CCSCNE 2021

Student Poster Abstracts

The Consortium for Computing Sciences in Colleges

The Twenty-Fifth Annual
Consortium for Computing Sciences in Colleges
Northeastern Conference

Scheduled to be held virtually from April 16 - April 17, 2021

Organizer:

Ramapo College of New Jersey
Mahwah, New Jersey

Enabling Effective Visualization Creation in High School Teachers & Students

Matthew Spaulding
Computer Science
Rhode Island College, Providence, RI

Faculty Advisor
Sally Hamouda
Rhode Island College, Providence, RI

Given that 40% of students are visual learners [1], visualizations are crucial for students' retention on key issues and topics. With current instruction mostly coming in the form of auditory and textual learning—written and spoken word—students who are visual learners underperform [1]. With the rise in technology and big data, data and visualization literacy are becoming necessary skills for students preparing to enter the workforce in any field. However, available tools for creating visualizations are limited in accessibility due to their steep learning-curves, download requirements, sign-in requirements, and paywalls. While these tools offer expansive collections of fancy charts, they fail to deliver basic charts in a way that is simple to access for students and teachers. Additionally, available tools provide methods to create charts but provide little support to help users find relevant and effective datasets. For high school teachers, finding datasets that engage and connect with students is a difficult task. Our free online web-tool is built on a survey gathered from Rhode Island high school teachers during a data visualization workshop that enables users to create interactive and customizable charts in three clearly defined steps, using either data uploaded by the user or sample data that is local to the Rhode Island and surrounding New England area.

[1] - Hannah K. Ricketts, Alexa M. Salsbury, David R. Bevan, and Anne M. Brown. 2018. Using Immersive Visualization Environments to Engage Students in Hands-On Learning. In Proceedings of the Practice and Experience on Advanced Research Computing (PEARC '18). Association for Computing Machinery, New York, NY, USA, Article 74, 1–5. DOI:https://doi.org/10.1145/3219104.3229274

Developing a Control Room in Virtual Reality (VR) to Improve Underwater Remotely Operated Vehicle (ROV) Piloting

Amy Phung, Cameron Wierzbanowski, Erika Lu, Everardo Gonzalez, Nathan Shuster
Franklin W. Olin College of Engineering
Needham, MA

Faculty Advisors
Lynn Andrea Stein
Franklin W. Olin College of Engineering, Needham, MA

Ben Erwin, Eric Martin, Kakani Katija, Steve Haddock
Monterey Bay Aquarium Research Institute, Moss Landing, CA

We developed a control room display in virtual reality (VR) that has the potential to streamline the operation of underwater remotely operated vehicles (ROVs). Typical ROV control rooms consist of a wall of fixed monitors, each displaying a separate piece of telemetry data. ROVs at the Monterey Bay Aquarium Research Institute (MBARI) have been equipped with a 180° 4K stereo-pair fisheye camera to aid in scientific study and generate educational content. Our project, an early prototype, uses live footage from this camera to compute the re-projections necessary to create an immersive, 1:1 scale, 3D view of the ROV's surroundings in VR. Providing ROV pilots with a wide field of view, without the distortion that results from projecting raw footage onto a flat screen, should improve pilots' spatial awareness during operation. To begin incorporating the functionality of existing control rooms into our application, we project real-time telemetry data into the 3D environment and provide multi-user support to allow pilots and scientists to collaborate in VR. Our app also includes hand and gaze-based controls that enable pilots to rearrange the displays, a feature not possible with fixed monitors. Since pilots have individual preferences for how these displays are arranged, we included an option to save their display preferences, allowing for seamless transitions between pilots. We demonstrated our application during the operation of MBARI's MiniROV in a 1 million liter saltwater test tank facility, and we anticipate testing our application at sea later in 2021. Our poster further discusses the potential impact of VR on ROV control, our pilot and scientist centered development process, and the key features of our prototype.

Deep Learning API with Django

Salah Zahran
Computer Information Systems
SUNY Buffalo State, Buffalo, NY

Faculty Advisor
Sarbani Banerjee
SUNY Buffalo State, Buffalo, NY

Deep Learning is a subset of Artificial Intelligence and Machine Learning which is a type of technology that can think intelligently similar to humans. The concept involves a deep processing and analyzing of data which leads computers to make the best possible decision based on patterns and implications in data analysis. Deep learning takes this concept to the next level by creating a similar structure to neural networks in the human brain and applies this concept to its models. It involves creating artificial neural networks in which there are multiple layers consisting of nodes which each contain a small fraction of all the data input. Deep learning applications are very commonly written in Python that will be used for the current project. Other software that will be used for this project is Django, which is a back-end framework for Python, and is widely used for creating Application Programming Interfaces (API) and backend web applications. This research project will be creating an API with Django and use it to run any Deep Learning models. API provides users with an interface that serves as a middle ground between users and the backend. They offer versatile ways of interacting with web applications and make it extremely easy for users to interact with backend apps as well. The project will consist of coding an API that can accept Deep Learning models as input and will then run and output the model's results in a convenient and clear way to the user.

Person Re-Identification: Tracking the World

Mac Johnson
Computer Information Systems
SUNY Buffalo State, Buffalo, NY

Faculty Advisor
Sarbani Banerjee
SUNY Buffalo State, Buffalo, NY

Person Re-Identification is the use of multiple videos or camera angles to track individuals over time. Someone walking down the street might be in several camera angles at once, but the task is to combine the videos or snapshots to show it is the same person. One problem with this idea is that pictures can have many different variations that affect the accuracy. The plan is to use multiple training libraries to find if this improves identification. Person Re-Identification is a rising topic in the Cybersecurity field and will have a lot of potential for progress. PyCharm will be used as the IDE for this program because all the learning processes are stored in libraries in Python code. There a couple of options regarding which learning dataset to use, CUHK01, iLIDS-VID, and RPIField being a few of them. The libraries will train the program with artificial intelligence to track a person and store images of them for future queries. Since most of the time libraries are used individually, there should be an increase in accuracy by combining them. The expectation is to see a clear advantage of training methods when used together and also identify the single most efficient library. The goal is to find a different combination of these training techniques to allow the program's artificial intelligence to be more accurate and adaptable. The results of this research project will show the differences and benefits of using multiple image libraries.

Vehicle Dash Cameras with Artificial Intelligence

Matthew M. Stranz
Computer Information Systems
SUNY Buffalo State, Buffalo, NY

Faculty Advisor
Sarbani Banerjee
SUNY Buffalo State, Buffalo, NY

The purpose of this research project is to create a vehicle dash camera that assists the user, police, and insurance company in easy license plate recognition and number retrieval in an event of an accident and/or hit and run. Right now, personal vehicle dash cameras are very limited and only record data, and since accidents are very chaotic, it can be difficult extracting the required information. To implement these specific aspects, the vehicle license plate will be detected by YOLOv4, a real-time object recognition system and machine learning model. Following detection, the number within the license plate will be filtered through OpenCV real-time computer vision and printed with Tesseract OCR optical character recognition engine. Currently, this is a software capability that is only found in government, police, drone, and stationary security cameras. Supporting software for this research will be with Google Colaboratory, Jupyter Notebook, Anaconda (Miniconda for Raspberry Pi), Nvidia CUDA Toolkit 10.1, and Git. Additionally, the hardware that would be needed is a Nvidia GPU, Raspberry Pi 4 Model B, a storage device micro SD card, Arducam day and night vision camera, Raspberry Pi 7" touch screen display, USB to USB Type-C with 15W car charger, and a housing unit that would incorporate the hardware along with a mounting for the windshield or the mirror. The final goal is to have all license plate numbers be enlarged and printed above the license plate, and all detections be saved onto the storage device.

Quick Browse

Pa Reh
Computer Information Systems
SUNY Buffalo State, Buffalo, NY

Faculty Advisor
Sarbani Banerjee
SUNY Buffalo State, Buffalo, NY

The goal of this research project is to create a Google Chrome extension called "Quick Browse," which will make navigation easier for all users on Chrome to browse around the web-pages, in less time. Google Chrome has been phenomenal, and many people prefer Chrome over another web browser because it is very user-friendly. However, the browser does not have all the best features available, so it is important to take advantage of the Google Chrome Extension tool to create the necessary changes, which is why "Quick Browse" is needed. This problem is important to fix because many users waste countless hours on the browser clicking around the webpage. The "Quick Browse" will speed up the process for the user to get to their destination webpage. It will allow a user to get more work done without having to spend so much time clicking around the browser. The technology involved in this project will be HTML, CSS, JavaScript, Jason, and Photoshop. Photoshop will be used to create an image, appropriate for the icon of the extension. HTML and CSS will be used to mainly display what is on the browser. JavaScript and Jason will be the code that works behind the scenes of the Extension. The result of these projects will save the user a lot of time. It will allow the user to get to their destination webpage without having to type in the web address all the time, which is much easier, and it saves a lot of time in the long run.

Traffic Light Control with Reinforcement Learning

Devanshi Malaviya
Computer Information Systems
SUNY Buffalo State, Buffalo, NY

Faculty Advisor
Sarbani Banerjee
SUNY Buffalo State, Buffalo, NY

The rapid increase in automobiles in the past few years has led to traffic congestions all over the world. This forces drivers to sit idly in their cars wasting time and fuel. Current traffic light control policies are not optimized which leads to people waiting in their cars for nonexistent traffic and extended travel time than necessary. In the US, on an average, people spend about 100 hours in traffic congestions per year. The current research project will focus on reinforcement learning to optimize the traffic flow to reduce the travel time of drivers. It can be done by building an environment where every intersection has knowledge about the number of the vehicles and their speed as they approach the intersection. Simulation of Urban Mobility or SUMO will be used to build a traffic simulator. Reinforcement learning works on state and action policies which will allow the traffic lights to make optimized decision based on their current state. It will balance the exploration and exploitation to make sure that the model is not overfitting and every lane is given importance according to how busy it is. For every state, it gets a reward if it reduces the travel time and the goal of the model is to collect as many rewards it can. Therefore, at the end this project will attempt to get the most optimized simulation. The Python packages used in this project are Keras, Tensorflow and OpenAI.

Analyzing Communications on Twitter after a Year into the COVID-19 Pandemic

Matthew Morgan
Department of Computing Sciences
SUNY Brockport, Brockport, NY

Faculty Advisor
Adita Kulkarni
SUNY Brockport, Brockport, NY

Social media is a crucial channel for sharing news and information in the modern media environment. During the COVID-19 pandemic, people largely used social media, in particular Twitter as a medium to express their struggles and opinions. Understanding the socioeconomic state by analyzing tweets will help in better planning and preventing socioeconomic distress during future epidemics. In this work, I collected tweets all over the world using the Twitter API from December 20, 2020 to January 16, 2021. I kept track of trending hashtags and keywords related to COVID-19 and collected tweets containing those hashtags and keywords daily. I grouped the tweets into three categories — 'General COVID-19', 'Vaccine', and 'New Strain'. Tweets containing hashtags like #COVID19, #covid, #corona, #coronavirus were classified into the 'General COVID-19' group, #Covid19Vaccine, #CovidVaccine, #ThisIsOurShot, #vaccination, #AntiVaccine, #NoVaccineForMe were classified into the 'Vaccine' group, and #mutation, #CoronavirusStrain, #newstrain were classified into the 'New Strain' group. The goal of this on-going work is to perform lexical category, word collocation, psycholinguistic and sentiment analysis of the collected tweets and compare the results with research published during the early days of the pandemic in order to understand its long lasting effects.

Comparison Performance Study of Round Robin Variations for CPU Scheduling

Joshua Marquis
Computer Science
Eastern Connecticut State University, Willimantic, CT

Faculty Advisor
Sarah Tasneem
Eastern Connecticut State University, Willimantic, CT

We present our research on a comparative performance study of variants of Round Robin (RR) CPU scheduling algorithms in a single processor system. The goal of the research was to determine which RR variant would have the best performance in terms of average turnaround time, average wait time, and the number of context switches. To achieve this we developed and tested CPU scheduling simulation programs and measured the performance benchmarks for multiple algorithms. This research is important because CPU scheduling is vital to operating system design, as it has significant impact on resource utilization and the overall performance of the system. RR being the most common CPU scheduling algorithm, a number of modifications have been designed to the traditional *RR* algorithm [1, 2]. This study examined several improved *RR* algorithms, namely *sorted RR*, *sorted and minimum quantum RR* [2], and the *sorted and minimum quantum RR thresholded (SMQRRT) variants* [1]. The performances of these variants were evaluated through discrete event simulation. Task service times and inter-arrival times were generated following uniform distribution and the simulation program generate processes while the CPU serviced them following *RR* policy. In this project we present results for 1000 processes and measured the average turnaround time, average wait time, and the number of context switches for each algorithm. Several simulation test outputs were recorded by varying inter-arrival and service time bounds, quantums, as well as context switch times. Based on the results generated, the algorithm with the best performance was the *SMQRRT* algorithm proposed by Shafi et al. [1]. This algorithm had the lowest average turnaround time, average wait time, and had the least amount of context switches of all the algorithms tested. The results from discrete event simulation suggest that *SMQRRT* can be used for improved performance.

References

[1] U. Shafi *et al.*, "A novel amended dynamic round robin scheduling algorithm for timeshared systems," *The International Arab Journal of Information Technology*, vol. 17, no. 1, pp. 90–98, Jan. 2020, doi: 10.34028/iajit/17/1/11.

[2] M. K. Mishra and F. Rashid, "An improved round robin CPU scheduling algorithm with varying time quantum," *International Journal of Computer Science, Engineering and Applications*, vol. 4, no. 4, pp. 1–8, Aug. 2014, doi: 10.5121/ijcsea.2014.4401.

A Ruff Day for a Dog Salon, a Way to Collect Data

Alexander Wagner
Computer Information Systems
SUNY Buffalo State, Buffalo, NY

Faculty Advisor
Sarbani Banerjee
SUNY Buffalo State, Buffalo, NY

The problem that exists in every company is how are they going to collect and store user data, then keep it safe under a login authenticator. For the case of a small dog salon, they don't have a lot of resources to do so. Using Django, Python and SQL they can make a simple website that allows users to enter, collect and show data that they've collected. This research project is meant to show how this is applicable for small businesses. The rationale in doing this project is that when you own a small business with 3-5 employees' information about customers and such data can be lost very easily. Prior research was done to fix this problem, but the difficulty was that the application was hosted on one computer that was shared with all the employees. The solution for this will be to launch a website and back-end data storage using Django so the employees can record information with their device of choice. The way this can be done is with Django to set up an IP address that the employees can connect to, a form that can be created with Python, and a database that can hold all the information. Also, some tools to help download the database in different forms such as comma separated values and a text document will be used. One can expect an application that runs on a website that is secured with a login client, and a means of data collection.

What to Trust When Searching for Health-Related Symptoms on Google

Nina Sachdev
Computer Science
Wellesley College, Wellesley, MA

Faculty Advisors
Eni Mustafaraj
Wellesley College, Wellesley, MA

Users who perform symptom searching on Google can often feel overwhelmed with the vast amount of online information available to them. In some cases, this can lead down the anxiety-ridden path of cyberchondria and health misinformation. Thus, the goal of this study is to provide users with clear steps they can take to consume trustworthy health information online. Through web scraping and machine learning techniques, data analysis can be done on the Google search results generated from medical and problematic health queries. The results identify authoritative health websites to click on, the components of a search results page to pay attention to, and the best way to formulate a health-related query. The findings suggest there is a prevalence of authoritative websites in both medical and problematic health search results, as well as an association between Google knowledge panels and legitimate medical queries. This will inform future work on classifying Google search results pages by their level of trustworthiness and assistance to symptom searchers, as well as aid further efforts to combat cyberchondria and health misinformation. The poster for this study will highlight the motivation behind focusing on health-related search results on Google, as well as give an in-depth analysis into key findings that are informative to symptom searchers.

Are those Ants?

Karanveer Gill
Computer Information Systems
SUNY Buffalo State, Buffalo, NY

Faculty Advisor
Sarbani Banerjee
SUNY Buffalo State, Buffalo, NY

This project will be studying how to detect people in images who appear miniscule. Small people can be identified as low as 20 pixels. This project has some inspiration from satellite images. It can almost replicate how a satellite image can detect a person. For example, if a picture was taken from a satellite, the identification of small persons will only be beneficial for investigations. This project will be written by using Python language and Jupyter Notebook will be used as the Integrated Development Environment (IDE). Windows Docker will be utilized to access a COCO annotator. This will be accessed within the Python program, referencing a JSON file. A dataset will be downloaded with help from this COCO annotator. This data annotation will be designed to detect tiny people. The purpose of this project is to immediately detect persons in a picture that are not in a crowded area. Images will mainly be aerial shots where the people in the image will look tiny in size. There will be an indication that the object, in this case the "tiny-person," is identified. This project will also attempt to zoom-in the photo, almost to replicate a zoomed-in camera or satellite image. With the identification of tiny persons, the object detected will turn out to be human. An objective of this research is to mimic a satellite image or camera identifying potential suspects.

Accessibility in the Classroom

Mark Sternefeld, Saad Khan, Heather Moses, Su Thit Thazin
Golisano College of Computing and Information Sciences

Rochester Institute of Technology, Rochester, NY

Faculty Advisors
Samuel Malachowsky, Daniel Krutz

Rochester Institute of Technology, Rochester, NY

When teaching students to create accessible software, instructors often use traditional learning methods such as articles or videos. Unfortunately, these frequently don't effectively reinforce best practices. Accessibility in software creation is important because 15% of the world population has one or more disabilities. Educators have indicated that it is hard to incorporate accessibility into already-packed courses, and many instructors may not feel comfortable teaching a topic that they themselves aren't well-versed in.

Over the last two years of this project, our results indicate that our proposed learning solution is preferred by students over traditional means. We investigated two student groups and compared their feedback. The traditional learning group received articles and presentations, whereas the interactive learning group utilized our developed learning labs. Our study indicated that students found the passive learning model overwhelming and not engaging as it involved a lot of reading. We examined the post-survey results from the question: "How useful was the lab?" and performed a statistical analysis using a Likert scale of 1 to 5. We also performed a sentiment analysis on qualitative feedback to determine the students' overall experience.

The results indicated that based on a scale of 1-5, the interactive learning group rated their lab experience significantly more useful ($< .0003$) than the traditional learning group rated their material. Additionally, based on the sentiment analysis, we found that there was a .94 confidence in positive sentiment for the interactive learning group compared to the .83 confidence in negative sentiment for the traditional learning group. In conclusion, our results indicate that students who used our labs found them to be both more useful and more enjoyable when compared to traditional learning.

Phorcys

Connor Blanchfield-Tomaszewski, Andrew Mahr,
Jordan Zimmitti, Samuel Zurowski

Tagliatela College of Engineering
University of New Haven, West Haven, CT

Faculty Advisors
Vahid Behzadan, Mehdi Mekni
University of New Haven, West Haven, CT

Penetration testing is a form of ethical hacking where an attack is performed against a computer network to find vulnerabilities. Due to the complexity of penetration testing, it is commonly done manually by trained cybersecurity professionals. Additionally, the lengthy amount of time it takes to conduct a penetration test results in high costs, and the large scope of these assessments can lead to human errors resulting in missed vulnerabilities. Our project, named Phorcys, is designed to conduct an automated real-world attack utilizing recent advances in machine learning to report strengths and weaknesses within a network. Phorcys utilizes deep Reinforcement Learning (RL) to automate the process of penetration testing. Once trained, The Phorcys RL agent can conduct penetration tests fully autonomously. Therefore, Phorcys' use of deep RL provides companies with a straightforward and cost effective approach to conduct high-quality and frequent penetration tests. At a high-level, Phorcys will start with a user who tells the agent the scope of the attack. It will then perform reconnaissance that will be ingested into the deep RL model for the given targets. The model will decide on what exploits to leverage in the process, and executes those exploits to compromise the target. After successfully conducting the assessment, Phorcys concludes by automatically generating a report of the penetration test.

Differentially Private Machine Learning for Breast Cancer Classification

Melinda Goda
Department of Mathematics and Computer Science
Concord University, Athens, WV

Faculty Advisors
Abdur Shahid
Concord University, Athens, WV

Over the last decade or so, Machine Learning (ML) has significantly changed the healthcare system. Breast cancer detection is just a single instance from a large pool of healthcare problems which has benefited fundamentally by ML. Breast cancer among women is one of the most common and deadliest cancers worldwide. Early diagnosis can help in timely treatment; leading to improved survival rates. ML has made the early detection of malignant tumors and the reduction in the possibility of mistreatment a reality. However, the vulnerabilities associated with the design of ML-based systems have not yet been fully understood. In the breast cancer classification problem, the ML model responsible for the classification is considered to be highly valuable intellectual property. Study shows that ML models have the capacity to capture information on their training data, and exploitation of the models could lead to a membership attack and training data leakage using the model's prediction. Therefore, protecting patient's privacy is highly important in healthcare problems, such as in breast cancer classification, and very little works have been done to meet this requirement. To address the privacy issues, differential privacy has gained significant attention to develop privacy-preserving machine learning models. In this work, we propose to utilize the differential privacy-based Logistic Regression and Naive Bayes models to classify benign and malignant tumors in such a way that individual's presence in a dataset is obfuscated. We implemented the two privacy-preserving models using the popular Wisconsin diagnostic breast cancer (WDBC) dataset and compare them with traditional non-private Naïve Bayes and Logistic Regression algorithms. We extensively study the relationship between the privacy requirement and model accuracy. The results show that it is possible to achieve high accuracy with both privacy-preserving models.

Hyperpass - A Unified Password Manager

Myles Trevino
Computer Science
Wheaton College, Norton, MA

Faculty Advisor
Martin Gagné
Wheaton College, Norton, MA

With the increasing use of digital services, people are accumulating more sensitive digital information than ever before. We present a password manager software that follows the qualities we believe to be essential to such an application: open-source, multi-platform, and unified codebase.

Existing options do not meet these requirements. Google's password manager, LastPass, and 1Password are all proprietary, keeping users in the dark about what exactly is being done with their personal information. Mozilla's Lockwise cannot be used on browsers other than Firefox, and KeePass has no official mobile application, making users unable to access the information they need when using certain devices or platforms. Bitwarden uses separate codebases for each platform, duplicating functionality in different ways, and resulting in an inconsistent user experience.

In contrast, our solution's source code is publicly available on GitHub, available for web, desktop, mobile, and as a browser extension, and uses a shared codebase across all platforms. With this approach, security-concerned users can view the application code, users can access their critical data on any device, and the core interface and functionality will be the same across all devices.

A unified codebase greatly impacts development, and this manifests in the quality of the end product. Functionality is only written once rather than separately for each platform, reducing potential sources for bugs. Fixes and improvements generally apply to all platforms rather than individual ones, making it easier to add new features and improve existing ones. As a result, development time is reduced and quality is improved. These benefits allow us to focus efforts onto creating a secure, intuitive, and visually stunning experience for all devices.

We hope releasing an application which follows these tenets will make digital security more accessible for everyone.

.

New Haven Admissions Robot (NHAR)

Robert Boutillier, Meghan Cichon, Erica Maggiore, Marcus Novoa, Jacob Sanchez

EECS Department

University of New Haven, West Haven, CT

Faculty Advisor

Mehdi Mekni

University of New Haven, West Haven, CT

Robots endowed with Artificial Intelligence Conversational Agents (Chatbots) capabilities have become more popular in business, as they can improve customer service experiences, reduce costs, and handle multiple users at a time with around-the-clock availability, reliability, and accessibility. However, this technology being used in education is still in its infancy. In fact, an Open House event is a keystone of a university's recruitment plan. Such events let prospective students catch a glimpse of campus life, culture, and accomplishments. However, planning an Open House that ensures every prospective family feels engaged is a challenge. To address this problem, we propose our New Haven Admissions Robot (NHAR), a smart virtual assistant robot that provides continuous and instant support to students and their families. Support includes accurate information about the university, the offered programs, the history, the campus, and the student life to name a few. NHAR uses a software solution that integrates IBM's chatbot (Watson), a React Native mobile application, a NodeJS web server, and Google's cloud-based Natural Language Processing (NLP) services, and more. Our poster presents the problem statement, our followed methodology, the software requirements engineering, associated design and architecture, and concludes with future perspectives of our project.

Smart Home Energy Services (SHES)

Alex Socha, Andrew DeMarco, Chris Dowd, Ehis Ekore
Computer Science
University of New Haven, West Haven, CT

Faculty Advisor
Mehdi Mekni
University of New Haven, West Haven, CT

The average monthly residential electricity bill in Connecticut is $127. That's 19% higher than the national average. Residential energy accounts for 12% of greenhouse gas emissions and 33% of total U.S. energy-related CO2 emissions. Moreover, low-income households face an energy burden three times higher than other households. This forces tough budget choices because home energy costs demand a larger portion of their income. There is an urgent need to help residents achieve measurable savings on energy bills and carbon emissions, while improving the comfort and indoor air quality of their home. In collaboration with Neighborhood Housing Services (NHS) and I Heart my Home, a startup especialized in revitalizing communities via social, economic, and environmental health, we propose Smart Home Energy, a scalable, portable, modular solution that help home owners assess their energy needs. Our solution addresses the following requirements: 1) characterize their physical buiding systems, 2) evaluate providers of retail energy supply and/or onsite generation, and 3) using computer simulation techniques, assess products or occupant changes that modify the timing of equipment operation. The underlying software architecture of our SHES solution is fully open source and includes the following tools: *Parametric Analysis Tool* (PAT), *Open Studio*, and *ResStock Analysis Tool*. Since *Scalability* and *Portability* are key software requirements requested by our partner, *Amazon Web Services* (AWS) will be used to deploy our solution and support large volume of data and concurrent access to our platform of several hundred of users. Moreover, a RasberryPi will host the simulation module to support our partner's professionals in the field.

aChord: Generating Chord Voicings for User-Defined String Instruments

Ben Bingham, Caleb Braddick, Jackson Reed, and James Stratton
Computer Science
Wheaton College, Norton, MA

Faculty Advisor
Martin Gagné
Wheaton College, Norton, MA

In order to play a chord on a string instrument, one must select a "voicing"—that is, a permutation of the members of the chord and arrangement of the fingers on the fingerboard. For each combination of chord and instrument/tuning, there is a unique, potentially very large set of voicings to choose from. Existing encyclopedic references are difficult to navigate, include only standard instruments and tunings, and offer a limited number of pre-determined voicings for each chord. To determine voicings without such a reference requires significant experimentation and music theory knowledge.

As an alternative, our mobile application, aChord, generates a complete set of voicings for a user-defined chord and instrument algorithmically, using a more comfortable, accessible interface. Chords may be submitted in the familiar "chord symbol" format in which they are printed on sheet music (e.g. "Cmaj7"), and will be parsed into a collection of notes. The resulting voicings are presented via live-generated fret diagrams in a standard format. The complete set of voicings is culled based on several criteria which can be specified by the user in accordance with their needs and the specifications of their instrument; for example, voicings which omit the fifth of the chord may be included to accommodate instruments with fewer strings, or otherwise removed.

User inputs are translated into chord and instrument objects which can perform basic musical operations such as adding an interval to a pitch. In the case of chord symbol input, each character of the symbol is parsed, looking for multiple common styles of notation. Then voicings are generated via a recursive algorithm that visits every combination of valid finger positions and saves the ones that comprise the complete chord and meet the user-defined criteria.

Academic Support Seeking Behaviors Differ by Gender in Coding Bootcamps

Rachel Beaulieu
Mathematics, Computing and Statistics
Simmons University, Boston, MA

Faculty Advisor
Nanette Veilleux, Sherry Seibel
Simmons University, Boston, MA

Coding bootcamps are an increasingly popular alternative to an undergraduate computer science degree. They are comparatively short and inexpensive, with an average duration of 14 weeks and an average cost of $14,000. The gender imbalance in traditional undergraduate programs is notorious, however coding bootcamps seem to be bridging the gender gap. Comparing the most recent government data to statistics from Course Report, an estimated 17-19% computer science graduates vs 34% of coding bootcamp graduates were women in the 2017-2018 academic year. The 2021 survey by Course Report showed women in coding bootcamps increasing to 41%. Since bootcamps appear to be attractive alternatives to traditional college degrees, we have conducted a qualitative study to investigate how women experience each of the two pathways.

Concentrating on bootcamp populations, this study found differences in the way that men and women navigate their academic experiences. Our research uses grounded theory to analyze over 11 hours of recorded interviews of 7 women, with a smaller comparison group of 4 men. One initial finding reveals that women and men seek academic support differently. Although both cohorts relied on peer study groups, women intentionally selected groups primarily composed of other women whereas men noted no preference. The smaller number of women therefore yielded smaller all-women study groups compared to men's nearly homogenous study groups. In addition, while all participants reported some positive experiences, over 50% of the women interviewed had at least some negative experiences during the course of their program compared to only 25% of the men. Seeking all women peer groups was one strategy women took to mitigate negative experiences. These findings further suggest that bootcamps do not provide a clear alternative to the unwelcoming environment women students cite in traditional undergraduate programs and that further research is necessary to identify their greater popularity.

How to Train Your Data Scientist

Meghan VanSchalkwyk
Computer Science
Rhode Island College, Providence, RI

Faculty Advisor
Sally Hamouda
Rhode Island College, Providence, RI

This research is about the topic of high school students and data visualization, more specifically what can be done in high school to prepare students for a career in data science? Currently there is little data on this topic as the field of data science is quite new. There is slightly more information available on data visualization, likely as students have seen some form of data visualizations, namely chart, bar, and pie graphs, before they reach the high school level. Currently the issue with what students have seen in data visualization is just simple graphs and images based on small data sets. They have not generally encountered visualizations representing very large amounts of data.

Therefore, surveys were given out to high school teachers and professors at the university level in the state of Rhode Island over the summer of 2020 about the topics of data science and data visualization. There were also surveys given out to professionals in the data science field and surveys for high school students who may be interested in a data science career. It also aimed to increase the number of teachers incorporating data science and data visualizations in their classrooms even for students who are not interested in data science as a career. This is considered important as data visualizations are everywhere, online, in the news, on social media.

The goal with this paper was to determine the level of knowledge about data visualization and to determine the best way to increase the level of knowledge and interest in data science and data visualization. From that baseline of what people know, combined with what the industry experts want to see in new data scientists it will be possible to develop curriculum to support high school students that are interested in data science and data visualization.