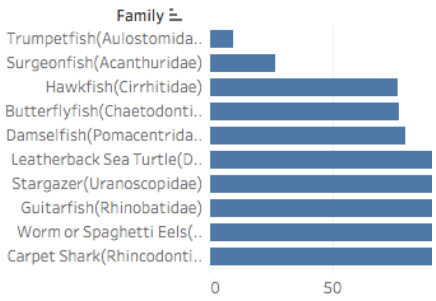


12. Recall from Biology, multiple species are organized into a family of similar organisms (Homo sapiens fall under the Hominidae family). E.g., *Brown Chromis* and *Bicolor Damselfish* are both grouped into the *Damselfish* family. The *species* with a *rank* of 1 is seen most frequently and the *species* with a *rank* of 547 is seen the least frequently.
    f. Generate a **new sheet**/graph that shows the *average* (avg) *rank* of the high-level *family* groups. Hint: you can sort the data within your charts to make it easier to answer the questions.

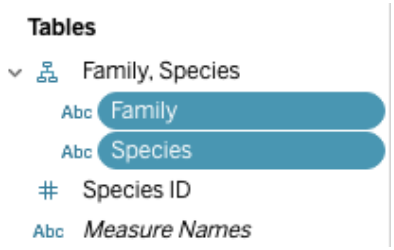




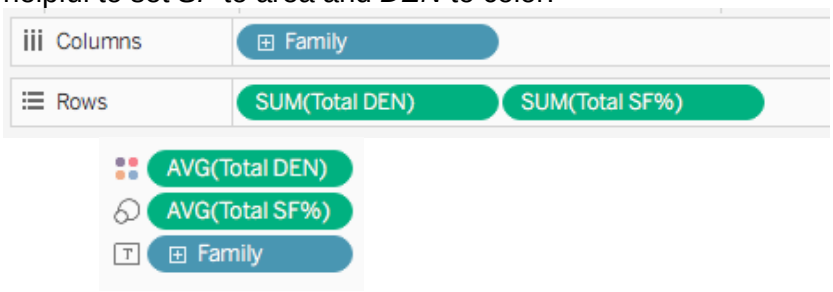
    g. Hint: generate a graph of *family* compared to Avg(*Rank*). Tableau defaults to calculating the Measure (Sum) instead of Avg – but that can be changed in the dropdown from *SUM(Rank)*.
    h. What are the **top 5 most likely families to be observed** - with the lowest Avg(Rank)?
    i. What are the **bottom 5 least likely families to be observed** - with the highest Avg(Rank)?

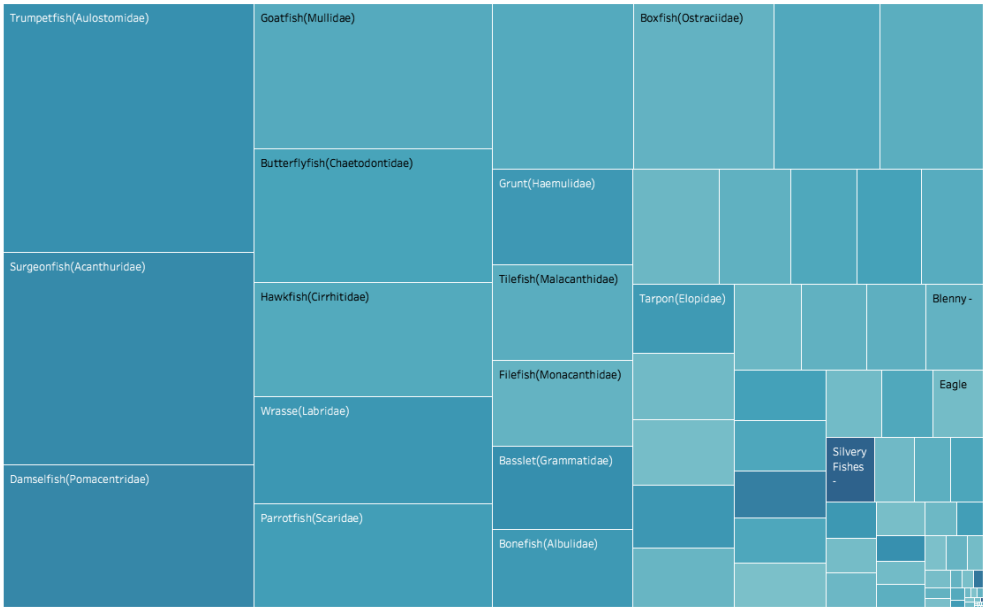


13. Dragging the *Species* dimension on top of the *Family* dimension in the left pane will create a hierarchy. Thus, the *Family* category will be the parent to the *Species* sub-category (Family,Species).



j. Generate a new Treemap sheet showing the average *total sighting frequency* and average *total density* for each *family,species* using size & color. It may be helpful to set *SF* to area and *DEN* to color.

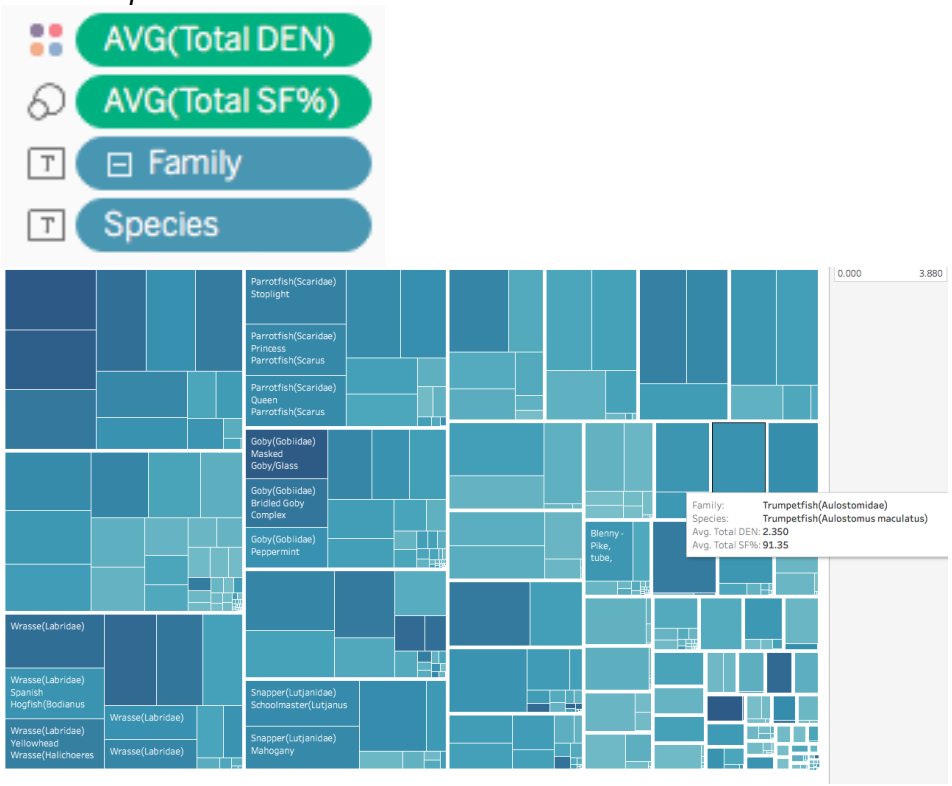




k. Which *families* have the **highest *densities***?

l. Which *families* have the **highest *sighting freq***?
m. **Patterns**: Does *density* seem to be roughly correlated with *sighting frequency*? E.g., *families* that are seen in higher abundance are also reported more frequently?
n. **Outliers**: Are any *families* observed with low frequency but high density? E.g., a *families* that is rarely seen, but when it is seen there are large numbers?
o. **Outliers**: Are any *families* observed with high frequency but low density? E.g., a *families* that is often observed but generally observed all on its alone.
p. Side note: Observe how Trumpetfish have the highest sighting frequency at the *family* level.

14. From the same style chart, click to expand the *family* category to show the breakdown for each *species*.

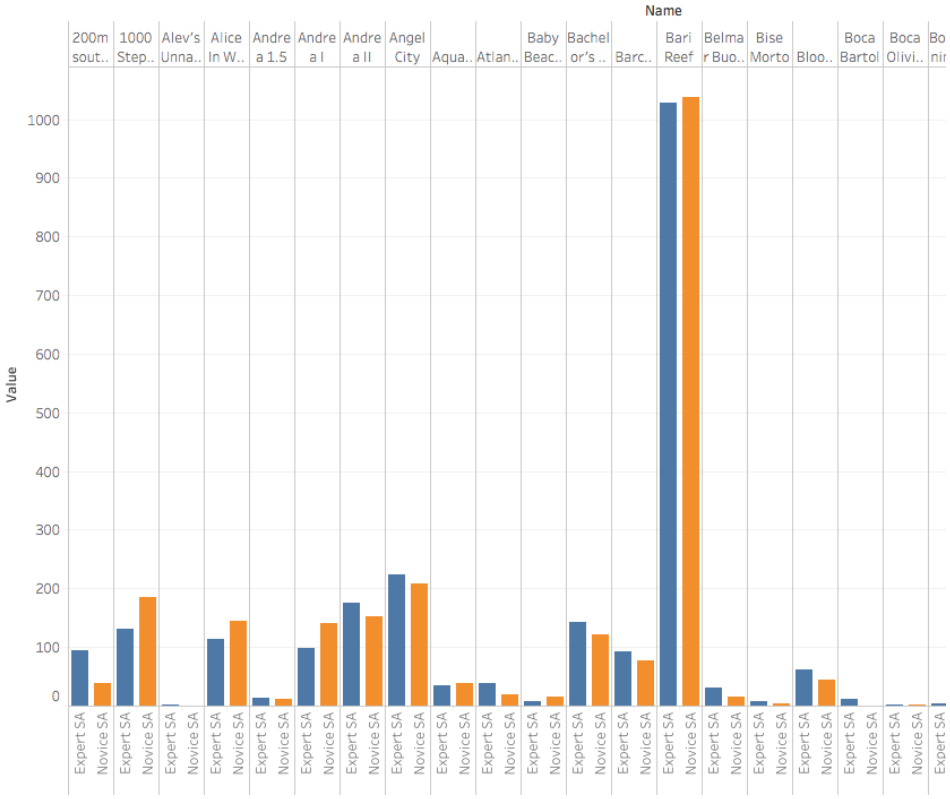


q. Find the Trumpetfish sub-category – it may take some hunting! Let's compare the Trumpetfish *family* to the Damselfish *family*. There is only one *species* of Trumpetfish in our data while there are 14 Damselfish *species*. Note from the prior questions that the Trumpetfish *family* is observed on 91% of surveys while the average Damselfish *family* member is only observed on 52% of surveys. In this view which shows all 547 species, the set of 14 *species* of Damselfish seem to be sighted much more frequently than the Trumpetfish!
r. In a few sentences, interpret these two charts to **explain this paradox**. Suppose a tourist is going snorkeling/SCUBA diving at this location and is confused if they will see any Damselfish or Trumpetfish underwater. *Describe to the tourist what they are likely to observe underwater given these last two charts*.

Next, let's explore the *Survey Counts* dataset.

15. **SA** columns indicate that the surveyor reported the ***S**pecies and their **A**bundance*, while **SO** columns indicate that the surveyor ***O**nly reported the **S**pecies* they saw and not the abundance level. **Name** refers to the name of a coral reef that was surveyed.

s. Let's figure out whether experts or novices are providing data for each reef. Make a **side-by-side bar plot** that compares the total (Sum) number of Expert SA reports to the total (Sum) number of Novice SA reports for each reef site.

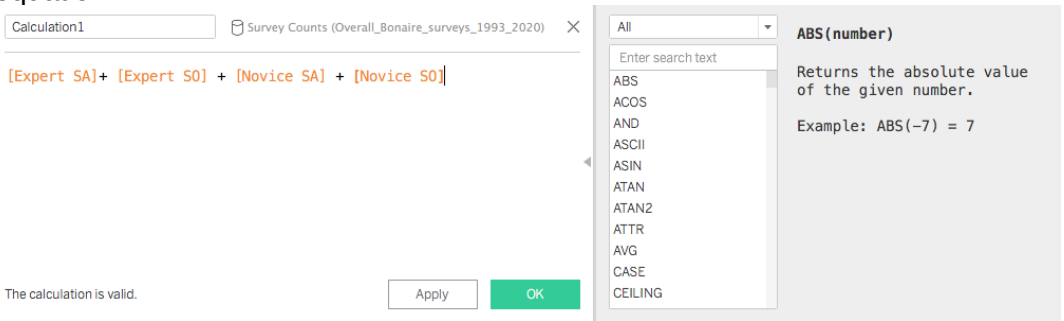


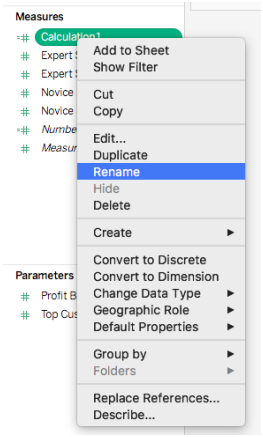
t. Let's assume that experts provide more accurate survey data than novices. Are there any reef sites that have significantly **more data from novice surveys than expert surveys**? If so, analysis regarding these sites may be incomplete or misleading! List any such reef names (hint: ~3-5 locations).

u. Side note: On average, experts report 80.86 species per survey and novices report 55.64 species.

16. Let's create a calculated field that counts how many total surveys are available per location, regardless of the person. To do this, we will add a new measure that simply adds the ExpertSA, ExpertSO, NoviceSA, and Novice SO columns.

v. From the top menu, select Analysis >> Create calculated field, and add the equation.



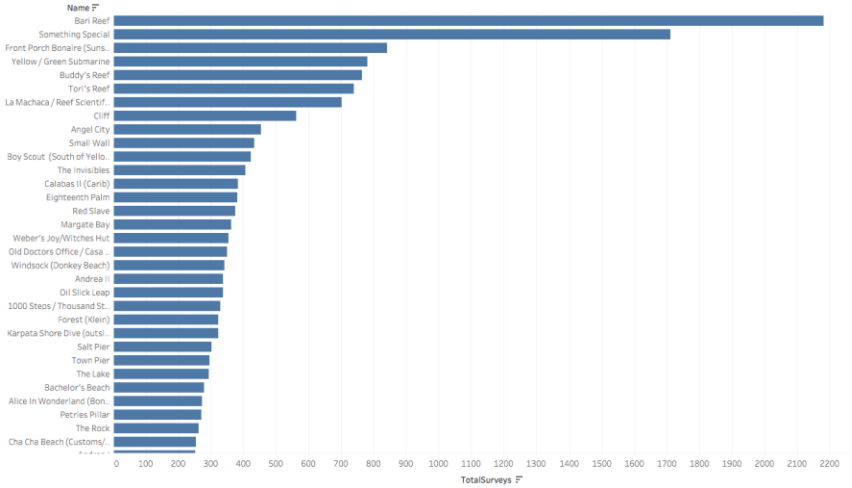


w. Rename Calculation1 to TotalSurveys.

x. Create a sheet of the *total surveys* per reef *name* using a chart type of your choice, e.g., bar.

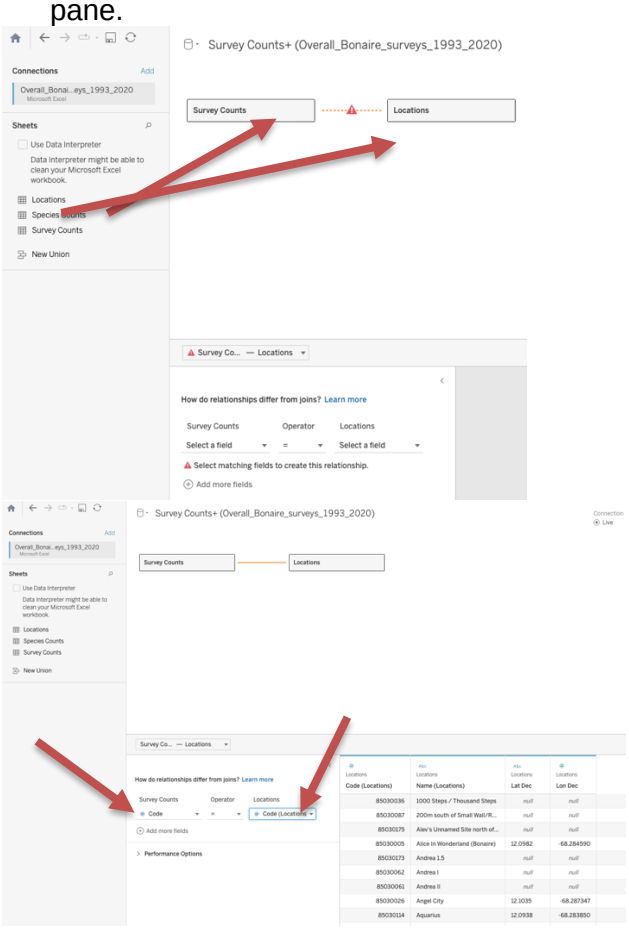
y. Sort the reefs by deceasing numbers of TotalSurveys in the top menu. This menu option can sort ascending, descending, or transpose (flip the rows and columns).

z. What are the **names of the top seven reefs** by survey count? Were any of these locations identified as being **possibly biased** due to too many novices from question 10?



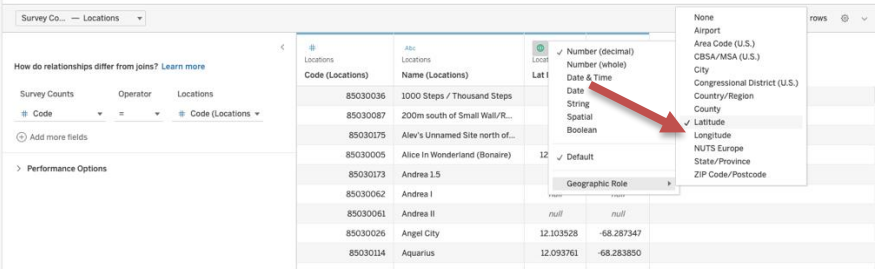
Next, let's place these datapoints on a geospatial map.

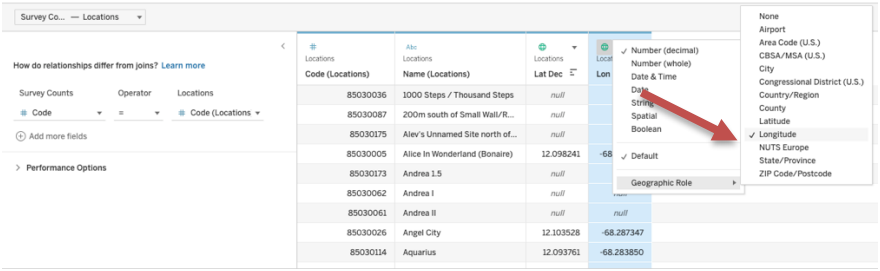
17. Add a new datasource out of the Excel file's '*Locations'* worksheet. Open the Excel file as a new data source. Drag both *Survey Counts* and *Locations* worksheets into the top pane.



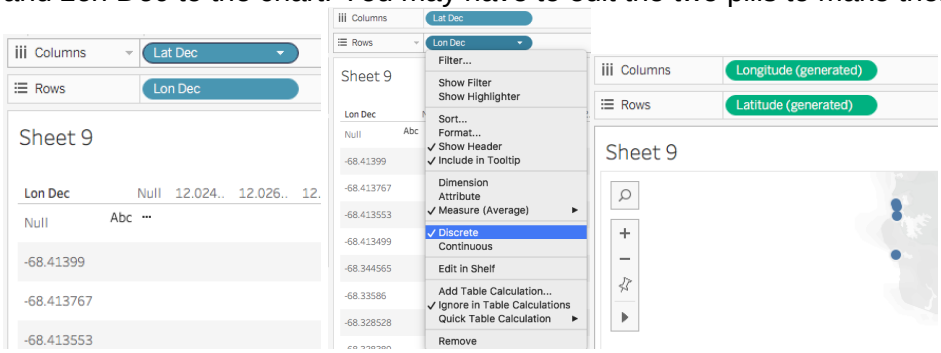
18. Rows from two separate datasources can be merged into a larger row as long as they both contain an identical identifier that can be matched. We are going to join the survey data for each reef location (a row in the *Survey Counts* table) with its actual Lat/Lon coordinates (from the *Locations* table). Edit the relationship to force Tableau to only match rows from the *Survey Counts* and *Location* datasets together if they discuss the same location *Code*. Each reef site gets its own unique *code* and *name*.

19. Edit both *Lat Dec* and *Lon Dec* columns so that Tableau interprets these as decimal numbers and as latitude and longitude data points. Also, select to use them as columns with actual Lat/Lon coordinates.

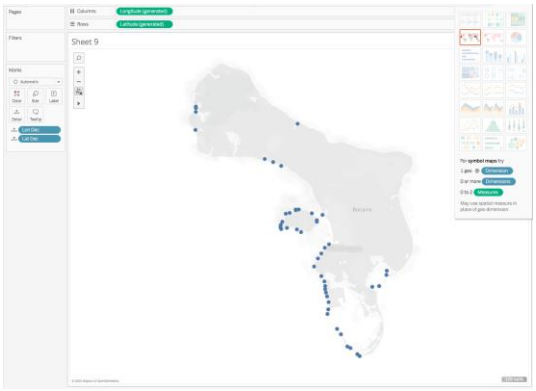




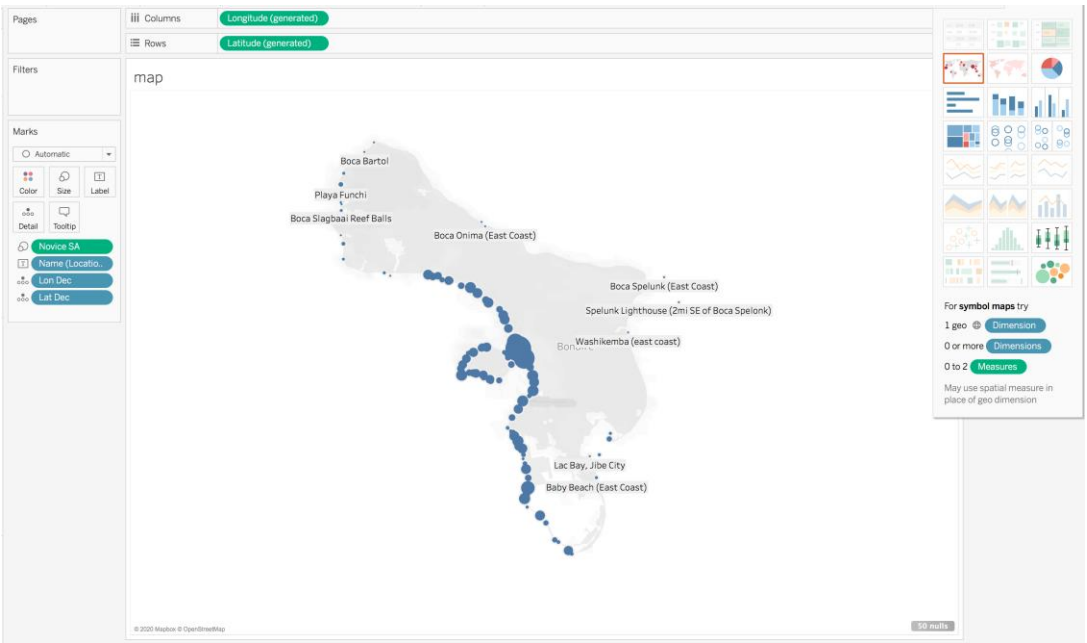
20. Create a new sheet for our map from our new *Survey Counts+* dataset. Drag *Lat Dec* and *Lon Dec* to the chart. You may have to edit the two pills to make them both Discrete.



21. If Tableau interprets your lat/lon coordinates correctly, it will generate a map with a point for each row.



22. Add *Expert SA* to set the size of each point and *Name* to each point's label. Hint: On the left Tables pane, you may have to select the *Lat Dec* and *Lon Dec* dropdown menus and *Convert them to Discrete*. When you add *Expert SA* to the *Marks* pane, you may have to change *Measure(Sum)* to *Attribute*.
23. **Where do experts** seem to conduct the most surveys?
24. Reflect+answer: Why might the Southwest coast of the island be **heavily surveyed and not** the East coast? Could this **bias the type and abundance** of species within the surveys of this trusted dataset?



25. Save and turn in your final Tableau project (.twbx file). Upload a Word or PDF file that contains your answers to each question along with screenshots of your Tableau charts. For full credit, text and screenshots are required for all questions above.

Teaching Case

Yours, mine and ours: Risk assignments, management, and tradeoffs on the road to driverless vehicles

Paul D. Witman
pwitman@callutheran.edu

Jim Prior
jr2dg.consulting@gmail.com

School of Management
California Lutheran University
Thousand Oaks, CA, USA

Scott Mackelprang
scottmackelprangemail@gmail.com

Abstract

This case study looks both backwards and forwards at real and potential incidents and risks created by the use of various levels of self-driving vehicles. The case provides background on autonomous vehicle technology, the legal and risk management frameworks involved, and a variety of scenarios for students to consider. The scenarios provide the foundation for discussion of autonomous vehicle introductions and operations, as well as considerations for how other new technologies may be launched.

Keywords: autonomous vehicles, technology rollout, risk management

1. INTRODUCTION

It is time to go to meet friends for dinner. You open the door of your car, get in, and tell the car where to take you. There are no controls other than the screen which shows the route you will be taking. The car drives up to the door of the restaurant, drops you off, and then finds a place to park. After dinner, you use your phone to ask the car to come pick you up and take you back to your home. When you are not actively using your car, you can lend it out to others to earn some extra cash.

Many people will reach an age or condition when they can no longer drive safely. This might be due to reaction time, vision, confidence, or physical limitations. It would be wonderful to still have the same freedom of movement that is provided by driving - which may be possible if we can get technology to do some of the work for us. But such vehicles are not here yet - and current versions still have quirks, like the Waymo vehicles that wrongly routed dozens of trips into a dead-end *cul-de-sac* in San Francisco, disrupting traffic for those residents (CBS News Bay Area, 2021).

The spectrum of automation that can be implemented in driver-assisted vehicles can best be understood using the Society of Automotive Engineers (SAE) classification system, illustrated in Figure 1. This model defines levels 0, 1, and 2 as Advanced Driver Assistance Systems (ADAS), which require a driver. It also defines levels 3, 4,

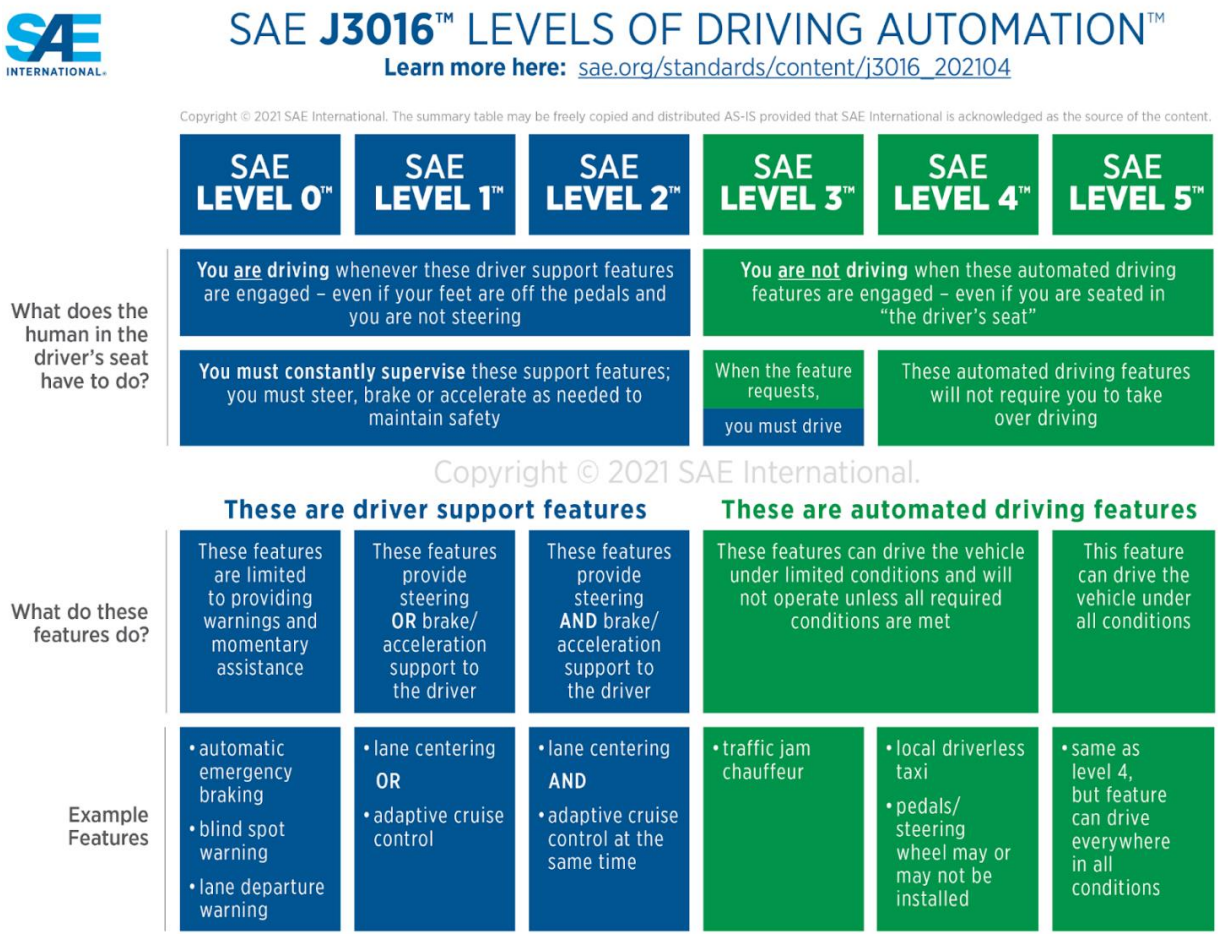


Figure 1: SAE Levels of Driving Automation (Society of Automotive Engineers International, 2021), used with permission

and 5 as Automated Driving Systems (ADS), which allow complete control of the vehicle by the technology under a variety of limited conditions, ultimately leading to a vehicle that requires no human driver (Society of Automotive Engineers International, 2021).

Many ADAS features are available in cars today that provide features such as adaptive cruise control, forward collision warning, automatic emergency braking, lane departure warning, lane-keeping assistance, and blind-spot warning as well as more advanced features such as lane-centering technology, lane-changing assistance, and self-steering systems. Drivers may already be developing dependencies on these capabilities and adjusting their driving habits because of them. However, there are currently no Automated Driving Systems available for consumer purchase in the U.S., though some are being used in tests for taxis and other such commercial operations (U.S. Department of Transportation, 2021).

There are many potential benefits from current and future levels of driving automation. Proponents expect to achieve increased safety for drivers and passengers along with other motorists and pedestrians, as well as increased mobility via expanded transportation options. They also expect economic and social benefits from a reduction in car crashes, less need for parking spaces/lots, increases in automated ride sharing reducing environmental impact, and efficiency and convenience with reduced traffic congestion (Litman, 2017).

Of course, the automation of driving brings challenges as well: legal liability and insurance systems to accommodate and cover accidents of autonomous and semi-autonomous vehicles, the need for new state and federal regulations around

risk assignments, and technology limitations that make fog and rain, or poor road markings more difficult to manage safely. Risks stemming from malicious hacking of autonomous vehicles can be anticipated, as can the potential for privacy issues as the vehicles coordinate their movements among themselves and share their observations about road conditions. It is also reasonable to expect people to use vehicles in ways that are good for them, while less good for society – e.g., sending their car away from a downtown location to find cheaper parking, saving money but creating traffic. All of these must be appropriately weighed against the economic and social value of entrepreneurship and innovation, lest we over-regulate.

Autonomous vehicles will change how accident liability is decided in the courts. Over time, individual court cases across the spectrum of SAE levels of driving automation will be treated as precedents contributing to autonomous vehicle case law. Judges and juries will have to contend with potentially complex technical issues. Was the accident due to a failure of a collision avoidance subsystem manufactured by a third party? Or a software failure of some kind? Did the driver fail in their ultimate responsibility? Or did the vehicle's automation fail them? Was the system not maintained properly? What do the current and future laws say about liability in those situations?

Not surprisingly, it takes a lot of technology to try to replicate and improve on the human driver. Vehicles need to know where they are in terms of geographical location, they need to know what is around them, how to react to those things, and how to interact with them appropriately. Many types of sensors are involved, as highlighted in Figure 2, below. Sensors may include things like visible light cameras, short- and long-range radio-based distance tracking (radar), ultrasound, and light-based distance tracking (called LIDAR). Data from all of these sensors is then integrated in high-performance computing devices that determine how to change speed and direction as required (Wendt & Cook, 2018). In addition to the sensors, vehicles need Global Positioning System (GPS) or Global Navigation Satellite System (GNSS) data from satellites, detailed maps (more detailed than are publicly available), and map updates based on recent data from other vehicles which may have discovered things like lane closures, etc.

To make sure that these systems, whether ADAS or automated driving, work properly, they go through many levels of testing by the manufacturer. Eventually, the vehicles need to be tested in a ”real-world” environment – the actual roads and traffic conditions that the vehicles need to operate in. This naturally introduces new risks to occupants of automated vehicles: an error

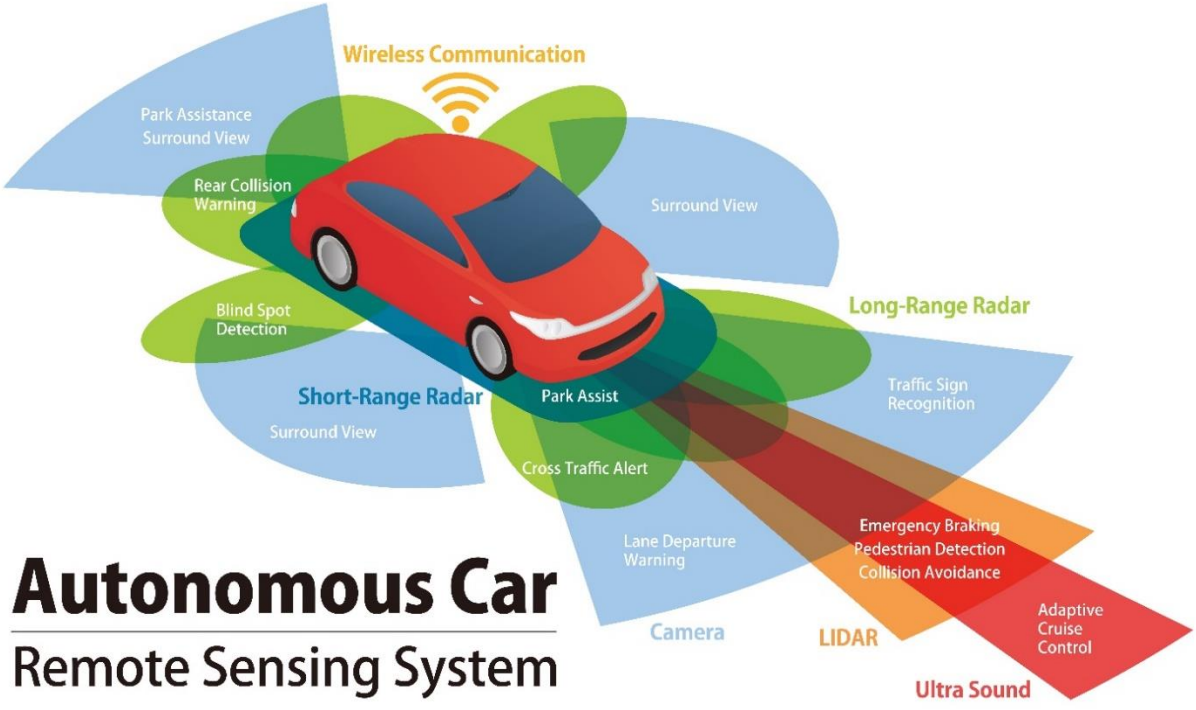


Figure 2: Sensor components of autonomous vehicles (metamorworks, 2022)

could injure passengers and drivers as well as damage the vehicle. It also introduces new risks for other nearby drivers and pedestrians who might be injured if automated vehicles make an error.

Risk is a concept that is not unique to business or technology - it is fundamental to many aspects of everyday life. Risk refers to the probability that a certain type of event will occur. Related to that concept of risk is the magnitude of impact that might occur should the event occur. Risks can result in positive or negative outcomes, though we most commonly focus on risk as being related to negative outcomes - injuries, damage, cost, and the like.

There are several fundamental approaches to managing risk. One can avoid the risk, transfer it to someone else, reduce it to acceptable levels or simply accept it.

> *Avoidance*: If a risk is determined to be too high, then you avoid the activity that creates the risk.
>
> *Transfer*: In many cases, the risk can be transferred to another party.
>
> *Reduction*: Risk reduction is a common step for processes or activities that cannot be avoided, and where the risk cannot be transferred to another party.
>
> *Acceptance*: In some cases, the best option is to accept a risk (Ahmed, 2017). If steps have been taken (as in the above strategies) to reduce or mitigate the risk, the remaining level of accepted risk is known as "residual risk."

The dynamics of the risk management being exercised in the advertising, sale, regulation, operation, and enforcement of autonomous vehicles are currently fluid and not well understood by the public. This will, undoubtedly, lead to challenges for the industry.

Ethical choices, for the manufacturer, the user, and other parties, are generally viewed through one of several different "lenses." These lenses include (among others) the fundamental "rights" of affected parties, a view of what represents *justice* in the broader context of merit and need, a utilitarian view that balances the resulting good vs. harm across all parties involved, and a "common good" approach that is similar to the utilitarian approach, but recognizes a mutual concern for all parties involved directly or indirectly (Markkula Center for Applied Ethics, 2021).

For the past several decades technology has advanced more rapidly than the societal and governmental structures to manage and coordinate them (Malan, 2018). For example, e-cigarettes and social media, which use advanced technology in dynamic economic and operational environments, both confronted society in advance of the regulatory structures to govern them. Manufacturers and users of autonomous vehicles are now beginning to face the challenges of managing their legal risks and liabilities without well-defined legal structures. Such risks can impact the rollout or adoption of autonomous vehicles. In at least one case, release of an otherwise-ready technology, Audi's Level 3 self-driving for traffic jams, was deferred largely due to lack of a legal framework to manage its risk and potential liability (Szymkowski, 2020).

2. DISCUSSION SCENARIOS

The following are several scenarios to consider as to how the introduction of this new technology has been or could be managed. The goal of these scenarios is to spark your thinking about the various factors that impact the deployment of new technologies. These factors might include issues of risk, liability, insurance, legal structures, ethical considerations, system testing and more. Each scenario includes a real or hypothetical situation, and the technology and other decisions and factors that affect it, and it asks you to consider the tradeoffs and decisions involved.

Risks, Broadly Considered

Mary heads off to work in her car, driving herself as usual. Two blocks away, Sam leaves about the same time heading to a destination close to yours, but lets his car choose the route, navigate, and control the vehicle's operation. Mary encounters Sam's car on the way, to her right at a four-way stop, arriving just ahead of Sam's car. Mary thinks she has the right of way, but can't make eye contact with Sam, even after honking. She waits, as does Sam's car – neither one immediately entering the intersection. When Mary decides after a few seconds to go ahead, she sees Sam's car start to do the same. After another false-start, Mary finally proceeds across the intersection, followed by Sam's car. Mary has just encountered an unexpected risk (not getting the usual cues from the other driver), which she has coped with by paying attention, carrying insurance, etc. All drivers take a number of measures to reduce the risk of adverse events happening, and to reduce the damage if they do, or to transfer some of the risk to others.

While there are many potential benefits to autonomous vehicles, benefits always come with some new risks – as with every new technology. Drivers may be more inclined to risk "driving" while intoxicated, since the car will "do most of the driving." Drivers have already, many times, bypassed "alert driver" detection systems (e.g., Dunn et al., 2021; Lambert, 2021), which defeats those controls.

These new risks also extend to other drivers, pedestrians, first-responders, cyclists, and road workers. Even though a self-driving car may be statistically safer than the average human driver, there have still been cases of those vehicles behaving erratically, in ways that are unexpected by their human-driven counterparts, or that require the ADAS driver to take control (Goodall, 2021; Smiley, 2022). Either of these may result in damage, injury, or death to the ADAS driver and to all those around them.

Among the many tradeoffs that arise with autonomous vehicles, there is a phenomenon called automation complacency (Azuma et al., 2022). When humans are responsible for operating a device that largely operates independently and successfully, it is common for the operator to lose focus, and to assume the device will always do the right thing, even when their own safety and the safety of those around them depend on their ability to retake control of the vehicle quickly.

Questions

- When a car manufacturer sells you a self-driving car, and you sign a contract saying that you understand that you are responsible for its safe operation, which risk management approach is the manufacturer using?
- Is that the "right" approach? How would you justify your answer?
- What are the ethical implications for the driver/owner, as well as for the public?
- If you as an autonomous vehicle owner want to reduce the risks associated with autonomous driving, what steps and decisions could you make to reduce the risk that your vehicle injures anyone, or damages property?

Cybersecurity Risks of Autonomous Vehicles

In 2015, two self-described hackers demonstrated an attack on a Chrysler Jeep Cherokee. In the attack, they used an insecure channel through the entertainment system of the vehicle to take control of air conditioning, radio volume, the picture on the dashboard screen, driving speed, and windshield wipers. The hackers were able to do this remotely, via a wireless Internet connection, and required no direct physical access to the Jeep's systems (Greenberg, 2016).

Now think ahead to a Level 4 or 5 autonomous vehicle, with no driver controls available. The vehicle navigates based on input from GPS satellites, maps downloaded from a central server, and the measurements of sensors in the vehicle, along with (perhaps) direct communication with other nearby vehicles.

Knowing that hackers are motivated by a wide range of goals (money, fun, power, bragging rights, etc.), hackers might attempt to conduct random attacks on cars, or might target a specific vehicle or vehicles to try to disrupt, frighten, or injure that vehicle's rider. Hackers might try to take remote control of the vehicle, as in the story above, or might try to disrupt its sensors in some fashion.

Questions

- Identify the navigating and driving control inputs that might be attacked, and some potential approaches to prevent or mitigate those attacks.
- How might a hacker attack a vehicle in a way that could generate income for the hacker?
- What do manufacturers need to do to mitigate the risk of cyberattacks?
- What privacy considerations arise when society is being driven around in autonomous vehicles?
- Should there be identification and authorization requirements for operating an autonomous vehicle? How might they be implemented to help prevent large scale vehicle theft in the event of a centralized vehicle controller breach? How might such control measures impact the ability of owners to lend their car to a friend?

Marketing vs. Rigorous Definitions of "Self-Driving"

Melissa has gone out to buy a new car and is considering buying a Tesla. She notes that it has a feature called "Autopilot." She's also considering a Honda, which has a feature called "Traffic Jam Pilot," and a Mercedes Benz, who call their feature "Drive Pilot." Melissa is at a bit of a loss to understand exactly what these features do, without looking deeply into the sales material.

What do you think Melissa sees when she reads the word *autopilot* on its own? Perhaps an image of an airline pilot sitting back, letting a system control an airplane in flight?

What about the Tesla feature called "Autopilot"? Same thing? Or is it a new concept since it is related to a car? What about "Drive Pilot" or "Traffic Jam Pilot"? How do those differ, if at all, from "Autopilot"?

These companies may be branding their conditional automated driving systems in this manner, hoping that you will think their systems are like those used in airplanes.

But are they? Are the circumstances the same? Commercial airline pilots give each other a wide berth in the sky, reducing the need for split-second decisioning. That is certainly not the case with automobiles. And commercial airliners have two pilots, backing each other up. Not so with automobiles.

Of course, companies place great value in their branding approaches, seeking differentiation from their competitors, and seeking to attract their target market. But they also owe potential customers clarity and transparency about their products and services, knowing that some customers will not read beyond the brand name. This is all the more true when the product takes over responsibility for driving a car, thus making life and death decisions.

All three of these "pilot" systems require certain conditions to be met for the automation to take over. Tesla claims that Autopilot will steer, accelerate and brake automatically within its lane (Tesla, 2022). Presumably, anything else requires driver intervention. Mercedes Benz' initial offering of Drive Pilot was limited to specific roadways in Germany that had been mapped to exacting three-dimensional detail (Mercedes-Benz Group, 2022). Honda's initial deployment of Traffic Jam Pilot is limited geographically as well (Honda European Media Newsroom, 2021).

But how might these and other companies be held accountable for truthfully stating and delivering on their marketing promises?

In January of 2020, the U.S. Department of Transportation and the National Science and Technology Council published the fourth in a series of reports on autonomous vehicles. One of the focus areas was protecting users and communities. Emphasizing safety, the report stated that the federal government will "enforce existing laws to ensure entities do not make deceptive claims or mislead the public about the performance capabilities and limitations of AV technologies including, for example, deceptive claims relating to vehicle safety or performance" (National Science & Technology Council & U.S. Department of Transportation, 2020). There are no new federal laws yet, but there is some existing consumer protection via the Federal Trade Commission's Truth-In-Advertising statutes.

U.S. states have actively been promoting and supporting autonomous vehicle development for over a decade. As of this writing, 38 states and the District of Columbia have enacted legislation or issued executive orders regarding autonomous vehicles. The legislative actions include authorization of studies, testing, and full deployment of autonomous vehicles (Governors Highway Safety Association, 2022). Yet there is still no specific regulation regarding unsubstantiated claims from manufacturers.

Ultimately, it will be incumbent upon lawmakers and regulatory bodies to appropriately protect the public as a whole.

Questions

- Consider the range of names chosen by automakers for their self-driving technologies: e.g., GM Supercruise, Waymo Driver, Tesla Autopilot and Full Self Driving, Audi City Assist, Subaru EyeSight, Mercedes Drive Pilot. Why do automakers (and other sellers) choose names for these features that could be misleading?
- How might the automakers change their choices to improve the clarity of what the feature actually does?
- Is Melissa legally or ethically required to read and understand the feature capabilities before putting them to use?
- Thinking from one of the ethical lenses mentioned above, what are the ethical implications of the names given to these features?

Vehicle Sharing

Joseph has gone home for the night in his Level 5 autonomous vehicle. It is 7 p.m., he will have dinner, talk with the family, and turn in for the night. Meantime, his car (a $40,000+ purchase) might sit idly in the driveway all night – eight to twelve hours of underutilizing that asset. Instead, Joseph puts that car to work. He signs up with a ride-sharing service, and his car can then go out on its own and pick up riders, take them to their

destinations, and return home in time to recharge and take Joseph to his next destination.

Since at least as far back as 1995, people around the world have started to embrace the modern version of the "sharing economy," as enabled by information technology (Puschmann & Alt, 2016). A growing sector of the sharing economy (e.g., Airbnb, Uber, Lyft, etc.) allows people to use a personal asset like a home or car to earn extra income. It may be that this same notion could extend into the use of the idle time of autonomous vehicles. Imagine you own a fully autonomous (SAE Level 5) vehicle. Beyond the nighttime rides in the story above, the car could also take the owner to work, and then it would not have to sit idle all day in the parking lot. Instead, it can pick up riders during the workday as well.

Owners might be able to choose limitations on how far the vehicle is allowed to travel, or how many people they are willing to let the car transport. For each ride that the vehicle performs, the owner earns a share of the income, while the service keeps the remainder to cover the cost of coordinating and managing the rides, riders, and vehicles. Presumably, some makers of autonomous vehicles will do better at providing highly specific navigation (getting to or from a specific building at a commercial site, etc.).

In a world where we begin sharing unsupervised access to our "personal spaces" (e.g., home, car, etc., that we routinely also use ourselves), this becomes fundamentally different from sharing space in your home (a la Airbnb) or in your car with you driving (a la Uber). The natural inclination to behave "properly" when the "renter" is with the "owner" diminishes if the owner is not present or visible, which would be the case for your car picking up rides while you sleep or work.

Questions

- What privacy issues are raised when you allow someone to use your car without your presence?
- What risks does the car owner take on in offering their vehicle this way?
- How can different car navigation capabilities be communicated to the passengers?
- What are the ethical considerations of vehicle sharing, both for the owner and for the "rider"?
- What are the pros and cons for society of enabling this sort of vehicle sharing?
- Thinking perhaps more maliciously, what other (perhaps criminal) uses might autonomous vehicles be put to?

Real Accidents with Self-Driving Vehicles

Late one night in Tempe, AZ, a test vehicle for Uber's autonomous vehicle program was traveling in fully autonomous mode with a backup vehicle operator behind the wheel. The vehicle's systems detected "something" ahead of it, moving toward the vehicle's path. The operator should have had full attention on the road, but was allegedly distracted by something else in the vehicle (Smiley, 2022). The on-board systems interpreted the "something" at various times as a bicycle, a person, and an unknown object, but did not alert the vehicle safety operator of the potential issue until it was too late, and what turned out to be a pedestrian was struck and killed.

Sadly, there is no shortage of incidents of vehicles operating autonomously (whether at Level 3 and above, fully autonomously as in this Uber case, or at Level 2 or below, where a safety driver is expected to be instantly available) that encounter accidents of some kind. The Uber incident above is just one of those. There are also numerous incidents of Tesla vehicles, running at Level 2, hitting parked emergency vehicles, perhaps due to the human-driven vehicle in front of them going around the emergency vehicle too quickly. There are also many cases where drivers have somehow bypassed the driver detection systems and let the car drive itself while the driver slept, or climbed into a different seat to "prove" how safe the car was (Dunn et al., 2021; Mak, 2021; Song et al., 2021). Other reports indicate that most accidents involving self-driving vehicles are caused by human error, such as rear-ending the autonomous vehicle (Jeffs, 2022).

Putting autonomous vehicles on the road requires the rollout of new technology that needs to operate safely on public streets, interacting with human drivers in other vehicles. Manufacturers of vehicles, and the related autonomous technology and sensors, need mechanisms and processes to safely gain confidence- and quality-building test miles without putting anyone at unreasonable risk. Even if the system is running in fully autonomous mode (e.g., Level 5), manufacturers need to test these vehicles in a public, "real-world" environment to be confident that they can trust the vehicle to make the right decisions.

Questions

- How does the operation of a vehicle at Level 4 or 5 change the dynamics of

interaction among vehicles? In other words, without a driver in control, what changes about the vehicle interactions?
- Thinking about the interactions that pedestrians have with vehicle drivers - what changes about those interactions when there is no "driver"?
- What steps could be taken by manufacturers to help mitigate the risk of putting these vehicles into the hands of the general public?

Insurance Industry Impact

In 2010, motor vehicle crashes in the United States caused 33,000 deaths, 3.9 million injuries, and 24 million damaged vehicles. The economic cost was calculated at $242 billion (Blincoe et al., 2015), including health care, lost work, repair costs, etc., and not all covered by insurance. Unfortunately, many studies have also shown that over 90% of car accidents are caused by human error (Treat et al., 1979).

One of the key anticipated benefits of ever-increasing use of autonomous vehicles is a reduction in human drivers and a presumed, corresponding reduction in accidents. That would be good, leading to a reduction in deaths, injuries, property, and financial losses.

But what happens to the automobile insurance industry? How should it change? How will it change?

Consider that the market value of the U.S. automobile insurance industry is $316 billion in 2022 (IBISWorld, 2022). That value comes from revenue, which starts with premiums charged to individual customers and businesses. Over time, with a reduction in accidents, insurance underwriters should adjust premiums downward to accurately reflect risk, thus reducing revenue.

Given that the transition to vehicles with automated driving capabilities (SAE Levels 4 and 5) will occur over many years, it is reasonable to assume that the automobile insurance industry can continue to survive and thrive while adapting to these new technologies. Discussions within the industry point to various other models for ensuring autonomous vehicles that may be explored. For instance:

No-Fault Insurance

While the track record of no-fault auto insurance is a bit bumpy, there are still 12 U.S. states that use a limited version in which drivers must use their own insurance to pay for their injuries after a crash - no matter who is at fault. They protect themselves by purchasing Personal Injury Protection (PIP) coverage (Hurst, 2022). Given that fully automated vehicles will not be driven by humans, there may be an opportunity to explore this approach.

Manufacturers May Self-Insure

Given that liability for accidents caused by autonomous vehicles will often reside with manufacturers, they may choose to self-insure - the company sets aside a pool of money to cover any losses - perhaps to a certain amount - then backed by insurance.

Fleet Insurance

It is very likely that manufacturers will partner with other businesses to manage pools (fleets) of autonomous vehicles. This business model is well known in the auto industry with, for instance, bulk sales to rental car companies and large corporations. These business relationships often leverage fleet insurance policies, which provide liability insurance for a fleet of vehicles under one policy (Stanley et al., 2020).

Questions

- What other approaches might be taken to pay for accidents involving autonomous vehicles?
- How might technological innovation aid in determining liability with autonomous vehicle accidents?
- How might the cost of insurance coverage be affected by road conditions where an autonomous vehicle is used?

3. TOPICS FOR ADDITIONAL RESEARCH

The following questions are intended to be used for additional research topics, perhaps for in-class discussion, for student research papers, or other purposes. The questions sometimes extend beyond the specifics of autonomous vehicles and into broader areas of technology deployments.

Driver (Operator) Attention

In vehicles at Level 3 automation and below, the vehicle expects the driver to be ready to take over control at a moment's notice - either because the operator detects a problem that may not have been caught by the vehicle, or because the vehicle cannot determine the correct next step. Many accidents have resulted from operators becoming too complacent and not paying attention.

Figure 3: Clinical trials (MD Anderson, 2022), used with permission

Question

- How can the automated systems in the vehicle monitor and track operator attention to ensure that the operator remains ready to take over?

Unintended Consequences of Widespread Adoption of Autonomous Vehicles

"The world is full of tradeoffs" -- Anonymous

Proponents of autonomous vehicles expect them to:

- become widely adopted,
- be less accident-prone than human drivers, and
- offer the possibility that people can readily get from place to place in an autonomous taxi, rather than owning a vehicle of their own.

If the cost of a "ride" (e.g., from an Uber-like service without a driver) becomes low enough, you will not need to own a car to be able to have car-like freedom of movement. If your own car can navigate and move entirely on its own, it may not need to park near where it drops you off - it could drive somewhere else. Perhaps it can park more cheaply elsewhere; perhaps it will pick up other passengers until you ask it to return.

Questions

- What kinds of side effects might occur if the proponents' expectations come to pass?
- What industries might see increased revenues?
- What industries might see decreased revenues?
- What new industries might appear?
- What new public safety risks does this introduce?

Parallels to Medical Clinical Trials

In some ways, the rollout of an innovative technology might be like the rollout of new medications or vaccines. Autonomous vehicles have numerous interactions with broad groups of people. Those interactions could work well in some situations and cultures, and poorly in others, with perhaps serious consequences, much like medical treatments.

Consider the process for conducting "clinical trials." In general, new treatments and approaches go through a multi-phase trial process, with carefully selected groups of participants to ensure that the tests assess the right questions, and that appropriate levels of risks are undertaken and managed.

A simple summary of the clinical trial phases follows (illustrated in Figure 3):

- Phase 1 - Tests if a new treatment is safe, and how best to deliver that treatment.
- Phase 2 - Tests if the treatment generates the desired reaction from the target. Does it generate side effects?
- Phase 3 - Tests new treatment as

compared to existing treatment (is the new treatment measurably better?).
- Phase 4 - Start broad delivery and monitor results.

For comparison, you might consider the model proposed for rollout of fully autonomous (driverless) taxi service in San Francisco, CA, run by GM's Cruise technology unit (LaReau, 2022). Their permit allows for a phased rollout of the technology, initially limited to "a maximum speed of 30 mph, from the hours of 10 p.m. to 6 a.m. daily 'when weather conditions do not include heavy rain, heavy fog, heavy smoke, hail, sleet, or snow.'" The permit could be broadened later, perhaps extending service hours, geography, or conditions. All of this is being done after the technology has been tested with human operators on board to monitor behavior and take over if needed.

Questions

- How does the rollout of autonomous vehicles compare to the clinical trial process and phases?
- Is this a potential model that could be used to structure our thinking about technology deployments to the general public?
- Are there ethical considerations for autonomous vehicle testing similar to those of clinical trials?
- Where, if anywhere, does the analogy to clinical trials break down or not make sense?
- What are the societal costs (if any) to be borne if government delays or neglects the oversight of managing the liabilities of autonomous vehicle rollouts?

4. CONCLUSIONS

Innovative technologies are in development all the time, in many sectors of the economy. Some are isolated – for example, deployment of robotic devices in a factory, which directly affects only the workers around those robots, and those whose jobs might be affected. Others, though, have a multi-faceted collection of interactions with broader groups of people, up to the broadly defined "general public." In the case of autonomous vehicles, people can be, at various times:

- owners and operators of autonomous vehicles,
- paying passengers in autonomous taxis or buses,
- users who receive shipments or deliveries from autonomous vehicles,
- driving other less-autonomous vehicles that must interact with the autonomous ones, and
- pedestrians, cyclists, and others who also must interact with autonomous vehicles.

Getting all of those interactions "right" is not a trivial problem. It is possible to design and evaluate approaches to those interactions in an isolated "laboratory" setting. Before a full and broad rollout of the technology, manufacturers (and regulators, insurers, etc.) want to be confident that those interactions will work well, and under all circumstances. Can the vehicle navigate a snowy road? Can the vehicle maneuver through a police roadblock? Does the vehicle interaction at a 4-way stop intersection work correctly? What about roundabouts? Construction sites? Detours? Do driver differences across states, countries, and cultures all work "correctly"?

Gaining enough confidence that those approaches are correct requires many hours of interactions with large swaths of the population, to be sure that the interactions work, without fail, under all conditions. Those hours require taking a variety of risks, and it is critical for technology managers to think through those risks and engage the other participants in the process to mitigate those risks appropriately. We hope that these scenarios have provided things to consider about the challenges of delivering new technologies to the public, and about managing those risks appropriately.

Governments and regulatory agencies also need to define the legal framework in which these vehicles will be operated, during testing and after full rollout. Legal liabilities should be well-defined and appropriate for all the potential scenarios. Regulators should carefully consider the full range of risk tradeoffs and specify appropriate compensating controls to offset each of those risks. Those controls might include limiting the conditions in which these systems can be used, limiting the geography where they can be used, or other similar constraints to manage the risk to the public. Automakers need to be crystal-clear about, and actively enforce, the need for driver attention. Only then can we safely get from "I drive" to "it drives."

5. ACKNOWLEDGEMENTS

The authors appreciate the support and insights of various industry experts who provided insights for this case. The authors also appreciate the productive feedback provided by their peers, and by the reviewers and conference chairs.

6. REFERENCES

Ahmed, R. (2017). Risk Mitigation Strategies in Innovative Projects. In *Key Issues for Management of Innovative Projects*. IntechOpen. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3884687

Azuma, M. C., Giordano, F., Stoffregen, S., Klos, L., & Lee, J. (2023). It practically drives itself: autonomous vehicle technology, psychological attitudes, and susceptibility to risky driving behaviors. *Ergonomics 66*(2), 246-260.

Blincoe, L. J., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. (2015). *The Economic and Societal Impact Of Motor Vehicle Crashes*. U.S. Department of Transportation, National Highway Traffic Safety Administration https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812013

CBS News Bay Area. (2021). *Dead-End SF Street Plagued With Confused Waymo Cars Trying To Turn Around 'Every 5 Minutes'*. Retrieved June 9 from https://www.cbsnews.com/sanfrancisco/news/dead-end-sf-street-plagued-with-confused-waymo-cars-trying-to-turn-around-every-5-minutes/

Dunn, N. J., Dingus, T. A., Soccolich, S., & Horrey, W. J. (2021). Investigating the impact of driving automation systems on distracted driving behaviors. *Accident Analysis & Prevention*, *156*, 106152.

Goodall, N. J. (2021). Comparison of automated vehicle struck-from-behind crash rates with national rates using naturalistic data. *Accident Analysis & Prevention*, *154*, 106056. https://profile.virginia.edu/sites/g/files/jsddwu351/files/njg2q/struckfrombehind_1.pdf

Governors Highway Safety Association. (2022). *Autonomous Vehicles*. Retrieved June 8 from https://www.ghsa.org/issues/autonomous-vehicles

Greenberg, A. (2016, 1 Aug). The Jeep Hackers Are Back to Prove Car Hacking Can Get Much Worse. *WIRED*. https://www.wired.com/2016/08/jeep-hackers-return-high-speed-steering-acceleration-hacks/

Honda European Media Newsroom. (2021). *Honda launches next generation Honda SENSING Elite safety system with Level 3 automated driving features*. Retrieved May 27, 2022, from https://hondanews.eu/eu/et/corporate/media/pressreleases/329456

Hurst, A. (2022). *States with no-fault auto insurance and what it means*. PolicyGenius. Retrieved May 30, 2022, from https://www.policygenius.com/auto-insurance/states-with-no-fault-insurance/

IBISWorld. (2022). *Automobile Insurance in the US - Market Size 2004–2027*. Retrieved May 25, 2022, from https://www.ibisworld.com/industry-statistics/market-size/automobile-insurance-united-states/

Jeffs, D. J. (2022). *Autonomous Cars, Robotaxis & Sensors 2022-2042*. IDTechEx. https://www.idtechex.com/en/research-report/autonomous-cars-robotaxis-and-sensors-2022-2042/832

Lambert, F. (2021, May 5). Tesla driver repeatedly spotted in backseat on Autopilot is begging to be arrested. *Elektrek*. https://electrek.co/2021/05/05/tesla-driver-keeps-being-spotted-in-backseat-autopilot-begging-arrested/

LaReau, J. L. (2022, June 2). GM's Cruise will be the first to offer driverless taxis in a major U.S. city. *Detroit Free Press*. https://www.freep.com/story/money/cars/general-motors/2022/06/02/gm-cruise-approved-california-driverless-taxis/7487221001/

Litman, T. (2017). *Autonomous vehicle implementation predictions*. Victoria Transport Policy Institute Victoria, BC, Canada.

Mak, A. (2021, Aug 17). Why Teslas Keep Striking Parked Firetrucks and Police Cars. *Forbes*. https://slate.com/technology/2021/08/teslas-allegedly-hitting-emergency-vehicles-why-it-could-be-happening.html

Malan, D. (2018). The law can't keep up with new tech. Here's how to close the gap. https://www.weforum.org/agenda/2018/06/law-too-slow-for-new-tech-how-keep-up/

Markkula Center for Applied Ethics. (2021). *A Framework for Ethical Decision Making*. Santa Clara University. Retrieved June 10, 2022, from https://www.scu.edu/ethics/ethics-resources/a-framework-for-ethical-decision-making/

MD Anderson Cancer Center. (2022). *Phases of Clinical Trials*. Retrieved June 4, 2022, from

https://www.mdanderson.org/patients-family/diagnosis-treatment/clinical-trials/phases-of-clinical-trials.html

Mercedes-Benz Group. (2022). *The front runner in automated driving and safety technologies*. Retrieved May 30, 2022, from https://group.mercedes-benz.com/innovation/case/autonomous/drive-pilot-2.html

metamorworks. (2022). Autonomous Car Remote Sensing System. In 530214349 (Ed.). Shutterstock.com.

National Science & Technology Council, & U.S. Department of Transportation. (2020). *Ensuring American Leadership in Automated Vehicle Technologies - Automated Vehicles 4.0*. The Office of Science and Technology Policy. https://www.transportation.gov/av/4

Puschmann, T., & Alt, R. (2016). Sharing economy. *Business & Information Systems Engineering*, *58*(1), 93-99.

Smiley, L. (2022, April). "I'm the Operator": Aftermant of a Self-Driving Tragedy. *WIRED*, *30*(04), 54-69. https://www.wired.com/story/uber-self-driving-car-fatal-crash/

Society of Automotive Engineers International. (2021, May 23). SAE Levels of Driving Automation™ Refined for Clarity and International Audience. *SAE Blog*. https://www.sae.org/blog/sae-j3016-update

Song, Y., Chitturi, M. V., & Noyce, D. A. (2021). Automated vehicle crash sequences: Patterns and potential uses in safety testing. *Accident Analysis & Prevention*, *153*, 106017.

Stanley, K. D., Grise, M., & Anderson, J. M. (2020). *Autonomous Vehicles and the Future of Auto Insurance*. R. Corporation. https://www.rand.org/pubs/research_reports/RRA878-1.html

Szymkowski, S. (2020). Audi hangs up hopes for Level 3 partial automation system. Retrieved May 31, 2022, from https://www.cnet.com/roadshow/news/audi-a8-level-3-automation-traffic-jam-pilot-system/

Tesla. (2022). *Autopilot*. Retrieved May 27, 2022, from https://www.tesla.com/autopilot

Treat, J. R., Tumbas, N. S., McDonald, S. T., Shinar, D., Hume, R. D., Mayer, R. E., Stansifer, R. L., & Castellan, N. J. (1979). *Tri-Level Study of the Causes of Traffic Accidents*. https://deepblue.lib.umich.edu/handle/2027.42/64993

U.S. Department of Transportation. (2021). *Automated Vehicles Comprehensive Plan*. https://www.transportation.gov/sites/dot.gov/files/2021-01/USDOT_AVCP.pdf

Wendt, Z., & Cook, J. S. (2018). *Saved by the Sensor: Vehicle Awareness in the Self-Driving Age*. Retrieved June 9, 2022, from https://www.machinedesign.com/mechanical-motion-systems/article/21836344/saved-by-the-sensor-vehicle-awareness-in-the-selfdriving-age

Editor's Note:

This case was selected for inclusion in the journal as the EDSIGCON 2022 Best Case The acceptance rate is typically 2% for this category of case based on blind reviews from six or more peers including three or more former best papers authors who did not submit a case in 2022.

Teaching Case

Cybersecurity Assessment for a Manufacturing Company Using Risk Registers: A Teaching Case

Dr. Jim Marquardson
jimarqua@nmu.edu

Dr. Majid Asadi
masadi@nmu.edu

Northern Michigan University
Marquette, MI 49855, USA

Abstract

This case asks information systems analysts to assess the cybersecurity posture of a manufacturing company. The exercise works well as a group activity in an information systems course that addresses cybersecurity controls. The case introduces guidance from the National Institute of Standards and Technology, and learners develop work products consistent with the standards. The narrative provides high-level summaries of relevant cybersecurity standards. The case is based on a real company and actual projects, but the company name and specific details have been fictionalized and made more abstract to make this case relevant even when specific technologies evolve. Through this experience, students will learn the importance of a defense-in-depth strategy for securing information systems.

Keywords: cybersecurity controls, risk management, teaching case, manufacturing cybersecurity

1. INTRODUCTION

Organizations confront new cybersecurity risks every day in today's computerized world. They mitigate risks using a cybersecurity plan. Information systems professionals must understand cybersecurity concerns when designing and evaluating systems because the confidentiality, integrity, and availability of information systems must be protected. Malicious actors, whether inside or external to an organization, can try to steal company secrets, hold systems hostage for ransom, deface websites, corrupt data, try to crash systems, and more. No single process or technology can be implemented to protect information systems. Rather, multiple measures must be employed to protect systems with a defense-in-depth strategy.

In this teaching case, the cybersecurity posture of a manufacturing company is evaluated. The case is based on a real company and actual projects, but the company name and specific details have been fictionalized. The case introduces a risk register as a risk management tool for identifying and assessing potential cyber risks.

2. ABOUT ACME

ACME designs and manufactures tables and chairs for use at schools, weddings, conferences, and organizations that require easy setup and teardown of seating arrangements. The company invested significant money into research and design (R&D) to develop tables and chairs that are both sturdy and light. ACME can produce tables that weigh significantly less and are more durable than similarly sized particleboard-based tables by using a patented combination of wood,

metal, and plastic materials. The use of several materials also insulates them to some degree from fluctuations in the price of raw materials. ACME can charge a premium for its products with the promise that they will last for years.

ACME realizes that it must continue to innovate to stay relevant in the market. Patents on some of their existing products will expire shortly. Competitors are designing tables and chairs just different enough to avoid patent infringement. It has been increasingly difficult for ACME to charge a premium for its products. Company officers have decided to invest significant capital into designing a new table and chair. ACME would also like to expand its product line with lecterns. The new tables, lecterns, and chairs are codenamed the T1001, L1001, and C1001.

Embarking on the design of new products has caused ACME to reflect on its cybersecurity posture. If competitors were to infiltrate the company's systems and steal the R&D documents, ACME fears that the business could be ruined. ACME would like to conduct a full cybersecurity assessment. ACME would like its cybersecurity posture documented in a risk register. Risk registers and their components will be described in the following section.

3. CYBERSECURITY FRAMEWORKS

There are many cybersecurity frameworks that integrate industry standards and best practices to help organizations manage their cybersecurity risks (Taherdoost, 2022). While all these frameworks aim to protect data and contribute to a stronger security posture, they also have their own unique characteristics. Two popular frameworks are the Cyber Security Framework published by the National Institute of Standards and Technology (NIST) and ISO 27001 published by the International Organization for Standardization (ISO).

NIST created the Risk Management Framework and Cybersecurity Framework to help U.S. federal agencies and private organizations better manage cybersecurity risk. NIST's frameworks are open and readily accessible to the public. NIST documents often remind readers that the content is to be used to guide risk management processes rather than serve as a checklist of best practices. Therefore, organizations must think critically and use sound judgment when adopting NIST's guidance.

ISO 27001 is an internationally recognized standard that concentrates on security in information systems management (ISM). It is possible for organizations to become ISO 27001 certified through formal audits. Small- and medium-sized businesses and startups usually start their cybersecurity plans with NIST and work their way up to ISO 27001 as they scale (Alshboul & Streff, 2015).

ACME is a mid-size business based in the United States and would like to become a supplier to the United States Department of Defense. Department of Defense suppliers must comply with NIST standards. Therefore, the remainder of this case focuses on NIST documents and recommendations. The remainder of this section describes relevant NIST documents and how they can aid in the risk management process and development of risk registers. NIST makes these documents available without cost on its website, https://csrc.nist.gov/publications.

NISTIR 8286: Risk Registers

Column	Description
ID	A sequential number that identifies a specific risk
Priority	The relative criticality of the entry (e.g., low, medium, or high)
Risk Description	A brief overview of the risk, often stated in terms of cause and effect
Risk Category	A set of categories consistent with other risk registers in an organization
Likelihood	Probability of the event happening (1=low, 10=certain)
Impact	Consequences if the event happened (1=negligible, 10=catastrophic)
Exposure Rating	A multiplication of likelihood and impact
Risk Response Type	Approach to risk (i.e., accept, transfer, mitigate, avoid)
Controls	Technical, operational, managerial, or physical controls that mitigate threats

Table 1: Risk Register Columns

Organizations use risk registers to document threats, the likelihood of threats occurring, threat severity, and controls put in place to mitigate threats. NIST Interagency or Internal Report (NISTIR) 8286, *Integrating Cybersecurity and Enterprise Risk Management (ERM)* contains guidance on creating risk registers (Stine, Quinn, Witte, & Gardner., 2020). Table 1 summarizes

key risk register columns. An example table with sample records is included in Appendix A.

NIST 800-30: Threat Identification

NIST Special Publication 800-30, *Guide for Conducting Risk Assessments*, contains a taxonomy of threat sources and events (Joint Task Force Transformation Initiative, 2012). Appendices D and E in the NIST document list dozens of threat sources and sample threat events. Organizations can refer to NIST 800-30 to determine if they omitted relevant threats. Populate this information in the "Risk Description" column of the risk register. The list below summarizes threat sources and gives examples of each threat:

- Adversarial
  - Outside hacker (e.g., a hacker with no affiliation with the organization)
  - Trusted insider (e.g., a company executive)
  - Privileged insider (e.g., an authorized information technology administrator)
  - Competitor (e.g., another company)
  - Nation-state (e.g., state-sponsored hackers)
- Accidental
  - User (e.g., filing clerk)
  - Administrator (e.g., database administrator)
- Structural
  - Computer network outage (e.g., power supply failure)
  - Temperature controls failure (e.g., overheating in the data center)
  - Operating system failure (e.g., memory leak consuming all resources)
- Environmental
  - Natural disaster (e.g., flood)
  - Power outage (e.g., long-term outage due to major disaster)

The following are example threat events consistent with NIST 800-30 in Appendix E:

- Successful phishing attack: A competitor pretends to be a supplier and obtains detailed product specifications from the company.
- Successful denial of service attack: An adversary points a botnet at the company website to overwhelm the web server with requests.
- Earthquake at company headquarters: An earthquake disrupts power and network connectivity.

NIST 800-39: Risk Responses

NIST Special Publication 800-39, *Managing Information Security Risk: Organization, Mission, and Information System View* contains a "Responding to Risk" section that describes how risk can be accepted, avoided, mitigated, or transferred (Joint Task Force Transformation Initiative, 2011). These response types map to the "Risk Response Type" column in the risk register. Key risk responses are summarized below.

- Risk Acceptance: An organization chooses to engage in an activity if the risk is within its tolerance. For example, the risk of damage from a tornado can be low or high depending on a data center's geographic location. A company might accept the risk of tornado damage if it is in a tornado-prone area and the organization has a high risk appetite. An organization might accept the risk if the likelihood of a tornado is low and the organization has a low risk appetite.
- Risk Avoidance: An organization chooses not to engage in an activity because it is above its risk tolerance. For example, a company might decide not to co-locate its equipment in a shared data center because of privacy concerns.
- Risk Mitigation: Actions to reduce risks to an acceptable level. Common cybersecurity controls, like requiring multifactor authentication, can help to mitigate risks. Risk mitigation efforts are the most common entry on risk registers.
- Risk transfer: Organizations shift risk to another party. In cybersecurity, risk transfer is often implemented through buying cybersecurity insurance. An organization might implement security controls but still purchase cybersecurity insurance because the cost of a major data breach could be very high.

NIST 800-53: Controls

NIST Special Publication 800-53, *Security and Privacy Controls for Information Systems and Organizations*, details different controls that can be implemented to manage risk (Joint Task Force Interagency Working Group, 2020). This can be used to populate the "Risk Response Description" column of the risk register. Sample controls are listed below.

- All Wi-Fi access points require the use of a modern encryption protocol.
- Vulnerability scans will be run on all internal systems quarterly.
- All employees receive acceptable use policy training on an annual basis.
- A wireless guest network will be segmented from the employee network.
- An electronic badging system controls access to all buildings.

4. ACME'S CYBERSECURITY POSTURE

Security of the R&D process is of paramount importance to ACME. The product designs contain proprietary information that will be patented. Corporate competitors and suppliers should not have access to the design specifications until ACME files patents. The public should not see prototypes of the T1001, L1001, and C1001 until the models are ready for purchase. To protect the R&D process, ACME has instituted several safeguards. The company president and the Chief Information Security Officer (CISO) review these safeguards annually.

Campus Security

The ACME headquarters campus is open to the public. People can park in the company parking lot without passing through security. All guests must check in with the reception desk in the main building. Guests present photo identification and sign a log maintained by a security guard. Once guests are signed in, they must wear a sticker that identifies them as guests. Company employees escort visitors for the duration of their stay on the company campus. ACME issues smart cards with photo identification to all employees. Company policy dictates that employees present their smart cards whenever they are on campus. A central system controls access to buildings using smart cards and electronic locks.

The campus contains three buildings. The main building has areas for hosting guests, product demonstrations, and offices for office workers such as accountants, information technology staff, and company officers. The manufacturing building is the largest building on campus. It holds raw materials, finished goods, and the machinery required to build the products. The third building is the R&D building. Discussion of new designs, consumer research, prototype development, and testing is confined to the R&D building.

The R&D building has two entrances. The first entrance is a door for employees that automatically locks when closed. Employees must swipe a smart card on a smart card reader which grants them access. The door has a motion detector inside the building that automatically unlocks when an employee exits. The second door is a large roll-up door used for large materials and machinery. This door is locked from the inside with a padlock. Only two employees—the chief of R&D and the company president—have keys to the padlock. Bollards surround the R&D building. A water-based fire sprinkler system protects the R&D building from catastrophic fire damage.

Security cameras monitor the interior of the main campus building and the manufacturing building. No cameras are in use inside the R&D building. Security cameras also monitor the exterior of each building on campus. The camera data feeds are sent to the data center, to a server kept in a large utility closet in the main building. Policy prohibits the use of photographic equipment such as smartphones with built-in cameras in the R&D building.

Visitor Policies

ACME implements a strict visitor access policy since visitors can steal intellectual property, collect information, become injured in hazardous manufacturing areas, or threaten the safety of employees and other visitors. All suppliers, contractors, and delivery personnel are subject to this policy. Controls such as requiring an appointment, check-in, check-out, visitor badges, and guest internet network are part of this policy. No visitors are allowed into ACME's R&D building unless authorized by a department manager. This rule includes the company employees during off-duty hours too. The requests for permission to enter the building must be made at the front office. Personal visitors, including friends and family, are not permitted to access the building during or outside normal business hours. An appropriate associate must escort the visitor to the building.

Authentication and Authorization

Once a year, the company president reviews smart card access logs. The logs contain the employees' information and the location where access was attempted. The president manually scans the logs for access that might be unusual. The president also determines if access is appropriate for all employees based on their job duties. If a change to access levels is necessary, the president submits an access change request to the head of information technology operations who then makes the necessary change.

Technology for Research and Design

The engineers responsible for R&D do their work using Windows computers. The computers

connect to the internet to allow the employees to research existing patents, price materials, and conduct market research. The internet connection is separate from the internet connection used by the other buildings on campus. The network in the R&D building is segmented from the rest of ACME's network. The computers inside the R&D building can access the computing infrastructure inside the same building or the internet, but no other computer on ACME's campus. All computers' external USB ports are disabled unless authorized by a department manager. The computers employ screen-saver passwords and privacy filter screens.

The floor plan is very open and not conducive to having a dedicated telecommunications closet. The internet service provider's cable modem sits on a shelf in a corner of the R&D building. The cable modem is connected to a router that only has wired connections. The router connects to a unified threat management (UTM) device. The UTM device is then connected to an unmanaged switch. All workstations in the R&D building connect to the unmanaged switch. The workstations are manually configured with IP addresses to point to the UTM appliance for their default gateway. The UTM provides basic malware prevention, intrusion detection, and web filtering. Full disk encryption is enabled on all computers in the R&D building. Though most computers at ACME have users authenticate with a central Active Directory server, the strict network segmentation of the R&D building prevents those computers from accessing the central Active Directory infrastructure. Employees log in to the workstations in the R&D building with a local account using a username and password. The employees must update their passwords every three months.

All computers in the R&D building can access a central file server. The server allows anybody who knows the IP address to connect with full access to read and write files. Because physical access to the R&D lab is controlled, no authentication is required to access the file server. Employees use the file server to share files and collaborate. The file server is backed up using an external hard drive weekly, just like the workstations.

Employees make weekly full backups of blueprints, market research, and other files critical to the R&D process. Employees make backups by copying files to external hard drives. They store the external hard drives in a locked cabinet in the R&D building. Again, only the chief of R&D and the company president have keys to the cabinet. They maintain three weeks of backups.

R&D employees use computer-aided drafting tools, email, and basic office productivity software for the majority of their work. At times, they need the ability to install software, so they have been given administrative access to their computers. Like all ACME employees, R&D employees must still abide by the company's acceptable use policy, which states that they should not install software without proper licensing. Penalties for violating the acceptable use policy include censure and termination of employment. The R&D employees receive cybersecurity awareness training annually to prevent and mitigate cyber-attack risks.

5. STUDENT ASSIGNMENT INSTRUCTIONS

Assess ACME's cybersecurity posture by creating a risk register. Develop the risk register by evaluating cybersecurity threats and controls that could help mitigate those threats. Figure 1 shows the cybersecurity framework used at ACME for cyber risk assessment.

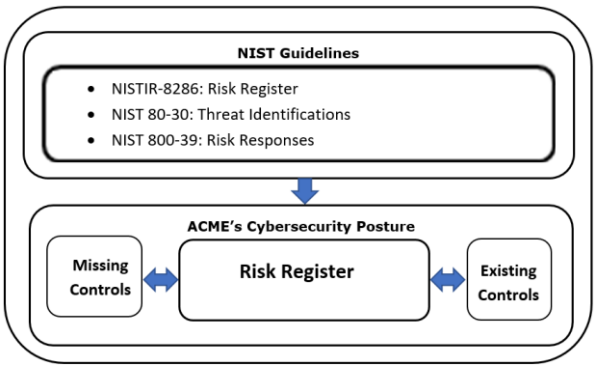


Figure 1: ACME's Risk Assessment Using a Risk Register

Professionals can employ a control-driven or event-driven approach to complete the risk register.

Control-driven Approach

In a control-driven approach, the existing controls are documented, then the adverse cybersecurity events that those controls mitigate are described. For example, a company might have fencing around its campus perimeter. Fencing is a known physical control, so it is clear that the company wants to keep people out. In the risk description column, document the specific reason why people need to be kept out. The control-driven approach largely focuses on what organizations are already doing.

Threat-driven Approach
Instead of focusing on controls already in place in the control-driven approach, the threat-driven approach emphasizes thinking about what could go wrong. Once the adverse events have been described, controls can be identified. For example, employees might be able to install software on their computers which could lead to them installing malware. A control could be restricting the ability to install software by non-administrative users. Gathering cybersecurity professionals with a variety of expertise together to brainstorm can help to find the essential controls and to identify the potential threats in the company. The threat-driven approach helps identify missing controls.

As company environments grow increasingly complex, professionals utilize proven methodologies capable of guiding a comprehensive, systematic assessment of cybersecurity threats. For example, some common methodologies such as OCTAVE (Operationally Critical Threat, Asset, and Vulnerability Evaluation™) and TVA (Threat-Vulnerability-Asset) have been utilized to facilitate the identification of critical IT resources, the threats to those IT resources, and the identification of related system vulnerabilities (Mejias, Shepherd, Fronmueller, & Huff, 2019).

Risk Ratings
Once the risk description and controls have been documented, the remaining columns in the table can be completed.

1. For ID, enter a sequential number.
2. For priority, enter low, medium, or high based on your subjective assessment.
3. For risk category, choose confidentiality, integrity, or availability. Assume that ACME had formally adopted these categories.
4. For likelihood, choose a number from 1-10, 1 representing the lowest and 10 representing the highest likelihood.
5. For impact, choose a number from 1-10, 1 representing the most negligible impact, and 10 representing the most critical.
6. For exposure rating, multiply likelihood and impact.
7. For risk response type, determine if the risk response type is to accept, transfer, mitigate, or avoid.

Key Questions
It is important to be thorough on a risk register. Failure to identify relevant threats may lead organizations to develop insufficient controls. Failure to document existing controls might make some organizations believe that current practices are a waste of resources. Developers of risk registers should ask themselves the following three questions until they feel satisfied that no significant items are missing.

- What threats exist that have not yet been documented?
- What controls does the organization employ that have not yet been documented?
- What additional controls should we put in place to mitigate risk?

As a risk management tool, the risk register helps cybersecurity professionals and organization leaders agree on the proper approach to cybersecurity.

6. CONCLUSION

Cybersecurity is a process, not an end state. Part of that process requires professionals to evaluate threats and controls that mitigate those threats. This paper asks students to assess the cybersecurity posture of a manufacturing company via risk registers. Risk registers help organize and prioritize cybersecurity resources. As a communication tool, risk registers help information technology professionals and organization leaders reach a shared consensus on the role of cybersecurity in achieving organizational objectives. NIST provides detailed guidance on completing risk registers and other risk management processes.

7. REFERENCES

Alshboul, Y., & Streff, K. (2015). *Analyzing information security model for small-medium sized businesses.* AMCIS 2015 Proceedings.

Burke, A. (2011). *Group work: How to use groups effectively*. Journal of Effective Teaching, 11(2), 87-95.

Joint Task Force Interagency Working Group. (2020). *Security and Privacy Controls for Information Systems and Organizations* (Revision 5). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-53r5

Joint Task Force Transformation Initiative. (2011). *Managing Information Security Risk: Organization, Mission, and Information*

System View (NIST SP 800-39; 0 ed., p. NIST SP 800-39). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-39

Joint Task Force Transformation Initiative. (2012). *Guide for conducting risk assessments* (NIST SP 800-30r1; 0 ed., p. NIST SP 800-30r1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-30r1

Mejias, R. J., Shepherd, M. M., Fronmueller, M., & Huff, R. A. (2019). *Using Threat Vulnerability Asset (TVA) Methodology to Identify Cyber Threats and System Vulnerabilities: A Student Field Project Case Study.* Business Education Innovation Journal, 11(1).

Stine, K., Quinn, S., Witte, G., & Gardner, R. K. (2020). *Integrating Cybersecurity and Enterprise Risk Management (ERM)*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8286

Taherdoost, H. (2022). *Understanding Cybersecurity Frameworks and Information Security Standards—A Review and Comprehensive Overview.* Electronics, 11(14), 2181

Appendix A

Sample Risk Register Entries for Students

The following example show sample entries for a risk register. This can be provided to students in a template.

ID	Priority	Risk Description	Risk Category	Likelihood	Impact	Exposure Rating	Risk Response Type	Controls
1	Low	Employee leaks records to the media resulting in reputational harm	Confidentiality	2	4	8	Mitigate	All employees must sign a non-disclosure agreement that a lawyer has vetted
2	High	File server hard disk fails leading to data loss	Integrity	2	9	18	Mitigate	- Files will be backed up to the cloud nightly - Restoration from backup tested quarterly
3	Medium	Customer data breached by hackers leading to costs to contain the breach and legal fallout	Confidentiality	2	7	14	Transfer	- Cybersecurity insurance policy purchased - Data breach playbook created

Guide for Each Column
Assume that ACME has adopted the following standards for completing its risk register.

- ID: Sequential numbering
- Priority: Low, Medium, or High
- Risk Description: What can go wrong
- Risk Category: Confidentiality, Integrity, or Availability
- Likelihood: 1-10 (1=low, 10=high)
- Impact: 1-10 (1=10, 10=high)
- Exposure Rating: Likelihood multiplied by impact
- Risk Response Type: Mitigated, accepted, transferred, or avoided
- Controls: Technical, operational, managerial, or physical controls that mitigate the threat

Digitally Prepared for Success? Technology Skills of Incoming First-Year College Students

Elizabeth McCarron (retired)
elizabeth.c.mccarron@gmail.com

Mark Frydenberg
mfrydenberg@bentley.edu

Computer Information Systems Department
Bentley University
Waltham, MA 02452, USA

Abstract

Students today are often assumed to be digitally literate and prepared for college, but that is not always the case. Introductory technology courses at the college level provide students with the technology and digital skills necessary for them to succeed in college. This study, conducted at a small New England "business school" that also offers liberal arts degrees, sought to determine which digital skills most incoming first-year college students have. The study examines which digital skills and competencies students acquire before entering college, and if those skills are adequate to achieve academic success in college. Guiding the study are these research questions: RQ1) What digital devices are students using and with what frequency? RQ2) What digital literacy competencies (DLCs) do incoming first-year students possess? RQ3) How do first-year students self-rank their DLCs, and how does that compare to individual task self-ranking? RQ4) What DLCs do incoming first-year students perceive as the most important to know, and do they have those skills? Results showed that most K-12 students are learning on Macs or Chromebooks and use the Google Suite of applications. They may need to switch to Windows machines and/or Microsoft Office products upon entering college. Almost 16% reported taking no high school courses covering computer topics. Only 34% of incoming first-year college students consider themselves "tech savvy" and most rank themselves at the beginner or intermediate level on digital literacy competencies. They believe the three most important skills to know are spreadsheets, programming, and word processing.

Keywords: Digital Literacy, Information Technology Literacy, Competencies, Digital Skills, First-Year College Students

1. INTRODUCTION

Most students entering 4-year colleges as first-year students are considered Generation Z, born from 1997 onward (Dimock, 2019). Technology has been ubiquitous in their lives - there has always been an Internet, a World Wide Web, smartphones, and WiFi. They have always had access to social media, on-demand entertainment, and virtual reality. They are constantly connected via mobile technology, desire instant gratification, and expect immediate response time. According to the Pew Research Center (Dimock, 2019), 95% of teens have access to a smartphone, and 45% say they are online "almost constantly." Gen Z are also the first of the digital natives.

The term "digital natives" (Prensky, 2001) refers to "native speakers" of the digital language of technology who have been surrounded by and immersed in technology since birth. This immersion infers that Gen Z are fluent in digital skills across a wide range of technologies and are proficient in creating and using digital content (Smith, Kahalke, & Judd, 2020). It is easy to assume that these digital natives "possess knowledge and skills that allow them to handle information and communication technologies (ICT) tools in a 'natural' way" (Sorgo et al., 2016). Yet access to these tools and technologies does not guarantee digital literacy.

Digital Literacy can be defined as "the ability to use information and communication technologies to find, evaluate, create, and communicate information, requiring both cognitive and technical skills" (Digital Literacy Task Force, 2013). It encompasses various competencies including basic computer skills, network literacy, digital problem solving, information literacy, and media literacy (Vanek, 2020).

Many people assume that because today's college students are digital natives, they are digitally literate and perfectly prepared for the technological rigors of academic coursework; this implies that they do not need a basic "computers 101" course in college. Many post-secondary faculty have experienced that this is not always true. This study sought to determine whether first-year college students are as digitally literate as everyone assumes they are by examining which digital skills and competencies students have acquired before entering college, and if those skills are adequate to achieve academic success in college.

The following sections present a review of relevant literature, describe the methodology used, report the results, and discuss the implications of the findings. A review of the literature explores four main topics: digital native stereotypes; effects of the digital divide; digital literacy skills needed by college students to be academically successful (especially during a pandemic), and digital literacy vs. information technology literacy.

2. DIGITAL NATIVE STEREOTYPES

Much of the literature regarding digital natives references the highly-cited paper by Prensky (2001). He states that "the single biggest problem facing education today is that our Digital Immigrant instructors, who speak an outdated language (that of the pre-digital age), are struggling to teach a population that speaks an entirely new language" (Prensky, 2001). He instigated a "digital native debate" as some educators and institutions panicked over revisiting curricula and increasing technology assets and resources while others advocated more investigation. Bennett, Maton, & Kervin argued that the debate was a form of "moral panic," not "empirically and theoretically informed" (2008). Prensky's paper is not research-based and includes no data or evidence to support his claims. Several studies have refuted Prensky's claims (Evans & Robertson, 2020; Judd, 2018; Smith et al., 2020).

A quantitative study of Chinese teenagers (Li & Ranieri, 2010) shows that while the majority of the participants had personal computers and Internet access at home, their overall performance on an Instant Digital Competence Assessment was "pass", not "good" or "excellent". Results showed large disparities in participant digital competence. Another 2010 study (Kennedy et al., 2010) showed that digital natives are not a homogenous group when it comes to digital competencies; through cluster analysis the researchers identified four distinct groups: power users, ordinary users, irregular users, and basic users.

In a 2016 study of Slovenian university students, Sorgo et al. (2016) found that being a digital native did not predict information literacy. A 2018 literature review study revealed various definitions of digital natives "with no specific clarification or research-based rationale" and "little connection between a student's age and digital skills and increased learning". In addition, "much of the research suggests that students' digital competence may be much lower" than that of their professors (Creighton, 2018). In addition, the concept of a "digital native" does not consider the many students of Gen Z age who lack the same technology knowledge and resources as their peers due to the digital divide.

3. DIGITAL DIVIDE

The digital divide encompasses more than having access to the Internet. Developing digital literacy skills requires access to technological resources. Network inequities include quality of Internet service, affordability, and skills necessary to obtain information (Parker, Santos, & Dancy, 2021). Hardware inequities include aging and broken equipment as well as limited access to computers and peripherals. Software inequities include operating systems chosen, versions of

software used, and the ability to keep software up to date.

Many institutions put substantial resources into providing and upgrading technology, but not necessarily training students in how to best use that technology effectively. "The 'have nots' who, either by choice or circumstance, lack the skills and critical thinking needed to sort through the vast array of information, are excluded from fully participating in their education" (Andreae & Anderson 2012, p. 77) In fact, students who would fall under the digital native moniker "could be 'strangers' in the digital world as a result of disinterest, illiteracy, economic constraints, poor network connectivity, lack of electric power and inadequate practical accessibility" (Adjin-Tettey, 2020, p. 11).

A 2018 study of students at a minority-serving institution showed that they do not arrive at college with the technological skills required for academic success. Courses in computer applications showed efficacy at building those technical skills and were a positive experience for students, who believed the courses were "a necessary part of the college experience that benefits them academically and professionally" (Buzzetto-Hollywood, Wang, Elobeid, and Elobaid, 2018, p. 78).

Another 2018 study at a large Midwestern University in the US found that even with most students owning laptops and smartphones, a digital divide was experienced regarding consistent access to reliable technology. Approximately 20% of respondents experienced issues in accessing technology such as connectivity problems, unreliable hardware, data limits, etc. "Students of lower socioeconomic status and students of color disproportionately experienced hardships, and reliance on poorly functioning laptops was associated with lower grade point averages" (Gonzales, McCorry Calarco, and Lynch, 2020, p. 750).

Reliable computers and Internet connections are integral to accessing colleges' learning management systems. Yet a 2021 study showed that "trouble accessing the internet and connecting to coursework, reliance on tablets or mobile phones, and the need to share devices with others are all more common among students from low-income households and those who are Black or Latinx" (Parker et al., 2021, p. 9). The COVID-19 pandemic and the shift to emergency remote teaching made the digital divide more obvious and the need for digital literacy skills more critical. A study by the Midwestern Education Compact (Jaggars Motz, Rivera, Heckler, Quick, Hance, and Karwisch, Jaggars et al., 2021) showed that inadequate technology resulted in students struggling with the switch to remote learning. Between 16% and 19% of college students reported barriers to online learning including inadequate hardware or Internet connections. The inequities of the digital divide impact low income, first-generation, and minority learners disproportionately, and are a "hindrance to student success" (Buzzetto-Hollywood et al., 2018, p. 77). Colleges must also address the needs of students with varying levels of digital literacy.

4. DIGITAL LITERACY AND INFORMATION TECHNOLOGY LITERACY

Digital Literacy

The JISC (Joint Information Systems Committee) defines digital literacies as "those capabilities which fit an individual for living, learning and working in a digital society". Living, learning, and working in a digital society also involves applying digital literacy skills and using technology tools to solve problems and create solutions. Digital literacy includes the knowledge and skill necessary to create, access, evaluate, use, and share digital information. It is not determined by age, because being able to use a particular technology does not guarantee the "critical thinking or other skills to understand the social implications or risks associated with its use" (Johnson, 2018).

While Gen Z regularly use social media applications (YouTube, Instagram, SnapChat, TikTok, etc.) for personal and recreational purposes, "research shows that young people are not as critically or technically skilled as they need to be in order to be effective digital citizens" (Talib, 2018, p. 56). In addition to digital literacy, the current economy requires students to have Information Technology (IT) literacy.

IT Literacy

Information Technology literacy goes beyond merely using tools and devices, but also requires the ability to do so effectively. Students must have a basic understanding of how their tools and devices work, and how they work together to build solutions. Students must develop collaboration skills while using technologies that they will encounter in the workplace, and know and use these tools responsibly.

IT literacy requires "an understanding of the tools technology provides … and an understanding of the legal, social, economic and public policy

issues that shape the development of the infrastructure and the applications and use of the technologies" (Lynch, 1998). As digital tools are constantly evolving, students need a solid understanding of IT concepts to adapt to an ever-changing toolset. This understanding of fundamental technology concepts enables the transfer of skills from one context to another. For example, the student who knows how to write a formula in Google Sheets or format a document using Google Docs must be able to apply their conceptual knowledge when using Microsoft Excel or Microsoft Word. A student who is proficient in using a computer running macOS needs to be able to translate those skills to navigate a computer running Windows. Understanding fundamental technology concepts enables students to solve more complex problems with technology, to select the appropriate tools with which to build their solutions, and to be able to explain their choices and results to others.

Students must be creative thinkers, able to research information and apply their knowledge to new settings. Yet studies show that college students underestimate the importance of critical and computational thinking (Wing, 2006), demonstrating a deficiency when analyzing or navigating information (Andreae & Anderson, 2012).

5. METHODOLOGY

This descriptive research study sought to determine which digital skills and competencies first-year students acquire before entering college, and if those skills are adequate to achieve academic success in college. To explore this, the following research questions were employed:

- RQ1: What digital devices are incoming first-year students using and with what frequency?
- RQ2: What digital literacy competencies (DLCs) do incoming first-years possess?
- RQ3: How do incoming first-year students self-rank their DLCs, and how does that compare to individual task self-ranking?
- RQ4: What DLCs do incoming first-years perceive as the most important to know, and do they have those skills?

Participant Characteristics

Participating students completed this survey during the first three weeks of the Fall 2020 semester, as part of an "introduction to information technology" course. Demographic information collected from respondents is summarized in Appendix A, Table 1 which includes the summary demographics for the sample (n=440) compared to those of the population (n=936) as provided by the University Factbook compiled by the Institutional Research department. The sample population is fairly descriptive of the actual population based on all categories.

Data Collection

To answer the research questions, the following information was sought:

- Demographics including age, gender, ethnicity, first-generation status, location of high school
- Hardware information including college and high school computers used, how often devices are used, which devices are used for which tasks
- High school information including IT topics learned as part of a course, programming languages studied
- Self-ranking of whether they consider themselves "tech-savvy"
- Self-ranking of DLC application tasks (word processing, spreadsheets, presentation software, databases, email, operating system, cloud storage, web browser, digital media, other tasks)
- The three most important tech skills that students feel they need for success in college and career

To obtain the required data, the authors created an online survey to send to first-year students.

Instrument

The survey (see Appendix B) was built in Qualtrics and was easy for respondents to use via computer or smartphone. Question types included multiple choice, multiple answer, open-ended, and scaled responses using Likert-type scales.

Non-demographic questions were reviewed by several IT professors and authors of IT textbooks to affirm the researcher's beliefs of beginner, intermediate, and expert tasks were accurate. Digital Literacy Competency questions were created after reviewing the literature and standards created by JISC (2014), (IEEE 2021) , and information literacy value rubrics by AACU (2013).

Sampling Procedures

After IRB approval, online surveys were distributed to all students enrolled in a required introductory IT course at Bentley University, a small 4-year business university in New England in the Fall semester of 2020 (n = 608) during the

first semester of fully online instruction due to COVID-19. Students self-selected into the survey by agreeing to participate (n = 565, 92.93%). Respondents received no remuneration for participation.

Researchers excluded 168 records: the 43 non-participants, 60 non-completed surveys, 43 of respondents less than 18 years of age (excluded due to IRB requirements), 4 non-first-year students, and 18 transfer students who took an IT course at a previous college (3 records fell in multiple excluded groups). The analysis proceeded with 440 records (47% of population).

Sample Size, Power, and Precision
With a population of 936, sample size calculations show that 273 or more completed surveys were required for a confidence level of 95% that the real value is within ±5% of the surveyed value. This study's sample size of 440 meant that there is a 95% chance that the real value is within ±3.33% of the surveyed value (*Sample Size Calculator*, n.d.).

Data Analysis
Data were downloaded in Excel format, uploaded into SPSS for analysis, and re-coded from text to numeric values as necessary. Measures included demographics; high school computer and course information; devices and apps used and frequency of use; self-rankings of various IT tasks; agree/disagree statements; beginner, intermediate, and expert tasks for each of nine applications (word processing, spreadsheets, presentations, databases, email, operating system, cloud storage, web browser, and digital media); and technology skills beliefs.

6. RESULTS

Demographic Data
Demographic variables were tested using Pearson's correlations against the "tech savvy" variable and students' self-ranking on eleven digital literacies. There were no significant correlations found.

RQ1: What digital devices are incoming first-year students using and with what frequency?
Most students (n=424, 96.36%) planned to use a laptop to attend online classes. They expected to log in to classes from their dorm room (n=327, 74.32%) or off-campus home (n=101, 22.95%), using wireless connections (n=396, 90.0%). The majority (n=272, 61.81%) used a Windows computer, followed by those using a Mac (n=166, 37.72%).

In high school, the majority of students used a Mac (n=179, 40.68%), followed by a Chromebook (n=122, 27.73%), and a Windows machine (n=104, 23.64%). Most students used a home computer to complete homework (n=416, 95.55%), 18 (4.09%) used a computer in a lab at school, and 4 (0.90%) used a computer at a public library.

In college, Windows machines became the most used (n=269, 61.14%), followed by Mac (n=167, 37.95%). At college, the majority of students (n=203, 46.14%) purchased a computer on their own, followed by those who purchased through the school (n=166, 37.73%). Seventy students (15.90%) used a computer that they already had at home.

Many students (n=409, 92.95%) use an iPhone daily; 29 students (6.59%) use an Android or other smartphone daily. A majority of students use a laptop (n=420, 95.45%) daily. More than half rarely or never use other devices including smart watches (n=266, 60.45%), tablets (n=258, 58.64%), Raspberry Pis (n= 373, 84.77%), VR Headsets (n=365, 82.95%), or smart speakers (n=251, 57.05%).

Students overwhelmingly use mobile phones to send text messages (n=432, 98.18%), listen to music (n=423, 96.14%), take (n=430, 97.73%) and edit (n= 370, 84.09%) photos, and post to social media (n=420, 95.45%).

Laptop/Desktop computers are the device of choice for sending email (n=415, 94.32%), visiting websites (n=409, 92.95%), using a search engine (n=406, 92.27%), editing a document (n=430, 97.73%), and making online purchases (n=395, 89.77%). Students use both laptops/desktops (n=339, 77.04%) and mobile phones (n=330, 75.0%) to watch videos.

RQ2: What digital literacy competencies (DLCs) do incoming first-years possess?
To answer RQ2, students were asked which DLC topics were covered in courses they attended in high school. The results are shown in Table 1.

Sixty-nine students (15.68%) reported no computer topics learned in high school courses. For the 55 students who took programming, the top three languages were Java (n=30, 54.55%), JavaScript (n=21, 38.18%), and Python (n=14, 25.45%).

Topics Covered in High School Courses	*n*	%
Email	253	57.5
Using the WWW	169	38.41
Social Media	166	37.73
Spreadsheets	151	34.32
Presentation Software	135	30.68
Database Software	96	21.82
Digital Photography	90	20.45
Word Processing	74	16.82
Programming	55	12.5
Digital Video Production	50	11.36
Digital Audio Production	18	4.09

Table 1. Topics covered in high school courses taken

For social media apps, a majority of students use Instagram (n=282, 64.09%), Snapchat (n=309, 70.23%), and YouTube (n=141, 32.05%) multiple times a day. Fewer than 40% use Twitter (n=166, 37.73%). Most students never use LinkedIn (n=329, 74.77%) or Reddit (n=291, 66.14%). There was a fairly even split between students who never use TikTok (n=135, 30.68%) and those who use it multiple times a day (n=154, 35.00%).

RQ3: How do incoming first-year students self-rank their DLCs, and how does that compare to individual task self-ranking?

When asked to what extent they agree with the statement "I consider myself to be tech-savvy," 150 students (34.09%) agreed/strongly agreed; 165 were neutral (37.50%), and 115 disagreed/strongly disagreed (26.14%).

As shown in Table 2, most students ranked themselves as Beginner or Intermediate on all digital tasks before coming to college; web browsing was the only task where beginners, not experts, were the minority.

With most students self-ranking as Beginners or Intermediates, the knowledge of students in those categories was explored further. For six specific skills in each of the first nine digital literacy task categories, students ranked their knowledge of the skill as "don't know what it is", "heard of it, don't know how to do it", or "know how to do it". The skills list is provided in Appendix A, Table 2.

Combining the first two response options allowed comparison of how many students could not perform a task versus how many could. To answer the research question, the self-ranking (Beginner or Intermediate) for each of the nine digital literacy task categories was compared to the self-ranking of the two same-level tasks in the category to determine whether self-ranked beginners could complete the beginner level tasks and self-ranked intermediates could complete the intermediate level tasks.

For self-ranked Beginners, the majority could complete both beginner-level digital media editing tasks and one of the beginner-level database tasks as seen in Table 3.

For self-ranked Intermediates, the majority could complete all intermediate-level tasks in 4 of 6 categories, and only one intermediate-level task in Word Processing and Web Browsing as seen in Table 4.

Skill	Beginner		Intermediate		Expert		Total
	n	%	*n*	%	*n*	%	*n*
Word Processing	201	46.2	**205**	**47.13**	29	6.67	435
Spreadsheet Software	**222**	**51.2**	206	47.58	5	1.15	433
Presentation Software	177	40.88	**226**	**52.19**	30	6.93	433
Database Software	**327**	**76.4**	99	23.13	2	0.47	428
Email Software	112	25.75	**251**	**57.70**	72	16.5	435
Operating Systems	162	37.33	**217**	**50.0**	55	12.6	434
Cloud Storage	149	34.33	**228**	**52.53**	57	13.1	434
Web Browsers	90	20.79	**225**	**51.96**	11	27.2	433
Digital Media Editing	**216**	**49.7**	174	40.09	44	10.1	434
Online Collaboration	168	38.80	**199**	**45.96**	66	15.2	433
Online Calendar	**204**	**46.9**	183	42.07	48	11.0	435

Table 2. Self-ranking skills as beginner, intermediate, or expert.

		Beginners	Can		Cannot	
Skill	**Task**	***n***	***n***	***%***	***n***	***%***
Spreadsheets	Create Chart	220	102	46.36	**118**	**53.64**
	Sum a Range	221	46	20.81	**175**	**79.19**
Databases	Create Table	324	158	48.77	**166**	**51.23**
	Add Data to Table	323	**191**	**59.13**	132	40.87
Digital Media Editing	Crop Photo	213	**194**	**91.08**	19	8.92
	Flip Image	213	**195**	**91.55**	18	8.45

Table 3. Self-ranked beginner knowledge of beginner tasks.

		Intermediates	Can		Cannot	
Skill	**Task**	***n***	***n***	***%***	***n***	***%***
Word Processing	Add Page Numbers	205	**184**	**89.8**	21	10.24
	Use Format Painter	205	47	22.9	**158**	**77.07**
Presentation SW	Use a Theme	221	**208**	**94.1**	13	5.88
	Change Layout	221	**177**	**80.1**	44	19.91
Email	Add an Attachment	250	**240**	**96.0**	10	4.00
	Reply All	250	**201**	**80.4**	49	19.60
Operating System	Install/Uninstall	213	**161**	**75.6**	52	24.41
	Backup Data	213	**110**	**51.6**	103	48.36
Cloud Storage	Share Link	222	**145**	**65.3**	77	34.68
	Collaboratively Edit	223	**141**	**63.2**	82	36.77
Web Browser	Accept Cookies	222	**155**	**69.8**	67	30.18
	View HTML Source	222	70	31.5	**152**	**68.47**

Table 4. Self-ranked intermediate knowledge of intermediate tasks.

RQ4: What DLCs do incoming first-years perceive as the most important to know, and do they have those skills?

In an open-ended question, respondents were asked what three technology skills they thought were the most important to help prepare them for their college education and their future careers. Responses were loaded into a text analyzer (https://www.online-utility.org/text/analyzer.jsp) which provided a word count of each individual word (n=880). Words occurring less than 5 times were eliminated, as were extraneous words (articles, pronouns, adverbs, etc.) that did not reflect a digital literacy. The remaining 90 words were reviewed for similar meanings (e.g., "spreadsheet", "spreadsheets", "Excel") and combined where appropriate. The 10 highest word counts were: spreadsheets (n=291), programming (n=128), word processing (n=110), computer (n=103), security/privacy (n=96), email (n=87), presentations (n=81), Microsoft/Windows (n=64), data/database (n=40), and web/websites (n=29).

7. DISCUSSION

We theorized that while the students surveyed fell into the age range of digital natives, not all arrived at college with the digital literacy skills necessary to succeed. This section provides further discussion of the research questions, suggests implications of these results for practice, and recognizes limitations of this study.

RQ1: What digital devices are incoming first-year students using and with what frequency?

As reported, the majority of students (n=272, 61.81%) are using a Windows computer in college, followed by those using a Mac (n=166, 37.72%). In high school, the majority of students used a Mac (n=179, 40.68%), followed by a Chromebook (n=122, 27.73%), and then Windows machines (n=104, 23.64%). So, three-

quarters of incoming first-year students used a non-Windows machine before college. The number of Mac users remained relatively the same between high school and college, which means Chromebook users switched to Windows machines for college. The digital divide is evident for any student switching to a machine running a different operating system than they are used to using. Chromebook users will not have experience with Windows OS and Office software. At this school, Macs need to access a remote server to run Windows versions of apps, forcing students to use different versions of software. In addition, many students used the Google suite of applications software in high school and are now expected to learn the Microsoft office "equivalents" for completing coursework.

Almost 90% of students planned to use wireless connections for online classes, which would prove to be an issue with Zoom when being required to turn cameras and mics on during classes. Where they obtained their computer could also have ramifications. The 70 (15.90%) students using a computer that they already had at home could have issues with up-to-date software, virus protection, slow speed, not enough RAM, etc. In addition, students may need to share a computer with others in the household.

RQ2: What DLCs do incoming first-years possess?

Nearly 16% of students reported taking no computer-topic courses in high school. Various reasons could include courses not being offered, students not having room on their schedules, or schools not offering topics in which students are interested. As a result, the only skills that many incoming first-year college students have in online research, word processing, preparing presentations, and using email were developed in a "learn by doing" fashion without formal training. This leaves gaps in knowledge and skills.

A majority of students possess social media skills, reporting using Instagram (n=282, 64.09%), Snapchat (n=309, 70.23%), and YouTube (n=141, 32.05%) multiple times a day, which Pew Research Center (2018) reported as the most popular social media apps. Students' near-constant use of smart phones and social media requires "a knowledge of online social networks, how to learn from them and through them, and how to use them to access and disseminate information" (Vanek, n.d.). They are communicating and collaborating, but other DLCs will make a bigger contribution to academic success.

RQ3: How do incoming first-year students self-rank their DLCs, and how does that compare to individual task self-ranking?

The majority of students ranked themselves as Beginner or Intermediates in knowledge of the nine DLCs, corroborating findings of Kennedy et al. (2010) that students are not a homogenous group when it comes to digital literacies. Comparing the self-ranking with the same-level tasks in each category showed that in general, students were ranking themselves accurately. For the Beginner-level tasks, a slight majority (51.23%) could not accomplish the database task of creating a table. Only 21.82% covered databases in a high school course, and they may have been given completed tables to work with since the majority (59.13%) knew how to add data to an existing table. The majority of self-ranked beginners could not accomplish either of the spreadsheet tasks. This may also be due to limited exposure since only 151 (34.32%) of students took a high school course that covered spreadsheets. It is notable that both of these categories were listed in the top ten "most important things to know" results, spreadsheet at #1, and databases at #9. Corroborating Li and Ranieri (2010), we found wide disparity in knowledge for some DLCs.

RQ4: What DLCs do incoming first-years perceive as the most important to know, and do they have those skills?

Of the 10 "most important to know" skills that students identified, 5 are pre-college digital literacy skills - spreadsheets (#1), word processing (#3), email (#6), presentations (#7), and databases (#9) - for which the majority of students ranked themselves as beginners or intermediates. Therefore, these are digital literacy skills in which they know they could become more proficient.

Programming and creating websites were ranked #2 and #10 respectively. As previously reported, only 12.5% of students have prior programming experience and only 14.1% have created a website. So, most students do not have these skills but recognize them as important to know.

Assuming that "computers/Mac/laptop" and "Microsoft/Windows" refer to learning more about their college computer, and perhaps a new operating system, this tracks with more than a third of students (n=162, 36.82%) self-ranking as beginners regarding their computer's operating system. Regarding security/privacy, 64% of students (n=282) agreed/strongly agreed that they were concerned about online privacy,

and less than half (n=197, 44.77%) agreed/strongly agreed that they could explain different ways to protect their systems and information from unethical users. The lack of/desire for knowledge of first-years in these DLCs has implications for colleges and their IT course offerings.

8. LIMITATIONS

The study occurred at one four-year institution which may affect the generalization of findings to a larger population and therefore the external validity. Since the institution is considered a "business school", it may attract significantly different students than would be found at a non-business-focused institution. Students self-selected to participate in the survey via a link in their course management system, after being informed about the survey by their instructor. The respondents could have been influenced by thinking the study was part of the course or required by the instructor. In addition, there is always a chance that self-reported answers are inaccurate or incomplete.

9. CONCLUSION

This study has shown that students may need to learn new hardware, software, and operating system skills when transitioning to college. With almost 16% of students reporting taking no high school courses covering computer topics, colleges cannot make assumptions about student digital literacy. Far from being technology experts, only 34% consider themselves "tech savvy" and most rank themselves at the beginner or intermediate level on DLCs. The digital native stereotype is harmful to students who, due to various circumstances including the digital divide, do not arrive at college with the digital literacy competencies needed to succeed academically. Many will need the scaffolding of a basic IT literacy course to ensure they are introduced to the skills and knowledge they will need to succeed in higher-level courses and persist to graduation. Institutions should capitalize on the fact that students believe spreadsheets, programming, and word processing skills are important to know for their college and professional success and ensure that these skills are offered formally via courses and supported with labs, tutoring, and other resources. The authors recommend that information systems educators become aware of students' limitations in DLCs and that educational institutions provide formal and extracurricular opportunities for students to develop the competencies they are lacking so that they are prepared for future courses that require these skills.

A near-term goal for future research is to replicate this study at different types of institutions (community colleges, large research institutions, etc.) to determine whether similar results are found. For example, unlike results from Buzzetto-Hollywood et al. (2018), this study showed no significant correlation between demographic factors and DLCs; other researchers may find a link.

In conclusion, the DLCs students learn before college may not be the skills they need in college. They need IT courses to help them develop the digital literacies they are lacking and strengthen and improve the ones that they have. The goal should be to develop IT literacy – applying their DLCs to solve problems and create solutions.

10. REFERENCES

Adjin-Tettey, T. D. (2020). Can 'digital natives' be 'strangers' to digital technologies? An analytical reflection. *Inkanyiso: Journal of Humanities and Social Sciences*, *12*(1), Article 1.

American Association of Colleges and Universities. (2013, July). *VALUE Rubrics*. AAC&U. https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-information-literacy

Andreae, J., & Anderson, E. (2012). Re-conceptualizing Access: The New Role of Information Literacy in Post-Secondary Education. *Comminfolit*, *5*(2), 74. https://doi.org/10.15760/comminfolit.2012.5.2.104

Bennett, S., Maton, K., & Kervin, L. (2008). The 'digital natives' debate: A critical review of the evidence. *British Journal of Educational Technology*, *39*(5), 775–786. https://doi.org/10.1111/j.1467-8535.2007.00793.x

Buzzetto-Hollywood, N., Wang, H., Elobeid, M., & Elobaid, M. (2018). Addressing Information Literacy and the Digital Divide in Higher Education. *Interdisciplinary Journal of E-Skills and Lifelong Learning*, *14*, 077–093. https://doi.org/10.28945/4029

Creighton, T. B. (2018). Digital Natives, Digital Immigrants, Digital Learners: An International Empirical Integrative Review of

the Literature. *Education Leadership Review*, *19*(1), 132–140.

Digital Literacy Task Force. (2013). *Digital Literacy, Libraries, and Public Policy*. https://alair.ala.org/handle/11213/16261

Dimock, M. (2019, January 17). Defining generations: Where Millennials end and Generation Z begins. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/

Evans, C., & Robertson, W. (2020). The four phases of the digital natives debate. *Human Behavior and Emerging Technologies*, *2*(3), 269–277. https://doi.org/10.1002/hbe2.196

Gonzales, A. L., McCrory Calarco, J., & Lynch, T. (2020). Technology Problems and Student Achievement Gaps: A Validation and Extension of the Technology Maintenance Construct. *Communication Research*, *47*(5), 750–770. https://doi.org/10.1177/0093650218796366

IEEE. (2021). IEEE Standard for Digital Intelligence (DQ)–Framework for Digital Literacy, Skills, and Readiness. *IEEE Std 3527.1-2020*, 1–47. https://doi.org/10.1109/IEEESTD.2021.9321783

Jaggars, S. S., Motz, B. A., Rivera, M. D., Heckler, A., Quick, J. D., Hance, E. A., & Karwisch, C. (2021). The Digital Divide among College Students: Lessons Learned from the COVID-19 Emergency Transition. Policy Report. In *Midwestern Higher Education Compact*. Midwestern Higher Education Compact. https://eric.ed.gov/?id=ED611284

Johnson, M. (2018, October 1). *Digital literacy: A must in age of digital transformation, cyberattacks*. https://www.hrreporter.com/focus-areas/hr-technology/digital-literacy-a-must-in-age-of-digital-transformation-cyberattacks/299126

Joint Information Systems Commission. (2014, March 6). *Developing digital literacies*. Jisc. https://www.jisc.ac.uk/guides/developing-digital-literacies

Judd, T. (2018). The rise and fall (?) of the digital natives. *Australasian Journal of Educational Technology*, *34*(5), Article 5. https://doi.org/10.14742/ajet.3821

Kennedy, G., Judd, T., Dalgarno, B., & Waycott, J. (2010). Beyond natives and immigrants: Exploring types of net generation students: Types of net generation students. *Journal of Computer Assisted Learning*, *26*(5), 332–343. https://doi.org/10.1111/j.1365-2729.2010.00371.x

Li, Y., & Ranieri, M. (2010). Are 'digital natives' really digitally competent?—A study on Chinese teenagers. *British Journal of Educational Technology*, *41*, 1029–1042. https://doi.org/10.1111/j.1467-8535.2009.01053.x

Lynch, C. (1998). *Information Literacy and Information Technology Literacy*. https://www.cni.org/publications/cliffs-pubs/information-literacy-and-information-technology-literacy

Parker, A. G., Santos, J., & Dancy, K. (2021). Online Isn't Optional: Student Polling on Access to Internet and Devices. In *Institute for Higher Education Policy*. Institute for Higher Education Policy. https://eric.ed.gov/?id=ED612947

Prensky, M. (2001). Digital Natives, Digital Immigrants. *On the Horizon*, *9*(5), 6.

Sample Size Calculator. (n.d.). Retrieved July 12, 2022, from https://www.calculator.net/sample-size-calculator.html?type=2&cl2=95&ss2=118&pc2=50&ps2=146&x=71&y=19#findci

Smith, E. E., Kahalke, R., & Judd, T. (2020). Not just digital natives: Integrating technologies in professional education contexts | Australasian Journal of Educational Technology. *Australasian Journal of Educational Technology*, *36*(3), 1–14. https://doi.org/10.14742/ajet.5689

Šorgo, A., Bartol, T., Dolničar, D., & Boh Podgornik, B. (2016). Attributes of digital natives as predictors of information literacy in higher education. *British Journal of Educational Technology*, *48*(3), 749–767. https://doi.org/10.1111/bjet.12451

Talib, S. (2018, February 7). *Social media pedagogy: Applying an interdisciplinary approach to teach multimodal critical digital literacy*. https://journals.sagepub.com/doi/full/10.1177/2042753018756904

Vanek, J. (2020). *Issue Brief: Digital Literacy*. LINCS | Adult Education and Literacy | U.S. Department of Education. https://lincs.ed.gov/professional-development/resource-collections/profile-8856

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33. https://doi.org/10.1145/1118178.1118215

Appendix A. Tables

Demographic	Population	Sample
Number		
	936	440 (47.00%)
Age		
18-20	905 (96.69%)	432 (98.18%)
21 or over	1 (0.11%)	8 (01.82%)
Gender		
Female	368 (39.32%)	166 (37.72%)
Male	568 (60.68%)	273 (62.04%)
Prefer not to answer	n/a	1 (00.23%)
First generation?		
Yes	172 (18.38%)	49 (11.14%)
No	764 (81.62%)	391 (88.86%)

Demographic	Population	Sample
Ethnicity		
Nonresident Alien	104 (11.11%)	n/a
Black	39 (4.17%)	14 (3.18%)
Hispanic	101 (10.79%)	29 (6.59%)
Native American	3 (0.32%)	1 (0.25%)
Asian	84 (8.97%)	54 (12.27%)
White	563 (60.15%)	295 (67.05%)
Middle Eastern/North African	0	9 (2.05%)
Multiracial	35 (3.74%)	35 (7.95%)
Unknown	7 (0.75%)	3 (0.68%)
High School Location		
New England	557 (52.96%)	279 (63.41%)
Outside New England	258 (31.35%)	113 (25.68%)
Outside the U.S.	104 (14.72%)	47 (10.68%)
Unknown	17 (1.82%)	1 (00.23%)

Table 1. Population Demographics vs. Survey Respondent Demographics

DLC	Beginner	Intermediate	Expert
Word Processing	Format text using bold, underline, or different fonts and sizes to change the appearance of a document	Add a picture to a document	Perform a mail merge
	Add page numbers to a document	Use the format painter	Use the references feature to create a bibliography
Spreadsheet	Write a formula to calculate the sum of two numbers if they are equal, and calculate their product if not equal	Create a basic pie, line, or bar chart	Create a pivot table
	Write a formula to calculate the sum of a range of cells	Write formulas in excel using references such as A1	Write formulas to look up values in a table
Presentation Software	Use fonts and styles to change the appearance of a document	Modify a slide's background	Apply slide transitions and animations
	Insert shapes or images	Modify a slide's layout pattern	Apply a theme to your presentation
Database	Create a table to store data	Create a report	Relate information in two different tables
	Add data to a table	Add a primary key	Find records that match a specific condition (such as all employees from Connecticut)
Email	Send a message	Send a message with an attachment	Send a message encrypted for security
	Delete messages from your inbox	Reply to everyone who received a message	Organize messages in folders
Operating Systems	Create a new folder	See how much free storage remains on your hard drive	Zip or compress a file
	Copy a file from one folder to another	Install or uninstall a program	Back up files on your computer
Cloud Storage	Access your files in the cloud from your computer or mobile device	Upload your files to the cloud from your computer or mobile device	Share a link to a document stored in the cloud with someone else
	Edit a document stored in the cloud collaboratively at the same time as someone else	Specify folders on your computer to store on the cloud	Synchronize your files stored in the cloud across multiple devices
Web Browser	Bookmark a website	Accept cookies	View the HTML source code of a website
	Get directions using an online mapping service	Build a search query to limit search results using modifiers	Clear your browser's cache
Digital Media	Crop or resize a photo	Cut a section from a video clip	Publish a video to YouTube
	Flip or rotate an image	Add a title or credits to a video	Make a podcast

Table 2. Digital Competency Skills List

Appendix B. Survey Instrument

Willingness to Participate

You are invited to participate in a research study on the use of technology as it relates to IT education. You were selected as a possible participant because you are enrolled in IT 101. Please read this form and ask any questions you may have before agreeing to be in the study.

This study asks about your experiences learning about and using technology from high school to the present. If you agree to be in this study, you will be asked to answer several questions about your technology experiences. Your responses will be recorded and downloaded for analysis. The survey should take approximately 15 minutes to complete.

Click YES to participate. Click NO to withdraw from this survey.
o Yes I will participate
o No I will not participate
Skip To: End of Survey If Click YES to participate.

Demographics

In which section of IT 101 are you enrolled?
▼ Select Your Section and Instructor (drop-down list of sections and instructors)

From where do you plan to attend this class?
o my dorm room on campus
o my home or apartment off campus
o another location on campus

Which device do you plan to use most often to connect to your online classes?
o mobile phone
o laptop computer
o tablet
o desktop computer
o other (please specify) _____

Which best describes the Internet connection you plan to use most often to connect to your classes when joining them online?
o wireless
o wired connection
o cellular / mobile phone carrier
o I don't know
o other (please specify) _____

How old are you?
o Under 18
o 18-20
o 21 or over
Skip To: End of Survey If How old are you? = Under 18

With which gender identity do you most identify?
o female
o male
o transgender female
o transgender male
o gender variant/non-conforming
o other (please specify) _____
o prefer not to answer

What is your race or ethnicity (please select all that apply):
▢ White
▢ Hispanic, Latino, or Spanish origin
▢ Black or African American
▢ Asian or Asian Indian
▢ Native American or Alaska Native
▢ Middle Eastern or North African
▢ Native Hawaiian or Other Pacific Islander
▢ Another race, ethnicity or origin

Are you a:
o First-Year Student
o Second, Third, or Fourth-Year Student
o Transfer Student

Display This Question:
If Are you a: = Transfer Student

Did you take a technology class at your previous college?
o Yes
o No

Display This Question:
If Did you take a technology class at your previous college? = Yes

Did [college] accept credit for your technology course at a previous college?
o Yes
o No

The computer that you currently use for school is a:
o Mac
o Windows Computer
o Chromebook
o Other _____

The computer that you currently use here at college was:
o Purchased through college's program
o Purchased on my own
o One you had at home and/or used in High School

Where did you attend High School?
o New England (MA, NH, CT, RI, VT, ME)
o Elsewhere in the US
o Outside of the US

Are you the first person in your family to go to college?
o Yes
o No

High School Experience

Where would you most often use a computer when completing homework assignments in high school?
o A computer in a lab at school
o A computer at the public library
o A computer at home

Which of these topics, if any, did you learn as part of a course in high school?
- ▢ Programming
- ▢ Web Design / Making Websites
- ▢ Digital Photography
- ▢ Digital Video Production
- ▢ Digital audio production
- ▢ Computer Science
- ▢ Word Processing
- ▢ Spreadsheets
- ▢ Presentation Software
- ▢ Email
- ▢ Databases
- ▢ How to use the World Wide Web
- ▢ Social Media

Display This Question:
If Which of these topics, if any, did you learn as part of a course in high school? = Programming

Which programming languages or environments did you learn in high school? (Check all that apply)
- ▢ Alice
- ▢ Scratch
- ▢ Java
- ▢ JavaScript
- ▢ C or C#
- ▢ Python
- ▢ HTML
- ▢ Visual Basic
- ▢ Other _____

Which computer or device did you use most in high school?
- o Google Chromebook (Desktop or Laptop)
- o Windows Computer (Desktop or Laptop)
- o Mac Computer (Desktop or Laptop)
- o Linux Computer (Desktop or Laptop)
- o iPad
- o Android Tablet
- o Other _____

Did you take an AP Computer Science Test?
- o Yes
- o No

Did you work or volunteer as a help desk, computer lab assistant, or technology tutor while in high school?
- o Yes
- o No

Your Current Use of Computers and Devices

How often do you use these devices?

	Rarely/Never	Weekly	Daily
Iphone	o	o	o
Android phone	o	o	o
Other mobile phone	o	o	o
Smart Watch	o	o	o
Ipad or Tablet	o	o	o
Raspberry Pi	o	o	o
Laptop	o	o	o
Gaming Device	o	o	o
VR Headset	o	o	o
Smart Speaker (Alexa/Google Home)	o	o	o

Which device(s) do you use to perform the following tasks?

	Laptop/Desktop	Tablet	Mobile Phone
Watch a video	▢	▢	▢
Send email	▢	▢	▢
Send a text message	▢	▢	▢
Visit websites	▢	▢	▢
Use a search engine	▢	▢	▢
Edit a document	▢	▢	▢
Make an online purchase	▢	▢	▢
Listen to music	▢	▢	▢
Take a photo	▢	▢	▢
Edit a photo	▢	▢	▢
Post a photo to social media	▢	▢	▢

How often do you currently use the following social media apps?

	Once a day	Several times a day	One or a few times/week	One or a few times/ month	I never use it
Facebook	o	o	o	o	o
Instagram	o	o	o	o	o
LinkedIn	o	o	o	o	o
Reddit	o	o	o	o	o
Snapchat	o	o	o	o	o
TikTok	o	o	o	o	o
Twitter	o	o	o	o	o
YouTube	o	o	o	o	o

Your Tech Skills

Keep up the good work! You're about half-way there! This section has 11 questions related to tasks you might know how to complete with various software applications.

Rate your ability to use each of these applications before you came to college.

	Beginner	Intermediate	Expert
Word Processing	o	o	o
Spreadsheets	o	o	o
Presentation Software	o	o	o
Database Software	o	o	o
Email Software	o	o	o
Your computer's operating system (Windows or Mac OS)	o	o	o
Cloud Storage (Google Drive, Dropbox, OneDrive, etc.)	o	o	o
Web Browsers	o	o	o
Digital Media Editing (photos, videos, music)	o	o	o
Online and Social Collaboration	o	o	o
Online calendar	o	o	o

Rate your ability to perform these word processing tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Format text using bold, underline, or different fonts and sizes to change the appearance of a document	o	o	o
Add page numbers to a document	o	o	o
Perform a mail merge	o	o	o
Use the references feature to create a bibliography	o	o	o
Add a picture to a document	o	o	o
Use the format painter	o	o	o

Rate your ability to perform these spreadsheet tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Create a basic pie, line, or bar chart	o	o	o
Write formulas in excel using references such as A1	o	o	o
Create a pivot table	o	o	o
Write formulas to look up values in a table	o	o	o
Write a formula to calculate the sum of two numbers if they are equal, and calculate their product if not equal	o	o	o
Write a formula to calculate the sum of a range of cells	o	o	o

Rate your ability to perform these presentation software tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Use fonts and styles to change the appearance of a slide	o	o	o
Modify a slide's background	o	o	o
Apply slide transitions and animations	o	o	o
Apply a theme to your presentation	o	o	o
Insert shapes or images	o	o	o
Modify a slide's layout pattern	o	o	o

Rate your ability to perform these database tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Create a table to store data	o	o	o
Add data to a table	o	o	o
Create a report	o	o	o
Add a primary key	o	o	o
Relate information in two different tables	o	o	o
Find records that match a specific condition (such as all employees from Connecticut)	o	o	o

Rate your ability to perform these e-mail tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Send a message	o	o	o
Send a message with an attachment	o	o	o
Send a message encrypted for security	o	o	o
Organize messages in folders	o	o	o
Reply to everyone who received a message	o	o	o
Delete messages from your inbox	o	o	o

Rate your ability to perform these operating system tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Create a new folder	o	o	o
Zip or compress a file	o	o	o
See how much free storage remains on your hard drive	o	o	o
Install or uninstall a program	o	o	o
Copy a file from one folder to another	o	o	o
Back up files on your computer	o	o	o

Rate your ability to perform these cloud storage service (such as OneDrive, Google Drive, or Dropbox) tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Upload your files to the cloud from your computer or mobile device	o	o	o
Access your files in the cloud from your computer or mobile device	o	o	o
Share a link to a document stored in the cloud with someone else	o	o	o
Synchronize your files stored in the cloud across multiple devices	o	o	o
Edit a document stored in the cloud collaboratively at the same time as someone else	o	o	o
Specify folders on your computer to store on the cloud	o	o	o

Rate your ability to perform these web browser tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Accept cookies	o	o	o
Bookmark a website	o	o	o
View HTML source code of a website	o	o	o
Clear your browser's cache	o	o	o
Build a search query to limit search results using modifiers	o	o	o
Get directions using an online mapping service	o	o	o

Rate your ability to perform these digital media tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Crop or resize a photo	o	o	o
Cut a section from a video clip	o	o	o
Flip or rotate an image	o	o	o
Add a title or credits to a video	o	o	o
Publish a video to YouTube	o	o	o
Make a podcast	o	o	o

Rate your ability to perform these social media tasks before you came to college.

	Don't know what this is	Heard of it, but don't know how to do it	Know how to do it
Write a post to a blog or social media site	o	o	o
Set up a blog or website using a content management tool	o	o	o
Write a comment on a blog or social media post	o	o	o
Make a video call or participate in a video conference	o	o	o
Create an appointment on an online calendar	o	o	o
Specify which friends or groups of friends can see your posts to Facebook	o	o	o

Important to Know

Two more questions to go!

To what extent do you agree with each of these statements?

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Coding is a valuable skill to have.	o	o	o	o	o
I am concerned about my privacy online.	o	o	o	o	o
I would like to build mobile apps.	o	o	o	o	o
I consider myself to be tech-savvy.	o	o	o	o	o
I understand how technology works and know how to use it responsibly.	o	o	o	o	o
I can explain good practices for selecting a strong password for my accounts online.	o	o	o	o	o
I can explain different ways to protect my technology systems and information from unethical users.	o	o	o	o	o
I can explain basic practices that contribute to a website's accessibility to people with disabilities.	o	o	o	o	o

What three technology skills do you think are the most important that will help prepare you for your college education then into your future career?

_____
_____
_____
_____
_____