Data Science Outreach Educational Program for High School Students Focused in Agriculture

Jasmine Sami, Krystin Sinclair, Zachary Stein, Larry Medsker
George Washington University

Ethics Procedure:

The study was reviewed and approved by the George Washington University IRB. Written informed consent was obtained from the parents of all participants.

Abstract

Members of the Data Science Program at George Washington University (GWU) designed and implemented a tuition-free two-week summer camp at GWU for high-school students from the Washington Metro Area. The United States Department of Agriculture (USDA) Office of the Chief Information Officer and his staff were our main partners in the project. The goal was to use Open Data related to Science, Technology, Engineering, Agriculture, and Math (STEAM) as the means of giving students experience with real-world data analytics methods and tools, as well as insights into careers in data science and agriculture. GWU provided a prime location in the District of Columbia, expertise from the data science master's program, and experience in active learning pedagogy. This two-week long (July 23-August 3, 2018) project-based summer camp introduced participants to the idea of using data for making important decisions as applied to topics in food nutrition, forestry, and urban agriculture. The camp had a special focus on effective and compelling ways to visualize data. Students explored, analyzed, and reconfigured quantitative and qualitative data, using fundamental graphical principles to present their project-related findings. GWU faculty and students provided guidance on the kinds of questions that can be addressed with data, on the challenges of gathering data through interviews and surveys, and on the techniques for presenting compelling arguments based on data. The camp curriculum was designed for project-based active learning with the aim of all the students gaining skills with ArcGIS, Excel, Tableau, and ESRI's Collector application for creating GIS maps. An overall goal of the camp was to spark interest in data science using the STEAM context. The learning goals were that participants finishing the camp would be able to demonstrate basic knowledge of the methods of data science by applying them to specific problems in the STEAM domain area of the camp and presenting the results in a conference-style setting. Achievement of the learning goals was assessed using a survey tool and successful completion of projects, as judged by the faculty and staff. The positive results are discussed along with conclusions, limitations, and recommendations for future camps.

Rationale

The industrial revolution in technology is creating a massive amount of jobs in STEM fields. Hubs of technology like San Francisco, New York, and Washington, DC, are contributing especially to the need for workers with STEM skills and knowledge. The demand for STEM professionals throughout the country has only been increasing with time. The number of STEM occupations in the United States is projected to grow by 8.9 percent between 2014-2024. The average salary for entry-level STEM jobs requiring a BA or higher is $66,123 compared to $52,299 for non-STEM jobs. With high demand for STEM workers, a proper introduction to STEM in schools is needed (Bybee, 2015) and (Madden *et al*., 2013). Early education appears to be the most influential place to start.

In fall of 2018, approximately 50.7 million students attended public schools in the United States. Public schools in the United States are maintained at public expense for the education of children of a community that constitutes a part of a system of free public education commonly including primary and secondary schools. Public schools in high-income areas benefit from access to guidance counselors, school psychologists, personal laptops, and up-to-date textbooks. This is in stark contrast with schools in low-income areas with less resources, lower-paid teachers, dilapidated facilities and larger class sizes. Without proper educational resources, teachers, or funding, it is increasingly difficult to have comprehensive STEM curricula in low-income school districts. At the national level, low-income school districts spend 15.6 percent less per student than high-income school districts do. While the intentions of public-school education are great, more effort needs to be made to better STEM education for high-need students.

The role of the educator in the push for more students pursuing the STEM workforce is crucial. Educators from elementary through the high school level have the ability to impact educational interests of their students. Educators are given the task of making their content relevant to students. Without the proper instruction and curriculum design of STEM topics from teachers, students will not garner interest and reap the benefits of the job prospects in STEM. The interdisciplinary nature of data science has been piquing the interest of students and professionals alike. In the modern world we live in, data exists everywhere from social media to our fitness tracking devices. With the amount of data only growing, it is essential to educate students about data science and how they can use these skills in career pursuits. The goal of the USDA Open Data camp was to spark an interest in data science to students who have not had exposure to it before. We want to inspire educators to create their own educational outreach programs that tie in the dynamic fields of data science and other STEM concentrations. For further discussion, see (Levin and Tsybulsky, 2017a), (Lewis and Matsumoto, 2017), (Lombardi, 2007), and (NRC, 2013)

Learning Objectives

When the camp was discussed with the USDA partners, we identified a few central themes we wanted the curriculum to have. The camp's main focuses were data science and STEAM, a concept that isn't as widely known. These two concentrations are rapidly becoming the most in- demand fields throughout the country, but especially in the DC area. George Washington University and partners believed that creating a camp with these focuses in the curriculum would potentially inspire high school students to pursue degrees in these fields once they enter college.

Our approach started with identifying what the participants should be able to do after finishing the camp. Our experience with data science education led us to priorities-students learning how to demonstrate practical problem-solving skills and communication within teams, as well as being able to present results. Our assessments were then aimed at measuring competencies students gained in the key areas of our learning objectives. Figures in the Results section are intended to demonstrate competencies and the extent that learning objectives were accomplished.

In this first offering of the camp, a small-scale pilot study, our learning goals were limited to making sure the program met expected needs and provided an opportunity for high school students to explore data science with hands-on application. The information gathered and the success of the program point to expansion of the assessment to explore additional benefits. For example, we foresee as a next step an educational research approach to measure aspects such as students' understanding of the nature of science and on their attitudes towards science. The data gathered in our project reported here will provide a solid basis for expressing additional learning objectives and corresponding assessments.

Data Science

The idea of doing short summer camps on data science is timely and necessary because of the popularity, amount of misinformation, and the rapidly emerging needs in the workforce. High school students need in-depth and accurate experiences for making good career choices and getting a realistic start in the field. Data science education research is an emerging new field that can borrow from advances in physics education research and computer science research. High-level commissions are just starting to work on requirements and recommendations for data science education, and experts disagree on the best approach in content and pedagogy. Meanwhile, educators can use advances and lessons learned from other areas of STEM education such as active learning and project-based learning. (Lewis and Matsumoto, 2017) The approach in the work reported here built on our experience in these related areas of STEM education research and on experience implementing a very successful data science graduate program.

Curriculum relating to data science began with a brief introduction as to what it is and how it is used. A final project was due at the end of the two-week camp which would display their knowledge of the programs and concepts introduced. Beginning with Microsoft Excel allows students to get a basic understanding of formulas, which can translate into other programs. In addition, understanding the basics of things like pivot tables and visualizations would be massively helpful when it came to analyzing and visualizing data. After the basics of Excel was learned, the more advanced programs were introduced. ESRI ArcGIS and Tableau were the programs used throughout the camp. ArcGIS is a program that works with maps and spatial reasoning to understand particular locations through data. Tableau is a data visualization program that creates unique visuals to display data. These programs would allow students to get creative in how they work with data. The term 'data science' can appear broad and vague, but introducing these programs allows the students to see the innovative ways data science is used.

STEAM

STEM has been revered and well known throughout the educational space since 2001, when NSF created the acronym (Koonce 2011). In the new word STEAM, the 'A' in represents agriculture. The need for STEM professionals in the United States has been increasing in tandem with our advancing technologies, yet the amount of STEM graduates we produce has not matched this demand. The camp was intended to introduce high school students to these concepts in a relevant way. Lumping agriculture in with STEM is an interesting addition considering this camp was held in a very urban area, Washington D.C. It is often falsely assumed that cities do not have a lot of agriculture, but this is not true. Urban agriculture is alive and well in many cities throughout the country.

Pedagogical Framework, Camp Design, and Structure

Pedagogy

The pedagogical design and learning environment were developed using our extensive experiences from physics education research (PER) and data science curriculum development. Key features were the use of active learning principles and project-based learning methods. Research in PER, including our own research and grant work, show the importance of students gaining deep knowledge and data analysis skills, accompanied by applications to real- world projects (Teodorescu, Bennhold, Feldman, and Medsker, 2012). Scientific communication includes collaboration in teams as well as learning presentation skills for showing results of projects. Our class sessions were designed with sequential use of brief lectures, student practice and immediate feedback to the students. The learning environment was a modification of the principles we use in Scale-Up classrooms. See (Beichner and Saul, 2003), and (SERC, 2018). The team involved in planning, curriculum design, outreach included different stakeholders: faculty and students from George Washington University's data science program, professionals from the US Department of Agriculture, and software professionals from Tableau and ESRI. The external colleagues were extremely valuable in making sure we had the latest in data analysis and visualization tools, as well as providing interesting examples of their use in current applications in urban agriculture. These external partners were also invaluable in discussing career opportunities in data science and agriculture.

Camp Design

Creation of the curriculum was a collaborative process that began in January of 2018. Through the faculty's varying research interests, certain lessons were divvied up between instructors. There were a few key concepts that needed to be addressed in the curriculum by USDA's camp standards, such as the urban forestry, food nutrition and urban agriculture. There had to be a project completed by the end of the camp as well. Aside from those components, faculty had free range as to how the USDA's goals and vision for the program were translated into lesson plans.

Several iterations of the curriculum were hashed out before there was a final, polished body of work for the students. Edits and assistance from GW Teach students, an undergraduate program preparing Science, Technology, Engineering and Mathematics (STEM) majors to become teachers, proved to be very helpful. Curriculum development was the influential aspect of the camp, as it makes or breaks the quality of the experience for the students. How STEM is traditionally taught in schools has been the subject of much research and debate. The goals for the instruction of this program was to create lessons that were relevant and most importantly, interesting to the students.

The camp curriculum consisted of lectures, lessons and exercises regarding the three main topics of Urban Agriculture, Urban Forestry and Food Nutrition as well as how to handle data. The focus of the first week was more lecture based and in-class exercises to understand the topics and the tools. The focus of the second week was to have the students group up and deliver a 10-15-minute presentation where they analyzed data on a topic that they found interesting.

The days were structured relatively the same. Each day was from 9 am till 4:30pm. The first thirty minutes of each day the students came in took the seats, set up their name tags and checked in with instructors. The next two and a half hours were used for the morning session, each day there was a specific topic that was focused on for this time block. Then students had a free period for lunch from 12-1. The next two hours were for the afternoon session, where each day an instructor took lead and had lecture with in-class exercises. The last half hour of the day was a wrap up so that students could ask any questions. There were variations on this schedule based on the goals of the day. For example, some days either the morning or afternoon session would be split in half to cover two different topics. Other days the afternoon session was a continuation of the morning. Our schedule is shown below.

1. **July 23** - Introductory lessons about data and food nutrition
2. **July 24** - Urban Forestry, Open Tree Map, Microsoft Excel
3. **July 25** - Urban Agriculture and Tableau. Afternoon session with USDA speaker
4. **July 26** - Analyzing and visualizing data using ESRI ARCGIS
5. **July 27** - Campus tour and visit to the garden. Create proposal for final project
6. **July 30** - Finalize project proposal and begin data analysis
7. **July 31** - Final data review and work on data visualization
8. **Aug 1** - Design Studio and Presentation Review
9. **Aug 2** - Design Studio and Final Presentations Review
10. **Aug 3** - Final Presentations, Evaluations, and Celebration

Example of a Camp Activity

One example of an activity completed during the program was the in-class lecture on data collection. The details are in the Supplementary Documents file for this article.

Student Presentations

The student groups chose the topics that they wished to focus on. The topics ranged from trends in American diets, asthma, tree canopy coverage in cities, farmers' markets and gentrification of neighborhoods. The students were given freedom to design any type of research question that they found interesting as long as they were in the realm of our three main research areas. Student got into groups and submitted a one-page proposal by the final day of the first week. See Supplementary Documents for details.

Outreach

In January, outreach became a key priority in making the camp successful. The minimum number of accepted students was 15, with the maximum at 30. In past USDA STEAM Camps, there had been difficulties with getting a sizable number of applicants. Deliverables were created to have a variety of ways of outreach to target high-school students. Through both the USDA and GW, faculty flyers, phone calls, emails, and visits were used to reach out to local educators and STEM organizations. Email messages were sent to all DC public libraries to ask if the flyer could be posted on bulletin boards, as well as to all principles of DC public high schools. Additional organizations were part of the outreach as well; such as the YMCA, GW Teach, Big Brothers Big Sisters, Boy Scouts of America, Girl Scouts of America, Future Farmers of America, 4-H, STEM for Her, Black Data Processing Associates, Girl Inc., D.C. Community Centers, and private high schools in the DMV area. The organizations were selected based on their ability to relay to the message to high school students.

The final group of 25 participants in the camp are our research population, and the small number is a limitation for generalization of the results. The larger sample of 56+ who applied was, however, representative of the larger audience because of the particular schools and organizations we targeted in the advertising: students in DC and nearby suburbs with similar demographics, including a common theme of being in schools that meet the federal guidelines for high-need districts. Many more similar students could have been admitted, but due to limited resources we had to restrict to 25 participants. The process of restricting the number included guiding the decision process to increase diversity of demographics and educational backgrounds. Thus, the significance of the assessment results is much better than suggested by the participant group size.

Application Process

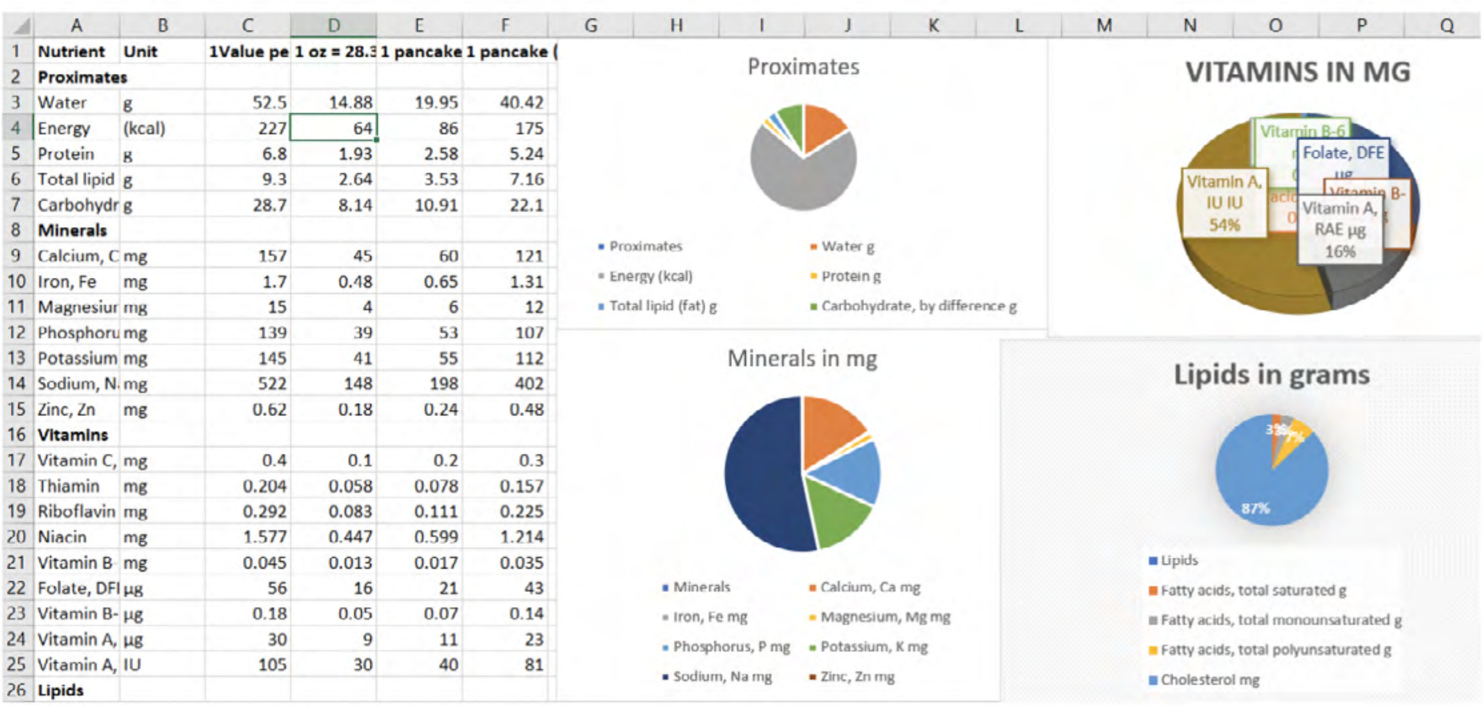
From March until May, students sent in their applications to be evaluated for the camp. The final count of applicants was 56 students. We created a website to field applications and for prospective students to find more information. Students were required to fill out the applications via the blog site hosted through the GW Data Science Program. The application required information including name, age, school, experience with data, and research interests. These questions were used to evaluate if they were eligible to participate due to age and grade. The questions regarding experience with data practices and research interest were used to understand their level of interest and concentration in the camp. Students also submitted an essay addressing three questions about why they wanted to attend, their knowledge of data and knowledge of the USDA. The submission was a short essay of 300-500 words. All the components of the application were evaluated to pick the final cohort of students.

Students

Twenty-five students completed this program. They were between the ages of 14-17, with the average, median and mode age at 16. Students were mostly from Maryland, DC, and Virginia, with one student from Delaware. Private, public, charter and homeschooled students made up the diverse group. These students had a daily commute to the GW Foggy Bottom campus each day via metro, car and/or train. The students were broken up into five groups of five. GW faculty created these groups according to information submitted via students' applications such as research interest, gender and experience with data analysis. By research interest, there were two Food Nutrition groups, two Urban Forestry groups and one Urban Agriculture group. The groups were diverse based on gender, state, school, and prior knowledge.

Software Tools

A bulk of the camp was dedicated to learning and understanding software tools relevant in both data science and STEAM concentrations. Arguably, one of the most enticing aspects of the camp to parents and students alike is



	A	B	C	D	E	F
1	Nutrient	Unit	1Value pe	1 oz = 28.3	1 pancake	1 pancake (
2	Proximates					
3	Water	g	52.5	14.88	19.95	40.42
4	Energy	(kcal)	227	64	86	175
5	Protein	g	6.8	1.93	2.58	5.24
6	Total lipid	g	9.3	2.64	3.53	7.16
7	Carbohydr	g	28.7	8.14	10.91	22.1
8	Minerals					
9	Calcium, C	mg	157	45	60	121
10	Iron, Fe	mg	1.7	0.48	0.65	1.31
11	Magnesiur	mg	15	4	6	12
12	Phosphoru	mg	139	39	53	107
13	Potassium	mg	145	41	55	112
14	Sodium, N	mg	522	148	198	402
15	Zinc, Zn	mg	0.62	0.18	0.24	0.48
16	Vitamins					
17	Vitamin C,	mg	0.4	0.1	0.2	0.3
18	Thiamin	mg	0.204	0.058	0.078	0.157
19	Riboflavin	mg	0.292	0.083	0.111	0.225
20	Niacin	mg	1.577	0.447	0.599	1.214
21	Vitamin B	mg	0.045	0.013	0.017	0.035
22	Folate, DFI	µg	56	16	21	43
23	Vitamin B-	µg	0.18	0.05	0.07	0.14
24	Vitamin A,	µg	30	9	11	23
25	Vitamin A,	IU	105	30	40	81
26	Lipids					

Figure 1: Introductory lesson in learning the basic visualization functions of Microsoft Excel. Data was pulled from the USDA Food Composition Database. https://ndb.nal.usda.gov/ndb/search/list.

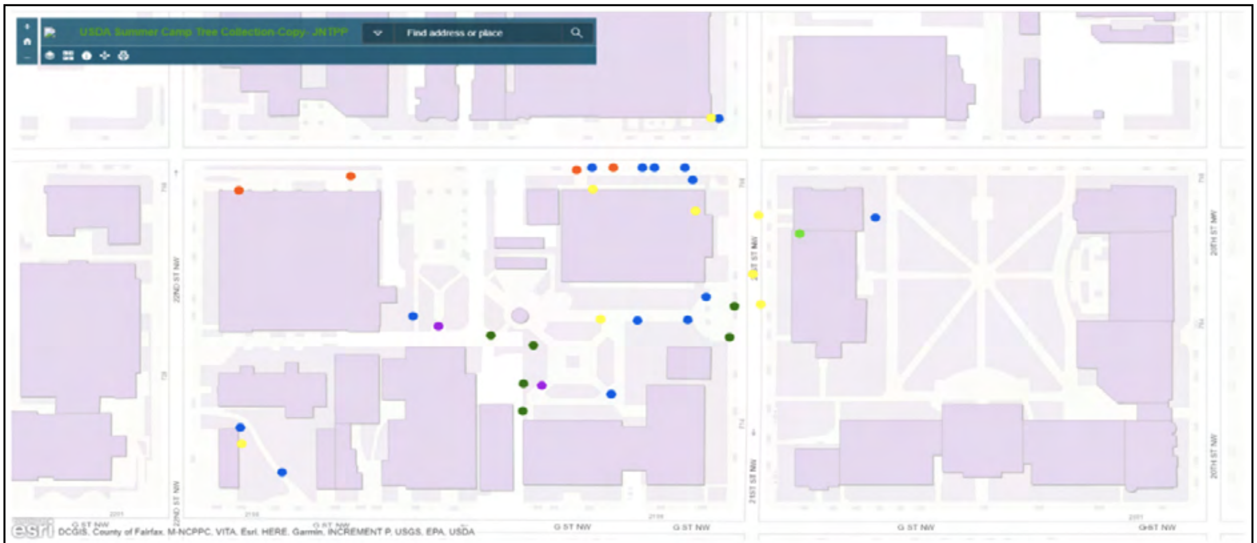


Figure 2. Mapped trees using Open Tree Map on the George Washington Campus created by camp students

learning these high-skill programs. Throughout the camp, multiple software tools were introduced during lessons and taught to the students.

The Microsoft Office Suite is a widely used family of software. The software within the Microsoft Suite that we found particularly important was Excel. While working with data, it is inevitable to work with spreadsheets, as it is the base on what makes a data project. Excel is a very commonly used spreadsheet program, throughout the workforce and in higher education. Failure to understand basic functions, commands or tools within Excel would severely impact the quality of the projects for the camp. Many students self-identified as having limited experience with the program through the evaluation given to them on the first day to assess their skills. Because of this demand, GW faculty created an in-depth presentation describing the basic functions within Excel and how they can be implemented. For example, they were taught how to use the various arithmetic functions. Students were also taught how to make visualizations in Excel through pie, bar, and line charts. This exercise showcased the simple visualizations that can be made just through Excel to represent data. This was a great warm-up for students before Tableau was introduced.

Tableau is used for visualizing data in innovative ways, particularly in the business sector. A representative from the company came in to give a presentation on the functions of the program, giving the students a real-world understanding on how the product is used. Data, in the form of a csv or Excel file, is uploaded into the program, where it can then be used for the easy creation of graphs. Tableau enables more fine control over how the graphs are displayed, which can create a more visually appealing way to explain data. Having knowledge of this program was integral in visualizing data in the students' final projects.

Mapping software was important for the students to learn given that the camp is dedicated to STEAM learning. Fine cartographical maps bring an understanding of scale and visualize data in a distinct way. Urban agriculture, urban forestry and food nutrition are given geographical context when mapping software is used. For these reasons, it seemed only natural to introduce mapping software to the students. The GW faculty incorporated the ESRI's ArcGIS and Open Tree Map into the mapping sessions. Open Tree Map is a program that allows the public to create a comprehensive and informative inventory of trees in local communities. See Figure 2. This mapping tool best visualized the prolific nature of urban forests, especially in the DC area. This program is a prime example of an open source collaborative mapping software. It is an example of some of the specialized software tools that are created for urban agriculture.

ArcGIS is a program that provides contextual tools for mapping and spatial reasoning. There are various platforms that ArcGIS is hosted through. In the camp the students used ArcGIS Online. This program has publicly available data sets that the students can then use for the creation of different types of maps. The students have multiple options for how maps can display the data for a presentation, much like Tableau. A representative from ESRI came to give a lecture to the class about the software and held a mapping exercise. Using the phone application, Collector for ArcGIS, students were instructed to map trees surrounding the George Washington campus. The application allows the user to indicate trees in urban areas by submitting information about species, size and location. See Figure 2. Users also could take a picture of the tree being recorded through the application. This exercise provided our students with a real-life example as to how collaborative data collection can impact urban areas.

Evaluation Process

An evaluation form was filled out by the students on Day 1, and the same evaluation form was used on the final day of the camp. This way, a comparison could be made to show what they thought they had learned over the duration of the camp. For a follow-on research project, our self-report evaluation survey could be tested for reliability, or consistency of measurement, using a test-retest method. The instrument had content validity, as judged by the faculty and staff. Students self-reported their skill levels in the different software that were used during the camp, as well as self-reporting their knowledge regarding how to handle data. This allowed us to see if the students gained confidence in their data analysis skills and in the use of certain programs. The students on the first day filled out the survey forms to the best of their abilities and were told that the instructors would be using this data to adjust the camp plans as needed based on the answers that the group provided. On the final day of camp, the students again filled out the evaluation to the best of their abilities and were told that the questions may seem familiar from taking the survey on the first day.

A view of the first day of camp evaluation can be seen in the table below. The final day of camp had a few additional questions to help us hear the opinions of the students on the program as a summative evaluation to enhance and improve the program for a following year. These questions are: What did you learn from this camp? Did you have a favorite lesson, if so please list? How did this camp match your expectations? and Would you recommend this camp to a friend?

An additional evaluation method was observation of the student projects and final presentations. The consensus among the faculty and staff was that all student groups ended the camp with remarkably professional project results and presentations. A key to this degree of success was to include individual practice sessions with formative feedback from the faculty and other students.

Results

The evaluations showed the students' skill levels and how they changed during the camp. It also provided an estimate of how well the students understand the different tools and how to work with data.

Self-Reported Knowledge

This first table below shows the self-reported knowledge of the students on a scale of 1-5, with 1 being the worst and 5 being the best, in a few different categories.

Items in the Pre-Survey Given on the First Day of the Camp				
Name:		Age:	School:	
Camp topic of interest:				
Define Data Science:				
List Excel functions:				
Explain the importance of data visualization:				
Rank your knowledge of the following (1-5, with 5 best)				
Excel:	Tableau:	GIS:	Data transformation:	Data storage:
Data collection:	Data limitations:	Presentation skills:		
Comments:				

Table 1. Items in the Pre-Survey Given on the First Day of the Camp

These categories are knowledge of Excel, GIS, Tableau, data transformation, data storage, data collection, data limitations and presentations skills. These questions were asked in the exact same manner on the first and final day of camp. Overall, there is a shift from lower numbers to higher numbers in all the topic areas. Some topics had more of a shift than others For example, 83% of the students said that their confidence in Tableau was a $^1/_5$ on day 1 and on day 10 only 4% of the students felt that they were still at a $^1/_5$ with 17% of the students feeling that they were at a 5/5 and 38% of the students at a $^4/_5$. However, the presentation skills showed less signs of improvement. This is mostly due to students feeling that they have excellent presentation skills before the camp began. Presentation skills went from mostly 3 and 4s to mostly 4s and 5s. The average self-reported knowledge level for all topics was a 2 on day 1 and a 4 on the final day. This is a 40% increase in the overall confidence that this group of students had in these topic areas. Overall, this shows that the students felt that they had learned about these different topic areas and developed the confidence to work within them. A stacked bar chart of this information is displayed in Figure 3 below.

Excel Skills

The following charts represent the answers to the questions on Excel. One question asked students to list Excel functions. This question was repeated in the same manner on Day 1 and Day 10. On Day 1, most of the students could not list a single Excel function. On Day 10, most of the students listed 7 different functions, and only 2 students were unable to list any. On both days the highest number listed was 8, and this went from 1 student on day 1 to 2 students on day 2. There are many more functions mentioned on the final day of camp than on the first day as seen in Figure 4 displayed below.

Student Expectations and Recommendations

The final charts show the questions that were added to the final evaluation. One was, "Did the camp match your expectations?" 33% of the students said that it exceeded their expectations, and 63% of the students said that it matched. The other question was, "Would you recommend this camp to a friend?" 92% of the students said yes.

Discussion, Implications, and Limitations

Evaluation Results

These survey results show that the students achieved the learning objectives to a high degree from this camp. The students came with little background and were able to quickly adapt to the material and work on their own and in groups to develop projects. The students gained confidence in working with data through this program. The use of projects, lessons and in-class exercises with supervision provided the support that the students needed to channel their energy and focus on the work, gaining confidence over time through exposure to the new material. The students performed very well in the creation and presentation of their projects, a result the faculty was happy to witness. The camp produced the intended result USDA and GW hoped for: students receiving an education in STEAM concentrations while applying concepts related to data.

The assessments of the camp indicate success in achieving the learning goals and usefulness of the camp for all the students. In future offerings, we plan to do a more detailed pre-post assessment along the lines of the process for the biological outreach program of (Tsybulsky, Dodick, and Camhi, 2017) and (Tsybuosky, 2018) with their success in measuring improvements in students' understanding of the nature of science and on their attitudes towards science. While the camp was successful, a few things should be kept in mind for future STEAM camps and implementations of our camp design by others. Since this was GW's first year holding this camp, faculty inevitably ran into aspects of the camp that could be perfected.

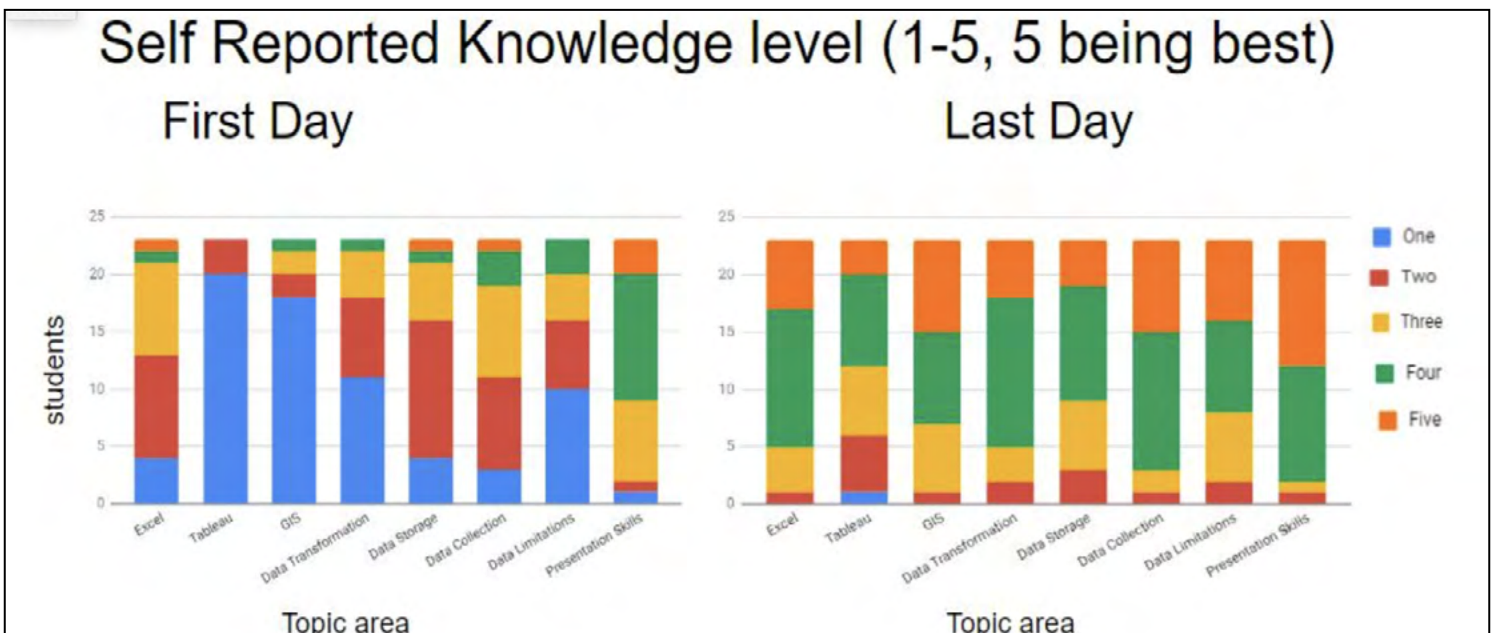


Figure 3: This stacked bar chart shows the self-reported knowledge value on a rating of 1 to 5. The same eight questions were asked on the first day and the last day of the program. The answers did shift from lower to higher over the duration of the camp, showing that the students feel that they gained skills in these categories.

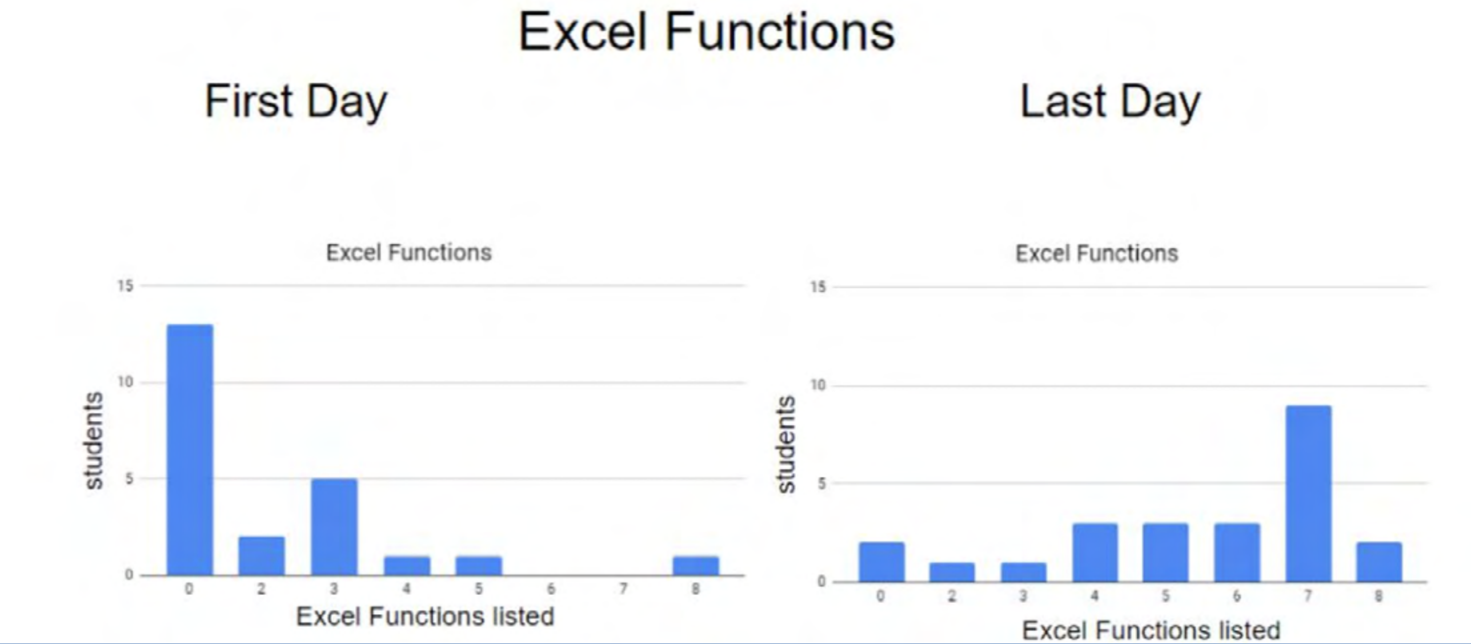


Figure 4. This bar chart shows the number of Excel functions that students were able to name. Students learned new functions during the program. Many students were able to name more functions on the last day of class than the first.

Application Process

With the initial outreach, GW faculty's main concern was to reach as many students as possible. There were more applications submitted than spots available; with this high demand, it is important to best optimize the procedures used to pick candidates. For the application requirements, it would be helpful to have more rigorous academic standards for evaluating the 11 applicants. In the future, GPA and teacher recommendations would be included as part of the application. It is possible that the application was perceived as easy, which may have led to students applying who did not a serious interest in participating in the program. Therefore, applying more strenuous application requirements may alleviate this issue.

Access to laptops

After orientation for the camp, faculty sent out a questionnaire to students assessing their access to laptops, the Microsoft Office Suite, and other supplemental information needed. Over 25% of the students did not have access to a laptop. Adding this question within the initial application would fix this issue and allow faculty to know the number of laptops needed beforehand.

Increased outreach to DC/VA schools

Through the outreach between the GW and USDA faculty, there was significantly more MD applicants than DC or VA. The camp's mission was to have students from all three areas represented. Perhaps in the future, outreach can be divided up within faculty by region, so there isn't a disproportionate number of students from one area.

Volunteers

In addition to the research assistants who instructed camp lessons, teaching aids were brought in to assist in class. This proved to massively helpful, as there were the same number of staff members as there were student groups. For future camps, having teaching aids from GW who are interested in a career in STEM education may find this opportunity especially appealing. Enlisting teaching aids earlier would assist faculty, as we could have them involved in the creative process and development of the camp.

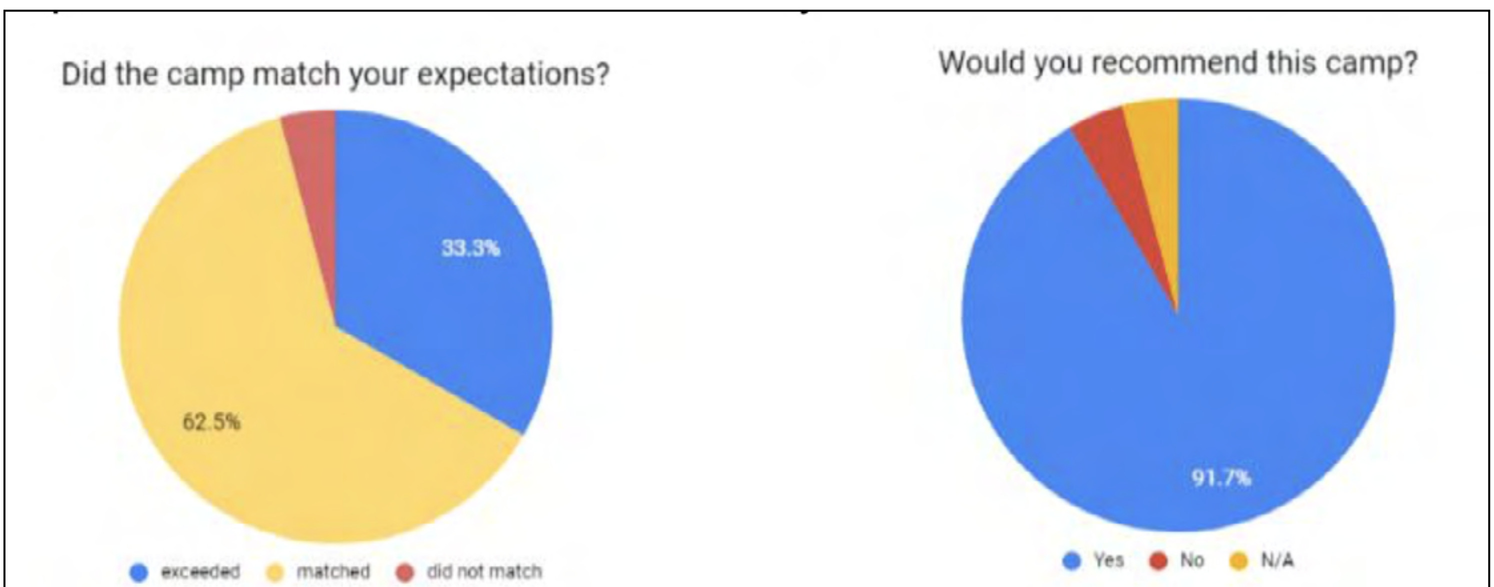


Figure 5. Results for additional questions in the final survey on the last day of the camp.

Evaluations

The evaluations format of having all questions regarding knowledge and skill be self-reported is a data limitation. This allows us an insight into the confidence level that the students had from before and after the camp. However, it does not provide a definitive answer on how much they learned and what an unbiased opinion would be of both their individual knowledge of the subject matter and their skill level when using the different software products. A potential future improvement would be to create a test on top of the self-reported section that could provide an appropriate grade for each student on the different topics that they are evaluated on. This would allow for an additional metric to measure how much the students learned from this camp.

Implications and Limitations

One limitation of the evaluations is that there are only 25 students. This is a small sample size. It is viewed that the larger samples have more reliable results, because it is easier to distinguish between a true variation and random noise. In general, the positive view of small studies is due to it taking less time and being easier to track the participants (Hackshaw, 2008). This program had a small number of participants due to limited resources. However, there were many more who applied. The pool of those who applied and those that were accepted had many of the same qualifications and characteristics, such as age range and education level. The full pool of applicants was generated through targeted advertising to get a set of applications from our target audience. Therefore, we assume that the students that completed the evaluation are representative of the type of student that would be interested in this type of a program. While, we are aware that this is a small sample, we do believe it to be representative.

Another limitation is that the survey evaluations are self-reported. Self-reported measurements were used because one of the goals was to see how the students' confidence in their skills changed over the course of the program. Self-reported information can be useful because it tends to be economical and quick to attain (Gonyea, 2005). These evaluations had both factual and attitudinal questions. Listing out all of the excel functions that one knows is factual. Stating how one feels one's own presentation skills are is an attitudinal question. Attitudinal questions are not easy to assess nor prove due to their subjective nature (Gonyea, 2005). In the future there can be added measures to the evaluation. That way the factual questions can be given further assessment. This type of evaluation only shows how the students view themselves. This is helpful information as teachers can use it to view how the students gain confidence in their work and if the students feel that they have learned from the program. A second evaluation form that asked students to go through a short data analysis project or a type of exam that had students detail out what excel functions are and to define a few or short answer questions on the different aspects of data handling can be useful in understanding what exactly students know. This type of exam can also be given on both the first day and the last day, so that the instructors can see exactly how much the students learned. This will also allow for instructors to understand where the students are lacking on the first day and make sure to cover those topics early on, so that students do not feel as if they are falling behind.

Acknowledgements

Thank you to all of these great organizations for working with us to put on an excellent camp: Chrishon Blackwell and Georgette Edmondson-Wright, GW Summer Program, contributed time and t-shirts to the USDA STEAM students. The CCAS Dean's office provided implementation logistic support throughout the camp. USDA personnel Bobby Jones, Dennis Crow, and Cynthia Larkins provided ideas, recruiting assistance, and constant support during the camp. Tableau professional Mike Boyle participated in sessions and Tableau donated software and exciting visualization presentations with Tableau products. ESRI professionals Nick Short and Kristen Hocutt donated software and interactive presentations using ESRI. The GW Libraries and Academic Innovation staff member Kean McDermott gave a great lecture on GIS.

References

1. (Beichner and Saul, 2003) Beichner, Robert & Saul, Jeff. (2003). Introduction to the SCALE-UP (Student-Centered Activities for Large Enrollment Undergraduate Programs) Project.
2. (Bybee, 2015) Bybee, R. W. (2013). The case for STEM education: Challenges and opportunities. Arlington, VA: NSTA Press.
3. (Gonvea, 2005) Gonyea, R. M. (2005). Self-reported data in institutional research: Review and recommendations. New Directions for Institutional Research, 127, 73–89. doi: 10.1002/ir.156.
4. (Hackshaw, 2008) Hackshaw A. (2008). Small Studies: Strengths and Limitations. European Respiratory Journal. doi. https://doi.org/10.1183/09031936.00136408.
5. (Levin and Tsybulsky, 2017a) Levin, I. & Tsybulsky, D. (Eds.). (2017). Digital Tools and Solutions for Inquiry-based STEM Learning. 371 pages. Hershey, PA: IGI Global. doi:10.4018/978-1-5225-2525-7.
6. (Levin and Tsybulsky, 2017b) Levin, I. & Tsybulsky D. (2017). The constructionist learning approach in the digital age. Creative Education, 8(15), 2463-2475. doi: 10.4236/ce.2017.815169.
7. (Lewis and Matsumoto, 2017) Lewis, A., & Matsumoto, G. (2017). Science Educator Professional Development: Big Data and Inquiry Learning. In Levin, I. and Tsybulsky, D. (Eds), Optimizing STEM Education with Advanced ICTs and Simulations (pp. 219-244). IGI Global.
8. (Lombardi, 2007) Lombardi, M. M. (2007). Authentic learning for the 21st century: An overview. Educause learning initiative, 1(2007), 1-12.
9. (Madden et al., 2013) Madden, M. E., Baxter, M., Beauchamp, H., Bouchard, K., Habermas, D., Huff, M., ... & Plague, G. (2013). Rethinking STEM education: An interdisciplinary STEAM curriculum. Procedia Computer Science, 20, 541-546.
10. (Moylan, 2008) Moylan, W. A. (2008). Learning by project: Developing essential 21st century skills using student team projects. International Journal of Learning, 15(9).
11. (NRC, 2013) National Research Council. (2013). Education for life and work: Developing transferable knowledge and skills in the 21st century. National Academies Press.
12. (SERC, 2018) Pedagogy in Action (2018) SCALE-UP https://serc.carleton.edu/sp/library/scaleup/index.htm.
13. (Teodorescu, Bennhold, Feldman, and Medsker, 2013) Teodorescu, R., Bennhold, C., Feldman, G., and L. Medsker, L. (2013) "A New Approach to Analyzing Physics Problems: A Taxonomy of Introductory Physics Problems," Physical Review Special Topics – Physics Education Research.
14. (Tsybuosky, 2018) Tsybulsky, D. (2018). Comparing the impact of two science-as-inquiry methods on the NOS understanding of high-school biology students. Science & Education, 27(7), 661-683. doi: 10.1007/s11191-018-0001-0. 14
15. (Tsybulsky, Dodick, and Camhi, 2017) Tsybulsky, D., Dodick, J. & Camhi, J. (2017). High-school students in university research labs? Implementing an outreach model based on the 'science as inquiry' approach. Journal of Biological Education, 52(4), 415-428. doi: 10.1080/00219266.2017.1403360.
16. (Tsybulsky and Levin, 2017) Tsybulsky, D. & Levin I. (2017). Inquiry-based science education and the digital research triad. In Levin, I. and Tsybulsky, D. (Eds), Digital Tools and Solutions for Inquiry-Based STEM Learning (pp. 140-165). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-2525-7.ch006.
17. Koonce, David A., et al. "AC 2011-289: What is STEM." American Society for Engineering Education. 2011.

Jasmine Sami is a first-year doctoral student in the Human-Technology Collaboration PhD at George Washington University. Jasmine has an MS in Data Science from George Washington University and a BS in Sociology at Catholic University of America. She has spent the past three years working on the NSF-funded Noyce Scholarship Grant, which provides funding for STEM undergraduate students to pursue a career teaching in high-need communities. Her research interests are in education technology and the future of work.

Dr. Larry Medsker has 30 years of experience in shaping cutting edge research on the impact data and analytical systems have had in areas including artificial intelligence, hybrid intelligent systems, and human technology support systems. His work in academia, industry and government settings includes leading the data science programs at both George Washington University and Merrimack College, as well as authoring four books, and more than 100 articles in the areas of Artificial Intelligence and Physics. Dr. Medsker is currently a Research Professor at George Washington University.

Krystin Sinclair is currently working as the Americorps VISTA Camden Coalition Health Data Analytics Associate. Krystin has an MS in Data Science from George Washington University. During her time at GW she worked as a Graduate Assistant for the Milken Institute School of Public Health. She is originally from New York, but most recently lived in Connecticut where she worked as an analyst in the insurance industry. Prior to that Krystin completed a Bachelor's Degree in Applied Mathematics and Statistics from Bryant University with a concentration in Spanish and a quadruple minor in Business Administration, Finance, Environmental Science and Biotechnology.

Zachary Stein is originally from San Antonio, and has spent a good part of his life dedicated to data science. He worked for years in a lab running simulations on how NMR spectrometer experiments worked and how that influenced the data that was collected. After a while, he decided to expand his skills by getting a degree in data science from George Washington University. He now works at Booz Allen Hamilton, where he applies those skills to help organize data for the Federal Government.

Data Science Outreach Educational Program for High School Students Focused in Agriculture

Supplemental Documents

Jasmine Sami, Krystin Sinclair, Zachary Stein, Larry Medsker
George Washington University

Table of Contents

1. Consent form

PARENTAL WAIVER AND ASSUMPTION OF RISK FOR CHILD VISIT TO GW

I am aware that ____ ______ ____ (my child, who is a minor, will participate in a day camp at the George Washington University ("GW") for a USDA Science, Technology, Engineering, Agriculture, and Math (STEAM) Summer Camp from July 23 through August 3, 2018, 9AM-5PM Monday through Friday ("Visit to GW") and the schedule of activities may include tours of GW, eating meals on GW's campus, analyzing open source data, the opportunity to learn from faculty while completing projects related to urban forestry, urban agriculture, and food nutrition, and other similar activities or activities incidental thereto. I understand that my child's participation in those activities, as a part of the Visit to GW, may have risks associated with them, including the risk of physical injury or sickness and property damage as well as other unknown hazards. I understand and agree that GW cannot be expected to control all of said risks. My child and I understand that any of the activities my child voluntarily elects to participate in are done so at our, and not GW's, sole risk and responsibility. While on GW property, my child agrees to abide by all rules and guidelines applicable to visitors to GW and its classrooms and to all other safety rules that may hereafter be given or explained to me or my child in connection with the Visit to GW.

ASSUMPTION OF RISK: Because my child's participation in the Visit to GW is voluntary, GW shall not be responsible for any injuries my child may incur, and I hereby assume all risks and all responsibility involved in my child's participating in the Visit to GW. To the best of my knowledge, my child is able to safely participate in the activities associated with the Visit to GW.

RELEASE: In consideration of my child's voluntary participation in the Visit to GW, I hereby expressly and knowingly assume all of the above risks and forever voluntarily release, discharge, waive, and relinquish, and covenant not to sue any and all actions or causes of action for personal injury, property damage, or wrongful death occurring to my child arising as a result of participating in such activity or activities incidental thereto wherever or however the same may occur and for whatever period said activities may continue. **I do hereby forever release, waive, discharge, and relinquish, and covenant not to sue for any action or causes of action, aforesaid, which executors, administrators, and assigns prosecute, present any claim for personal injury, property damage, or wrongful death against GW its employees, and agents for any of said causes of action, whether the same shall arise by the negligence of any of said persons or otherwise.** In other words, I agree that I and my child cannot sue or recover anything from GW or any of its trustees, officers, agents or employees if I or my child is injured in any way, or if anything happens to me or my child's person or property, including as a result of any claimed past or future negligence or carelessness of GW, or while preparing for, traveling to or from, or participating in any part of the Visit to GW. This release and waiver shall be binding on myself, my heirs, executors and administrators, and assigns. I further agree that this release and waiver shall be construed in accordance with the laws of the District of Columbia, and that if any portion is deemed to be invalid, the remainder of this release and waiver will still be binding.

I HAVE READ THIS "PARENTAL WAIVER AND ASSUMPTION OF RISK FOR CHILD VISIT TO GW" AND I FULLY UNDERSTAND THERE ARE RISKS OF INJURY INVOLVED IN THE USE OF GW FACILITIES AND EQUIPMENT AND THAT I, ON BEHALF OF MY CHILD, VOLUNTARILY ASSUME SUCH RISK.

Signature of Parent/Legal Guardian	Name of Visit to GW Participant
Printed Name of Parent/Guardian	Date

Emergency Contact Name(s) and Phone Number(s)*: ____________________

*If cellular number(s), may such numbers receive emergency text messages? Please circle: **YES/NO**

2. Survey Form Data Fields

Name:
Age:
School:
Area of Interest:

Define Data Science

List Excel Functions

Explain importance of data visualization

Rank 1-5 (5 being best)
Knowledge of Excel
Knowledge of Tableau
Knowledge of GIS
Knowledge of data transformation
Knowledge of data storage
Knowledge of data collection
Knowledge of data limitations
Presentation skills

What did you learn from this camp?

Did you have a favorite lesson, if so please list?

How did this camp match your expectations?

Would you recommend this camp to a friend?

Comments

2. Flyer for Advertising the Camp

2018 USDA OPEN DATA SUMMER CAMP

Questions?

Dr. Larry Medsker, GWU, lrm@gwu.edu
Bobby Jones, USDA,
Bobby.Jones@ocio.usda.gov

To apply visit us at
https://dataprograms.gwu.edu/usda-open-data-summer-camp/

OUR PROGRAM

USDA and the Data Science Program at George Washington University are teaming up to design and deliver a **FREE** "summer camp" in 2018 for high-school students (ages 15-17) in the Washington Metro Area that focuses on using Open Data related to Science, Technology, Engineering, Agriculture, and Math (STEAM).

CAMP OVERVIEW

This two-week long project-based Summer Camp will introduce participants to the use of data in the context of the USDA's work with food nutrition, forestry, and urban agriculture. Students will explore, analyze, and reconfigure quantitative and qualitative data, using fundamental graphical principles to present their project-related findings.

WHEN AND WHERE

The USDA Open Data Summer Camp will take place at George Washington University **In Corcoran Hall** from **July 23 – Aug 3 2018.** Camp hours are from 9 AM-4:30 PM. Transportation not provided.

3. Example of a Camp Activity

One example of an activity completed during the program was the in-class lecture on data collection. This activity started with the instructor asking the students to raise their hand if they answered yes to a question. These questions were Do you have a smart phone? Have you used excel before? and Are you tired this morning?. The yes and no answers were tallied on the white board. This way some basic data collection was done in a way that the students knew that they had collected data before. Then the instructors discussed the different types of data such as quantitative and qualitative. The instructors also discussed variable types such as nominal, ordinal, integral and ratio, as well as discrete and continuous variables and dependent and independent variables. This way students could learn about the different types of data that they might need to collect.

The lecture lead to small group discussions of different ways to collect these types of data. Students discussed surveys, quizzes, case studies, interviews, observations, testimonials among many other ways to gain information. Then as a class sampling methods and sampling error were discussed. The instructors then showed students are ways to search for data online. The students learned about Kaggle, data.gov, and other resources. Then the students broke into small groups to discuss what must be considered about data before any analysis can be done. The students came up with many ideas, such as making sure that it is reliable, valid, the source is trustworthy or credible and that it is from an open data set and not proprietary. The instructors then asked the students to discuss in small groups how to develop a SMART research question and what data would be needed to answer that question. This leads to the class discussing limitations of data, such as what can a researcher control, and how to think about bias. We followed with a lecture on how to create effective visualizations. The students were asked what they already knew about data visualization and they already knew that simple can be better and they were able to name many different types of charts and graphs. The instructors reviewed when to use the different types of charts and graphs and then how to make them in Excel or Google Sheets, depending on which platform the students preferred. This ended with an instructor lecture on how to search for patterns in data and how to use numbers to explain the whole story.

4. Example of a Student Proposal

The student groups chose the topics that they wished to focus on. The topics ranged from trends in American diets to Asthma and tree canopy coverage in cities to farmers markets and gentrification of neighborhoods. The students were given free rein to design any type of research question that they found interesting.

The students submitted a short proposal as a group by the final day of the first week. The proposal consisted of a research question, explaining what made it a SMART question and how they intended to answer it along with at least one dataset reference and scholarly article reference. The references allowed the instructors to know that the students had done preliminary research. This gave the instructors to read over the proposals and provide feedback before the students got too deep in the weeds of the project. The proposal on how trees can benefit Urban areas is shown below.

Do Trees Really Benefit Urban Areas?

Preface

To improve the quality of life by planting trees in the greater Chicago area; and analyzing data that affects the lake Michigan ecosystem. This research proposal will be looking at the effect trees have on the Greater Chicago ecosystem and urban environments throughout the world.

Population Density Compared to Tree Canopy Cover

The population density in the greater Chicago area as far north as the Evanston area and south to the John Sherman Park area will be studied. It has been found that the neighborhoods with more people have a greater percent of canopy cover except for in the most populated districts. These are the sections of the city will be focused on the most as they will give better quality results. The research project deals with improving the quality of life for the average Chicagoan through the wellbeing and health benefits associated with living near trees. "Chicago wants to increase its canopy cover from 14.6% to 17% by 2020." 1*

Air Quality

Through the course of the project or presentation, the effects of tree canopy cover on urban environments will be studied. One area that would be used is the amount of pollutants that various trees in an urban area have been able to remove from the air. Some of the sources that have been identified as potential resources are an article from the USDA and another from the University of Florida. Some of the issues with the sources are that not all of them presented all the raw data. This restricts how much information can be used since the data has already been cleaned. Though this can hinder the research process it also saves time because the data has been simplified. Another area that will be explored is Chicago and from findings it was an overall good air quality limit, and a 50/50% change of air quality from good to moderate, and unhealthy air was at a 0.8% minimum.(State of Illinois, Illinois Environmental Protection Agency) In conclusion to these percentages, Chicago is found to have an overall good air quality. 5

Erosion

Another important benefit to trees is their role in erosion control. Trees near water bodies help control water runoff and prevent excess sand from entering these bodies. Chris Walker said that the Illinois Beach State Park water level is on the rise. Before there was four hundred or five hundred feet of sand and now, they are lucky to have 6 feet of sand in a single stretch of beach. Another visual of this are the rock foundations under the buildings on park land. These foundations are becoming pulled under the waves and are no longer giving support to the buildings. Every year from 1939 to 2014 Feeder Beach has lost up to 8 feet of sand. This would add up to a total of 631 feet of sand lost over time. Starting in 2014 the amount of sand lost has drastically increased to 38 feet a year.

Work plan

We will handle looking at how tree canopies affect the health of different populations by comparing the frequency of particular diseases (Tuberculosis or Pneumonia) in areas with a lot of plant life to those which have a lower ratio of plants to the number of people. We will focus on air quality by comparing the number of toxins in places with more trees per individual to those that don't. The number of toxins removed by trees at various locations will also be compared to how much of these toxins are produced or released daily. We will focus on researching and identifying erosion changes around Lake Michigan. In order to manage our time effectively so that this project is completed on time, each person in the group will spend Monday and Tuesday wrapping research and creating the various visualizations to go with them. These with information from the articles would be compiled together in a Google Sheets presentation. Wednesday will be used to finish individual work and brief the group on all work has been done. The presentation should be done by Thursday and the rest of the day will be used for review, practice and checking.

S.M.A.R.T

Specific -- Our question is specific because it can simply be answered through thorough research and investigation. This is being done in the areas of Air Quality, Erosion and Health.

Measurable -- Research to answer the question can be measured by comparing different rates and percentages of each of the above areas in their various locations through the use of visual aids and representations in Tableau, Microsoft Excel or Google Sheets.

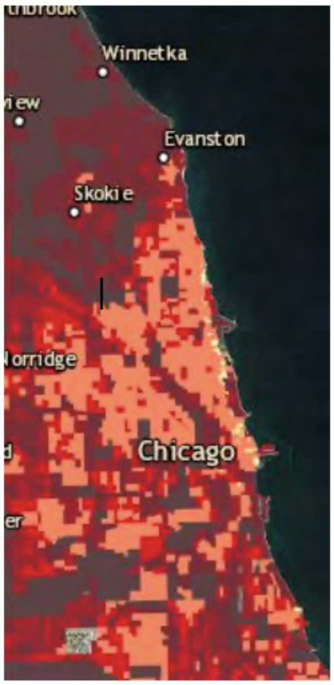
Achievable -- The answer from this experiment can be obtained after investigating or chosen areas to identify if there is a truly a difference between areas that have more trees per individuals than those that do not.

Relevant -- The results or findings from this project should not only educate people who lack knowledge in urban forestry but also show the importance of tree canopy in the urban area. This should encourage people in power to promote the growth of trees in urban environments.

Time Oriented -- The project has been planned and thought out such that extensive work is done without it being overbearing or too much to complete by Friday.

Sources

- (1*)2010 http://illinoisurbanwood.org/documents/Chicago_Urban_Wood_Report_v13.pdf - Chicago's Urban Forest Agenda: Result of Chicago's Urban Forest Effects (UFORE) model study 2007
- http://www.epa.illinois.gov/- Chicago air quality
- http://www.chicagotribune.com/news/ct-lake-michigan-sand-depletion-htmlstory.html#
- http://edis.ifas.ufl.edu/fr278#FIGURE%202
- Article by Francisco Escobedo, Jennifer A. Seitz, Wayne Zipperer, and Basil Iannone. The article provides recommendations but does not provide raw data, only information that had already been cleaned and tidied.



5. Example of a Group Presentation

The second week the students had a few hours each day to work as a group on the project to create a slide deck and presentation. The groups had multiple rounds of meetings with instructors to aid the students and provide feedback. They also had open classroom time for feedback -- where students were free to work individually or in groups and instructors would walk around from group to group asking how the project was going and giving slight direction, providing more information on how to use the software tools that were introduced during the first week, and helping find more resources and research materials. The instructors were able to use this time to make sure that all students had experience with working with data to find an answer to a question using various software tools. The final day of camp the students presented in front of the GW staff, USDA staff, family and friends of students in the program and the rest of the program. This allowed the research done by the students to be shown off as accomplishments to their parents and whomever else they wished to invite to the "graduation ceremony". One slide from the presentation on "Is there a link between Urban Agriculture and Gentrification"? is shown here.