Preliminary results from implementing a data driven team project in introductory risk and uncertainty analysis class for sophomore civil and environmental engineering students

Dr. Sotiria Koloutsou-Vakakis, University of Illinois at Urbana - Champaign

Dr. Sotiria Koloutsou-Vakakis holds a Diploma degree in Civil-Surveying Engineering (National Technical University of Athens, Greece), a M.A. in Geography (University of California, Los Angeles), and M.S. and Ph.D. degrees in Environmental Engineering (University of Illinois at Urbana-Champaign). She teaches undergraduate and graduate courses on Air Quality, Science and Environmental Policy, and Engineering Risk and Uncertainty. Her recent research is about gaseous emissions of reactive nitrogen from fertilized fields into the atmosphere and impacts on air quality and climate change, and implementing process and project learning in introductory fundamentals classes.

Preliminary results from implementing a data driven team project in introductory risk and uncertainty analysis class for sophomore civil and environmental engineering students

1. INTRODUCTION

Rapid changes in science and technology mandate that engineering education is updated to keep up pace with these changes. Computing has become an essential component of engineering education across disciplines, along with knowledge beyond the disciplinary boundaries [1] [2]. Such expanded outlook is needed for engineers to be able to understand the environmental and societal needs and contexts of the problems they are called to solve, and to evaluate the implications of the solutions they provide [3]. At the same time, engineering computational education has created a heavy (and often hard to manage) burden to computer science (CS) departments [4], pushing other departments to reorganize their own curricula to integrate computing throughout the undergraduate and graduate curricula. At the same time, there is an expanding volume of literature on professional identity and the importance of building it from the very early stages of one's education [4]. Combined consideration of all of these new challenges and needs, makes the idea of scaffolding [6] [7] through the curriculum, to be an appropriate metaphor.

In the "Body of Knowledge Outcomes" for civil engineering, the American Society of Civil Engineers (ASCE) includes: 1) risk and uncertainty as one of the technical outcomes; 2) experimental methods and data analysis as one of the engineering fundamentals outcomes; and 3) teamwork, leadership, communication, professional attitudes and ethical responsibilities, among the professional outcomes [1]. In overlapping paths, the National Academies of Sciences report on environmental engineering (EE), points to similar in-depth technical but also in-breadth communication and professional skills, to best prepare environmental engineers for addressing the "New Future", framed by increasing urban populations, changing climate and urgent needs for sustainable energy, food, water, production and waste management [2]. Data science is proposed as one of the in-breadth components of EE education.

Responding to the need to keep up with the rapid changes in science and technology, and recognizing the importance of professional skill development, a department curriculum updating effort is under way with three important focal points: computing, communication, experiential learning throughout the curriculum. In this paper, I focus on a required engineering risk and uncertainty (ER&U), sophomore level class for civil and environmental engineering (CEE) students. Because of its place at the base of the curricular scaffold, the ER&U class was the first to be updated. The recent major and interconnected updates to the class were: introduction of R as the computational tool and the introduction of a data driven team project on simple linear regression (SLR).

Project-based learning ranks high in the literature in terms of supporting learning [9] [10] [11] and professional identity development [12]. This paper focuses on the team project and shares my experience from its first iteration. Therefore, there is no deep historical perspective I can share and the current assessment is heavily leaning on qualitative data. I still believe this is useful information to share with the community, as a case study about what worked, what did not work and lessons learned moving toward the second iteration. This first assignment of the project was on exploratory grounds, aiming to overcome two major challenges: a) transfer to a big class of an activity and assessment type that the instructor has successfully used in smaller senior level classes; and b) the unpopularity of this specific class among students. Past surveys have pointed to two main reasons the class is unpopular: first, students do not immediately recognize this class as a CEE class, therefore association to CEE is a conscious, consistent effort throughout the semester; second, similar to what has been reported in the literature (e.g. [13]), the class is perceived as difficult, which mostly relates to lack of prior introduction to probabilistic thinking compounded by limited coding experience.

Description of the ER&U class
I will first provide a summary description of the ER&U class. This is one of two classes that CEE students are required to take within the CEE department, during their sophomore year, following fundamentals classes offered by other departments and an introductory CEE course offered in freshman year. Thus, the class is important in the CEE identity development process, and it comprises important part of the base of the educational scaffold, in the CEE curriculum. The class offers an applied approach to the fundamentals of probability and statistics. To relate the class with CEE, the vast majority of example and homework problems solved are drawn from the CEE context, with conscious effort to include representation from all CEE areas (e.g. structural, transportation, environmental etc.).

In recent years, the class was modified to a flipped model. Pre-lecture videos were prepared along with checkpoint questions and posted in the Blackboard/Compass2g class environment. The students are required to watch the assigned video and respond to the checkpoints before they come to class. This allows for more time in class for the students to work on example problems, following a brief overview of the main concepts of the day, by the instructor. Working on example problems is an activity that improves material comprehension and the component of the class, students have credited highest for assisting their learning. In terms of computational tools, until Spring 2019, students were using Excel but because of limitations with using Excel during exams, they were also instructed on using calculators and statistical tables on paper, for their exams. This created student resistance to use of Excel, with students complaining that they could not see why they were made to use Excel, when their exams required only a calculator and tables on paper. Therefore, one motivation behind the project implementation was to lead the students into discovering the value of computational tools. The project also created an appropriate carrier to transition to a more powerful for statistical analysis scripting tool, such as R. Use of a script language was a challenge on its own. CEE students are required to take an introductory computing class in the CS department, which currently covers coding in Python and MATLAB, but this class was not a pre-requisite for the ER&U class in the considered semester. Therefore, my assumption was that the student body of the class (similar to those in previous semesters) had limited coding experience. For this reason, R was introduced as an "smart app, useful to scientists and engineers" and it was integrated in a guided form, using base R functions, as needed for class purposes. R is taught through its use for the purposes of the class, not as an individual class topic.

2. DESCRIPTION OF THE PROJECT ACTIVITY
References to the benefits of implementation of projects in statistics courses date decades back (e.g. [13] [15]) and it is consistent with recommendations by the American Statistical

Association's Guidelines for Assessment and Instruction in Statistics Education that include integration of real data with a context and a purpose; fostering active learning; using technology to explore concepts and analyze data [16]. Though some studies have been rather inconclusive regarding a quantifiable value for 'authentic projects', with a study suggesting a higher benefit for higher achieving students, likely due to their higher ability to take ownership of the material and of the learning process [17]. The project implementation in this paper is oriented to CEE students, focuses on application of SLR and it shares some approach elements with [18] for non-statistics or mathematics majoring students. This implementation aspired rather ambitiously to a number of objectives: 1) encourage student involvement and overcome the student indifference barrier, by keeping students involved and connected with their team and the class throughout the semester; 2) promote student confidence in identifying and acquiring real datasets from existing databases and then using statistical analysis (specifically SLR) to answer questions students themselves selected; 3) facilitate material comprehension and promote critical thinking; 4) strengthen coding skills by leading students discover R's powerful computational environment, for a real world application; 5) strengthen written communication skills; 6) promote teamwork skills.

Enrollment was 103 students in the semester of first implementation. A semester comprises of 15 instructional weeks. The students were assigned by the random assignment algorithm in Blackboard/Compass2g platform, to 27 teams of 3 or 4 students. A project help period was held weekly, where student teams consulted with the instructor about the project. Consistent with in classroom scaffolding approaches [19], the project was assigned in 5 sequential parts that were aligned with the material in the class schedule. Each part was planned to have one or more special purposes toward the primary goals listed above. Parts 1 through 4 can be thought as parts of formative assessments, in the sense that students can correct, iterate and improve plans, based on instructor feedback. Part 5 was a summative assessment of the final project, after the previous steps have been completed and any misconceptions corrected.

Preparatory steps

1) Before the project was assigned, the concept of a model as a means of exploring relationships in a mathematical sense was introduced. For consistency with the level and materials taught in the class, students were informed that the ultimate goal of the project was building of a linear model to describe the relationship between two continuous numerical variables, which students were to choose during their research for Parts 1 and 2.
2) I presented to students what a Comma Separated Values (.csv) file is, emphasizing that this is a convenient data file format and the one to use in the class. I also explained how to open .txt and .xls or .xlsx files and save them to .csv format. Using only one file format is vital for keeping project management under control, as it eliminates the need of advice customized to variably formatted files.
3) I explained teamwork and team peer evaluation and teams were asked to divide labor that included one leader, with leaders rotating for the different parts of the project.
4) Teams are required be in touch with the instructor, weekly.

Part 1: Choose a topic and find the data. Assigned in week 1 and due in week 3.

Purpose: a) Students take ownership of the project: students to choose a topic of interest to them, with the expectation that this will keep their motivation and involvement high throughout the semester; b) Students learn to find and acquire data from existing databases.

Guidance: Student teams were asked to choose topics of relevance to CEE with example topics including: Infrastructure, Transportation, Energy (production, consumption, renewables), Food, Water Resources, Air Pollution, Climate, Technology (development, adaptation, use). The students were directed to use existing data in databases of governmental (local, state, federal) and intergovernmental (UN, UNEP, WHO, FAO, WTO) organizations, but they were not really restricted to data from these sources. Table 1 shows examples of suggested databases.

Deliverables: Submit one typed page with the following information:

1) Topic.
2) A 2-3 sentence justification why you find the topic interesting from a cee and/or societal perspective.
3) Information about the data you consider
    a. Source and link
    b. Brief description of the specific numerical variables you are considering.
4) Team peer evaluation.

Part 2: Refine the question. Assigned in week 4 and due in week 5.

Purpose: Students think about the topic they chose, in Part 1 and take a second look at the questions their data can address and kinds of relationships the data can possibly reveal. They have the opportunity to change their topic. No limit on the number of priority questions was set at this stage.

Guidance: Student teams work on this part in class, with the option to complete it after class, if they still wanted or needed to work on it. The process on this class activity was inspired by a Dec. 2018 ASEE (American Society for Engineering Education) webinar by the "Right Question Institute" [20]. The process comprises by successive steps of brainstorming, evaluating, prioritizing, finalizing research questions.

Deliverables: Submit

1) A record of team's in-class question generating process in the google form.
2) A max two-page (.pdf) file with the following information,
    a. Chosen priority questions as refined during the class session;
    b. A graph mapping (i.e., showing the correspondence) of your questions to the variables from the dataset you chose, that you believe hold the answers to your questions;
    c. A paragraph with a clear explanation, as to why you believe these are interesting questions and why you believe the variables you chose are relevant to the questions.
3) A single file containing the variables you will consider, in spreadsheet form (rows observations, columns variables) the file format should be .csv. The submitted file should

have a header row with simple (one or two words linked with an underscore) description of what are the contents of the columns, i.e. The variable names.

4) Team peer evaluation.

Part 3: Get acquainted with your variables: descriptive statistics. Assigned in week 7 and due in week 8.

Purpose: Student teams calculate descriptive statistics of their chosen variables using R.

Guidance: In addition to regular instruction a help session was dedicated to helping student teams importing .csv files to R. Students were required to use R functions: mean(), median(), var() , sd(), range= max() − min(), IQR(), quantile(), library(moments)/skewness() and kurtosis(), summary(), cov() and cor(). From this part, guidance was also provided on best practices for communicating information with emphasis on organization, clarity, neatness and use of citations.

Deliverables: Submit

1) Your R code (follow our class file naming convention). The script should run with the data file that you submitted in part 2. If you made revisions to the original data file by adding more data or re-organizing it, then submit your new data file (same filename rule). We need your real data file to test your R code.
2) A max 2 page, .pdf report with brief presentation of:
    a. Your data file header explained. Your column headings are good to be simple words for coding purposes. However, these simple coded names might be hard for others to decode. For this reason, in your report, create a list or a table, where in one column there is the column heading you use and in the other a full word explanation of what the variable represents with its units. For example: countr_pop: population of each developing country, in 1000 of people or garb_p_y: production of solid waste in tons per year.
    b. Tabulated summary statistics for all the variables you are considering in your data set, with a description in words of what the statistics are and (as applicable) what they indicate about the spread and shape of each variable.
    c. Anything else that you deem interesting about the variables you consider.
3) Team peer evaluation.

Part 4: Get acquainted with your variables: data visualization. Assigned in week 9 and due in week 11.

Purpose: Student teams explore different types of plots and the usefulness of alternate ways of visualizing data using R.

Guidance: Following relevant instruction, student teams were required to plot the following graphs using R library 'graphics', for one or more variables, as they thought they needed: box-and-whisker plot; parallel boxplots; histogram; relative cumulative frequency plot; normal quantile plot; basic scatterplot of two variables

Deliverables: Submit:

1) Your R code (follow our class file naming convention). The script should run with the latest data file that you have submitted. We need your real data file to test your R code.
2) A .pdf file with:
    a. All the plots you produced.
    b. Organize your graphs, number them (Figure 1, Figure 2 etc) and give each figure a figure caption of one to two sentences, briefly describing one main feature of the variable demonstrated by each graph. If you believe that a required graph type is not appropriate for one or more of your variablea, comment on it and explain why you think so.
3) Team peer evaluation.

Part 5: Simple linear regression and final report. Assigned in week 12 and due in week 15.

Purpose: Affirm 1) competence in formulating meaningful hypothesis regarding the relationship between two variables; 2) confidence and technical competence in using data and the SLR model; 3) ability to interpret results and reason about the SLR model; 4) ability to clearly communicate information and support importance of work; 5) ability to collaborate with others, including motivation and ability to take initiative.

Guidance: Detailed instruction and guidance was provided for 1) running SLR in R and testing the model assumptions; 2) for paper organization and structure. Example papers from ASCE journals were posted on the online class space and student teams were asked to model their papers after these example papers. The final paper was specified to be no more than 3 pages long, single spaced, in 2-column format, 11-point Cambria font with headings identify each section. The paper was required to include the following: Abstract, Introduction, Data and Methods, Results and Discussion, Conclusion, References. Figures and Tables should be neatly numbered, formatted, carrying a meaningful caption, and described in text. A detailed description was provided to students for the expected content of each paper section.

Deliverables: Submit:
1) Your R code (follow our class file naming convention).
2) A .pdf file of your final paper.
3) Team peer evaluation.

Table 1. Example databases suggested to students, as sources of data for their project. Student choices were not restricted to these example sites.

Topic area	**Data source**
US Climate Resilience Toolkit	https://toolkit.climate.gov/#climate-explorer
US Government open data	https://www.data.gov
OECD data by topic or country	https://data.oecd.org
Data on sustainable development goals (you can register with a google account to explore the data)	https://data.world/worldbank/sustainable-development-goals
US EPA air quality data	https://www.epa.gov/outdoor-air-quality-data
CDC-National Environmental Public Health Tracking Network	https://ephtracking.cdc.gov/DataExplorer/#/

World Health Organization	https://www.who.int/gho/database/en/
US Energy Information Administration	https://www.eia.gov/tools/
US EPA water data	https://www.epa.gov/waterdata
NOAA-National Centers for Environmental Information	https://www.ncdc.noaa.gov/data-access

3. PROJECT ACTIVITY ASSESSMENT

Table 2 shows mapping of the goals to the project parts. Assessment of this first implementation of a data driven project in a CEE class is mostly qualitative and is based on 4 sets of available information:

1) Grading of submitted work for each project part
2) Student peer teammate evaluation, required to be submitted at the end of each project part
3) Student evaluation of student experience with the project, submitted at the end of the semester
4) Informal assessment from instructor interaction with the student teams.

Grading of submitted work

For project parts 1-4, students were graded based on 5 rubric criteria: 1) completeness; 2) understanding of (correctness) and consistency with guidelines; 3) organization; 4) neatness; and 5) clarity. For part 5, 11 evaluation criteria were used, shown in Table 3. The final project grade was based on grading of the final report. Grades in the formative parts 1-4 were used as a means of informing students of the quality of their submissions and alerting them in case of serious omissions or misinterpretations. Based on my experience, students might skip reading teacher feedback. However, teams always noticed a low grade and responded by requesting help.

Table 2. Mapping of the goals to the project parts.

Goal / Project part	Students stay involved	Confidence	Comprehension critical thinking	Coding skills	Written communication	Teamwork
Part 1: Choose a topic and find the data	√	√	√		√	√
Part 2: Refine the question	√	√	√		√	√
Part 3: Descriptive statistics	√	√	√	√		√
Part 4: Data visualization	√	√	√	√		√
Part 5: Simple linear regression and final report	√	√	√	√	√	√
Final paper assessment criteria						
Paper presented in specified format					√	
Paper contains all required sections					√	
Content of each section is consistent with purpose of section and has reasonable amount of detail			√		√	
Only essential summary statistics and graphs are displayed			√		√	
Calculations and results explained in context			√		√	
Figures and graphs are interpreted in the text			√		√	
Discussion included the results of a hypothesis tests, appropriate conclusions that can be made about the population of interest and limitations based on methods or data			√		√	
Statements and conclusions are valid and logical on the basis of evidence from the data and from consideration of natural, social or economic implications or interconnection of the issue examined			√		√	
Citations are referred in the text and are neatly and consistently formatted					√	
R script is organized in a logical order and is well documented with headers for every section and comments for every step				√		
R functions are used correctly; function options are exploited to produce clear and well labeled figures and correct and most informative output				√		
Other assessment tools						
Team peer evaluations	√					√
Project evaluation survey	√	√	√	√	√	√
Informal observation from student-instructor interaction	√	√	√	√	√	√

Table 3. Mean team grade performance in the 11 rubric criteria for the final project. N=27 teams. (points assigned to individual criteria according to the scale: *Some effort evident*: 25 pts; *Steps taken but not completed*: 35; *Good with space for improvement*: 45 pts; *Excellent*: 50 pts).

Final paper assessment criteria	Mean score ± Standard deviation / Total possible points
Total	**502.3±26.2 / 550.0**
Paper presented in specified format	49.1±2.0 / 50.0
Paper contains all required sections	50.0±0.0 / 50.0
Content of each section is consistent with purpose of section and has reasonable amount of detail	42.6±5.3 / 50.0
Only essential summary statistics and graphs are displayed	45.4±5.0 / 50.0
Calculations and results explained in context	46.7±4.8 / 50.0
Figures and graphs are interpreted in the text	44.6±5.2 / 50.0
Discussion included the results of a hypothesis tests, appropriate conclusions that can be made about the population of interest and limitations based on methods or data	44.3±4.9 / 50.0
Statements and conclusions are valid and logical on the basis of evidence from the data and from consideration of natural, social or economic implications or interconnection of the issue examined	43.0±6.1 / 50.0
Citations are referred in the text and are neatly and consistently formatted	43.2±5.7 / 50.0
R script is organized in a logical order and is well documented with headers for every section and comments for every step	48.9±3.2/ 50.0
R functions are used correctly; function options are exploited to produce clear and well labeled figures and correct and most informative output	49.8±1.0 / 50.0

Teamwork: peer evaluation

To encourage good teamwork and accountability, at the end of each project part, each student was required to submit a team peer evaluation. Students who were not submitting the form for more than one time had a grade penalty in their project grade. The team peer evaluation was based on criteria adapted from Comprehensive Assessment of Team Member Effectiveness (CATME) (www.catme.org) but using an instructor compiled google form. Students scoring less than 3 on a 1-5 Likert scale, in more than 2 criteria, by more than one of their teammates were tagged and notified. The ultimate penalty was zero project grade, for students who consistently scored below 2 in all project parts, as rated by more than one team member. In cases of teams with problems, I was discussing with the team in order to realize if the circumstances of the complaints were justified or were motivated by personal antagonism. There was one instance that personal conflict was the case and that was obvious by the fact that the two students were rating each other low while the other two teammates were rating them average or good. This conflict was settled after discussion. Two students had to be penalized based on consistent negligence noted by all their teammates. These students were also not present when I summoned the team to discuss their problems. In both cases the problematic students were doing well enough in the other areas of the class (homeworks and exams), to obtain a final passing grade. In the project student feedback survey, 7% of the responders identified teamwork as a challenge for them (Figure 4).

Student feedback about the project

Students were asked to respond to an anonymous google survey that included the following questions:

1) Has the project helped you understand how you can use data to answer important questions? (rating on a 1-5 Likert scale) (Figure 1).
2) Which part of the project best supported your learning about working with data? (multiple answers allowed) (Figure 2).
3) Which aspect of the project did you find the most rewarding? (briefly explain). (Figure 3)
4) Which aspect of the project did you find the most challenging? (briefly explain).
5) How could I improve the project as a learning experience?

Questions 1 and 2 targeted a quantification of student feeling about the project. Questions 3 to 5 invited student free responses as a preliminary questionnaire that will help build more specific surveys in subsequent years. Responses revealed a positive attitude toward the project by about 56% of the students, with the rest 28% rating it in the middle of the scale and 16% rating the experience low (Figure 1). Students appreciated the last two parts of the project that were application of the statistical methods covered in class (Figure 2). Use of R, compiling the final paper and producing graphs were the activities rated highest as rewarding (Figure 3), while finding data (Part 1 of the project), using R and interpreting results were identified as the most challenging aspects of the project (Figure 4).

In their comments for question 5, students proposed more time allocated for the project in class, to make it better integrated so that it does not feel as ancillary to the class. Similar to response for questions 3 and 4, there were mixed comments about the use of R with emphasis on more R practice sessions. There were no negative, rejective or angry responses, which further support an overall positive student attitude toward this type of class activity.

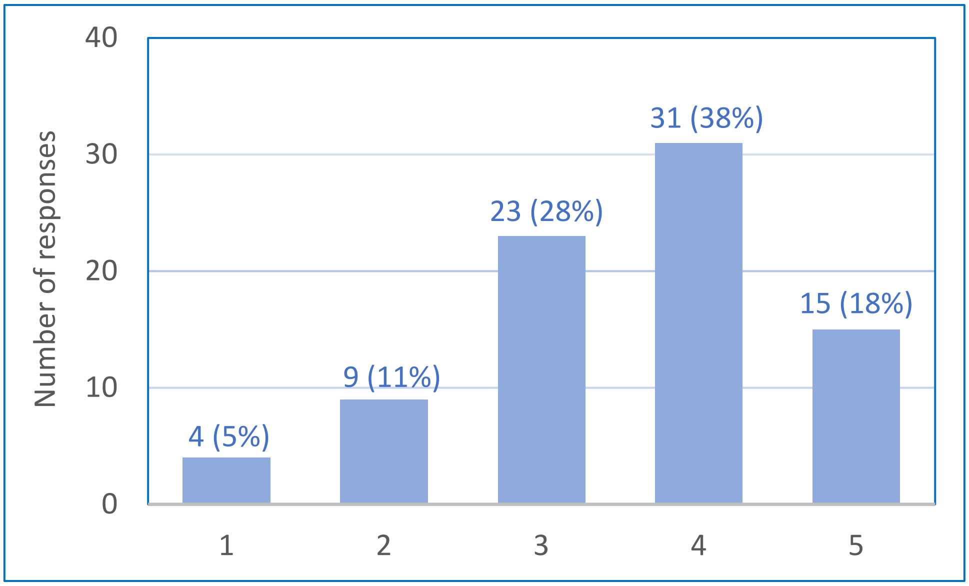

Figure 1. Responses to question 1: "Has the project helped you understand how you can use data to answer important questions?". 82 responses; (Scale: from 1 (No) to 5 (Yes, I fully appreciate why I had to go through the project process.)

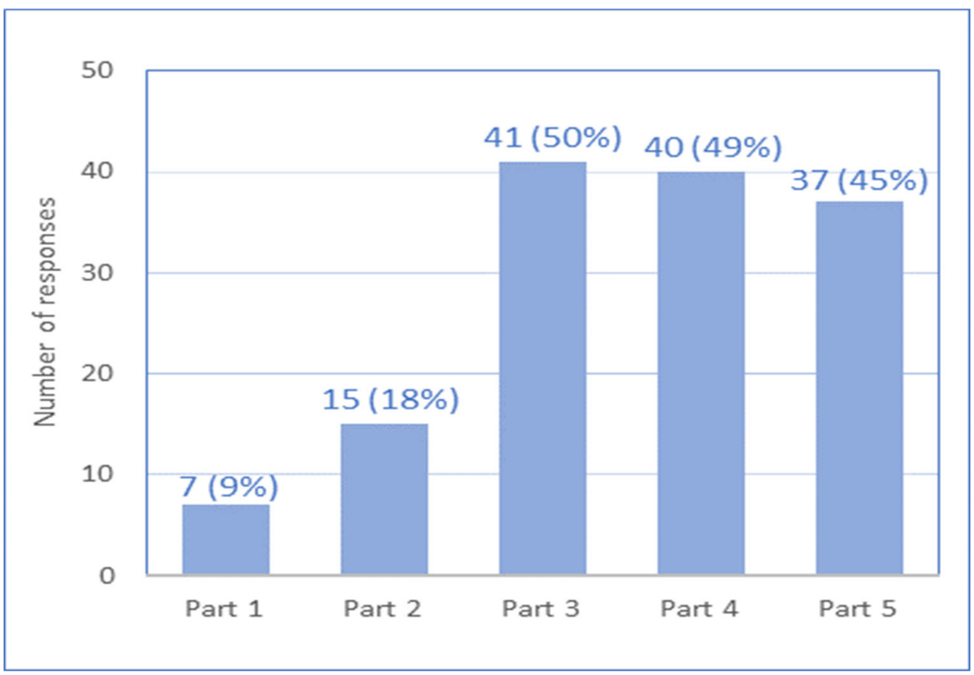

Figure 2. Responses to question 2: "Which part of the project best supported your learning about working with data?". 82 responses; This question allowed multiple answers and % percentage numbers, in the graph correspond independently to percentage of the 82 responders who chose a Part. (Part 1-Choose a topic and find the data; Part 2-Refine the question; Part 3-Descriptive statistics; Part 4-Data visualization; Part 5-Simple linear regression.)

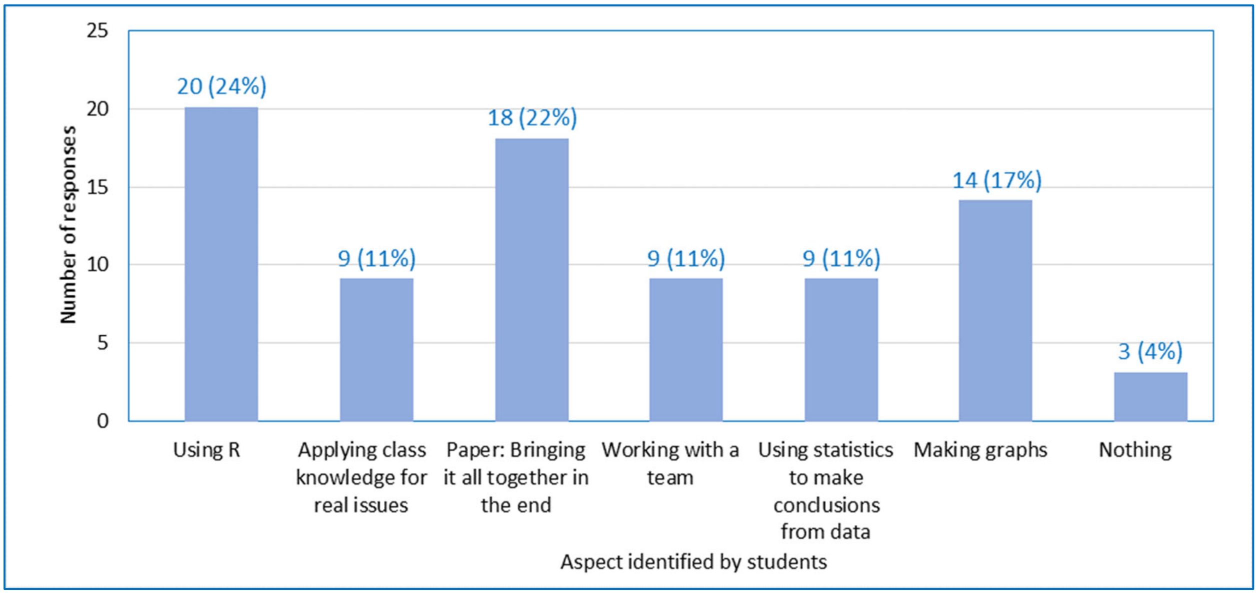

Figure 3. Responses to question 3: "Which aspect of the project did you find the most rewarding?". 82 responses. Identified aspects extracted from student free responses.

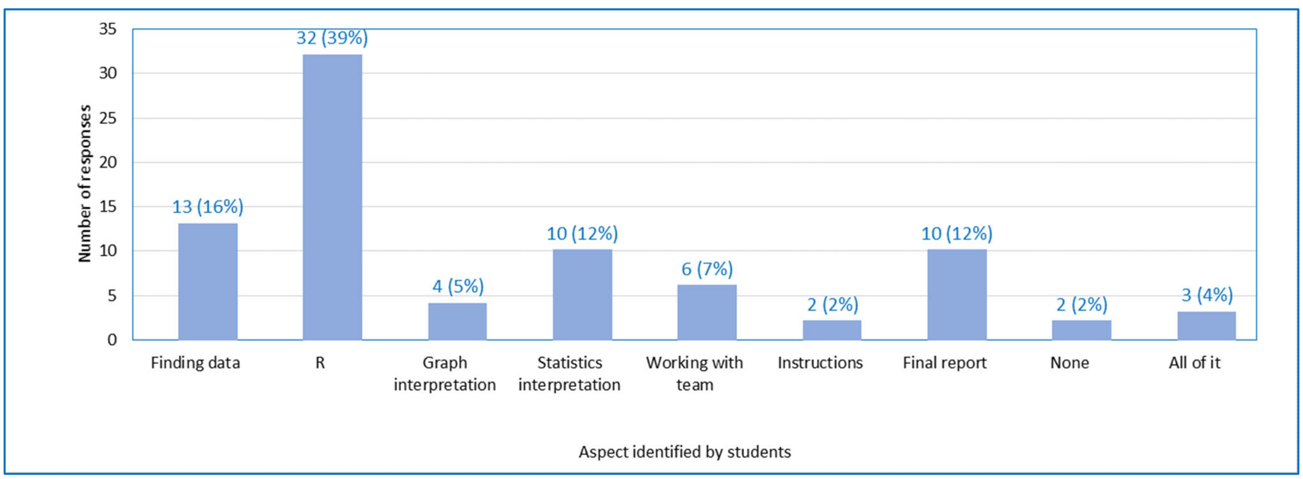

Figure 4. Responses to question 4: "Which aspect of the project did you find the most challenging?". 82 responses. Identified aspects extracted from student free responses.

Instructor assessment of activity from interaction with students

Here, I transfer informal experiential evidence from my interactions with project teams. It was clear that even though search for data by the students themselves has pedagogical benefits, it was the most challenging part of the project. In addition, students were particularly frustrated to discover that these real data had 'imperfections'. With the real data, the relationship between the dependent and independent variable was not always linear. Students were also frustrated to discover that the assumptions of the linear model were often violated. This gave opportunity for discussions about the limitations of the linear model and opportunity to test variable transformation.

From the instructor perspective based on graded work, peer teammate evaluations and informal assessment, the following areas of improvement were identified: 1) For this level of class, letting students choose topic and data while advantageous for been driven by student curiosity and interest, created challenges in terms of streamlining with the introductory material covered in this class; 2) A second issue apparent with about one third of the teams was lack of understanding of topic the team chose to investigate, combined with limited remedial capacity, due to lack of prior experience with researching a topic. This resulted in student teams not having basic contextual understanding of the issue they wanted to study, until late stages of the project, with some student teams trying to establish linear models between e.g., primary variables with non-linearly derived variables from these same primary variables; 3) Several teams started by splitting the work by project parts, instead of splitting tasks for each project part, as they were instructed to do. This was corrected every time it was detected, by requiring that all team members participated as a team to project consultations with the instructor.

4. CONCLUSION

In this paper, I described a preliminary assessment of a first attempt to integrate a project activity within an introductory Engineering Risk and Uncertainty class for Civil and Environmental Engineering students. Conclusions about achievement of the starting motivational goals can be summarized as follows:

Goal 1 - Students stay involved: From the team peer evaluations and participation of students in project related activities (project help hours, team meetings with instructor), this goal was achieved with only 2% students not been involved in the project.

Goal 2 - Confidence: This goal will need more targeted assessment in future implementations, as questions 1 and 2 of the project survey did not directly ask about personal confidence.

Goals 3, 4, 5 - Comprehension and critical thinking, *coding skills*, *written communication*: Even though final report quality and resulting grades provide positive evidence supporting achievement for all these three goals, responses to question 4, indicate students placed higher value in the technical skills they acquired through the project, and less on applying class knowledge to real issues.

Goal 6 - Teamwork: 7 % of students found teamwork challenging, with one instance that team needed intervention due to personal conflict. Better targeted survey will also help better assessment of this goal in future implementations.

This first implementation is going to be used for improvements in future classes in the following ways:

1. Providing specific datasets to work with
2. Integrating the computational environment (currently R) throughout the class, not only for the statistics part.

3. Allocating more class time dedicated to discussions about the relevance of the project to CEE applications and practice.

5. REFERENCES

[1] American Society of Civil Engineers, "Civil Engineering Body of Knowledge: Preparing the Future Civil Engineer, 3rd edition," Prepared by Civil Engineering Body of Knowledge 3 Task Committee, 2019. doi:10.1061/9780784415221.

[2] National Academies of Sciences, Engineering, and Medicine, "Environmental Engineering for the 21st Century: Addressing Grand Challenges". Washington, DC: The National Academies Press. 2019. https://doi.org/10.17226/25121.

[3] National Academies of Sciences, Engineering, and Medicine, "Data Science for Undergraduates: Opportunities and Options. Washington," The National Academies Press. 2018a.

[4] National Academies of Sciences, Engineering, and Medicine, "Assessing and Responding to the Growth of Computer Science Undergraduate Enrollments," Washington: The National Academies Press, 2018b.

[5] A. Godwin, "The Development of a Measure of Engineering Identity", American Society for Engineering Education, 123rd conference and exhibition, Paper ID #14814, New Orleans, LA, June 26-29, 2016.

[6] B. R. Belland, C. Kim, and M. J. Hannafin, "A Framework for Designing Scaffolds That Improve Motivation and Cognition," *Educational psychologist,* vol. 48, no. 4, pp. 243-270, 2013. doi:10.1080/00461520.2013.838920.

[7] R. D. Pea, "The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity", *The Journal of the Learning Sciences*, vol. 13, no. 3, pp. 423-451, 2004.

[8] R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/, 2013.

[9] P. C. Blumenfeld, E. Soloway, R. W. Marx, J. S. Krajcik, M. Guzdial, and A. Palincsar, "Motivating Project-Based Learning: Sustaining the Doing, Supporting the Learning," *Educational Psychologist,* vol. 26, no. 3-4, pp. 369-398, 1991/06/01 1991.

[10] E. H. Fini, F. Awadallah, M. M. Parast, and T. Abu-Lebdeh, "The impact of project-based learning on improving student learning outcomes of sustainability concepts in transportation engineering courses," *European Journal of Engineering Education,* vol. 43, no. 3, pp. 473-488, 2017.

[11] J. S. Vogler, P. Thompson, D. W. Davis, B. E. Mayfield, P. M. Finley, and D. Yasseri, "The hard work of soft skills: augmenting the project-based learning experience with interdisciplinary teamwork," *Instructional Science,* vol. 46, no. 3, pp. 457-488, 2017.

[12] M. E. Beier, M. H. Kim, A. Saterbak, V. Leautaud, S. Bishnoi, and J. M. Gilberto, "The effect of authentic project-based learning on attitudes and career aspirations in STEM," *Journal of Research in Science Teaching,* vol. 56, no. 1, pp. 3-23, 2018.

[13] N. Mvududu, "A Cross-cultural Study of the Connection Between Students' Attitudes Toward Statistics and the Use of Constructivist Strategies in the Course," *Journal of Statistics Education* [online], vol. 11 no. 3, 2003. Available at www.amstat.org/publications/jse/v11n3/mvududu.html. Last accessed Dec. 18, 2019

[14] J. Ledolter, “Projects in introductory statistics courses,” *The American Statistician*, vol. 49, no. 4, pp. 364-367, 1995.
[15] H. V. Roberts, "Student-Conducted Projects in Introductory Statistics Courses," in *Statistics for the Twenty-First Century*, eds. Florence Gordon and Sheldon Gordon, MAA Notes No. 26, Washington, DC: Mathematical Association of America, pp. 109-121, 1992.
[16] GAISE College Report ASA Revision Committee, “Guidelines for Assessment and Instruction in Statistics Education College Report 2016,” http://www.amstat.org/education/gaise, 2016.
[17] D. J. Spence, J. L. Sharp, and R. Sinn, “Investigation of Factors Mediating the Effectiveness of Authentic Projects in the Teaching of Elementary Statistics,” *Journal of Mathematical Behavior*, vol. 30, pp. 319-332, 2011.
[18] B. Bailey, D.J. Spence, and R. Sinn, “Implementation of discovery projects in statistics,” *Journal of Statistics Education*, vol. 21, no. 3, 2013, www.amstat.org/publications/jse/v21n3/bailey.pdf. Last accessed, Dec. 18, 2019.
[19] J. van de Pol, M. Volman, F. Oort, and J. Beishuizen, "The effects of scaffolding in the classroom: support contingency and student independent working time in relation to student achievement, task effort and appreciation of support," *Instructional Science*, vol. 43, no. 5, pp. 615-641, 2015. doi: 10.1007/s11251-015-9351-z.-z.
[20] The Right Question Institute, https://rightquestion.org/. Last accessed Dec. 19, 2019.