Gender Inequality in Computer Science: An Examination at the Undergraduate Level

By Paul Lucero, Suchita Lulla, and Panupoan Xiong

Abstract

As students at an engineering college, it is difficult not to observe the glaring gender gap that exists in science, technology, engineering, and math (STEM). Though many of the barriers that historically prevented women from entering male-dominated fields have been removed, the almost ubiquitous skew in the gender ratio in engineering and computer science classes toward men makes it clear that the work of fully incorporating women into these disciplines is not over.

In order to quantify the state of representation of women in engineering and computer science, we performed a study of the statistics published by the top 25 engineering universities (ranked according to the US News list of undergraduate institutions) as well as CU Boulder, the institution at which we are located. Trend-focused analysis of the data was then performed using the Python programming language for visualization and statistical inference.

We found that CU Boulder's gender ratio was consistently below the average gender ratio of the top 25 schools over the past 10 years. In addition, we found potential for a negative correlation between the rank of an engineering school and the proportion of women in engineering, but were unable to prove its statistical significance ($p >> 0.05$). Finally, we found that gender ratios in computer science tend to be more skewed toward men more than engineering overall; however, the computer science gender ratio has progressed toward 50% more rapidly than engineering overall in the past ten years.

Introduction

As society at large has progressed ever closer to gender equality, women are still woefully underrepresented in STEM fields. According to the 2011 census, women make up approximately 50 percent of the workforce; yet less than a quarter of STEM workers are women. Possible causes of this phenomenon currently being explored by diversity advocates are the minimal level of exposure to STEM and computer science in particular provided by the American public education system, the lack of role models in the STEM community to provide examples for potential female STEM workers, the continued perception of STEM fields as masculine, and the stigmatization of women in STEM fields as unfeminine.

Almost all disciplines within the umbrella of STEM are plagued by a lack of women. Introducing more women into STEM will tap the potential of a substantial subset of the population that has been thoroughly underrepresented. It will help women better themselves and allow them to help improve the lives of other women. Most importantly, women have vastly different experiences than men, and their perspectives are increasingly important in a world where science and technology affect the lives of a majority of people around the world.

In one study (Stout et al, 2011), when students were exposed to a same-sex expert in STEM, found that the students had an enhanced self-efficacy and enhanced commitment to pursue STEM careers. This study suggest that a lack of female role models may be hindering women from entering the field.

In another study by Heilbronner in 2012, “interest was cited as a major influence for occupational selection for both men and women.” In the American education system computer science is often not a required subject or it is not offered at all and so students entering college would have a very difficult time being interested in a topic they know nothing about . According to the same study, more women than men entered fields such as biology. In most educational systems, topics such as biology are a mandatory class that everyone takes, and will therefore gain at least a passing familiarity with before graduating. However, computer science classes are optional, higher level classes that a student may decide whether or not to take. This essentially means that students can be forced into making an uninformed or under-informed choice. Students that might thoroughly excel in STEM fields are thereby deprived of the opportunity to do, since they aren’t likely to be familiar with such fields, or make an informed decision not to enter these fields.

To date, most research conducted on this topic has been primarily focused on why women choose not to major in STEM which will no doubt be pivotal in solving this issue. However, research has yet to be conducted with regards to how different universities compare in their gender ratios, as well as how a university's prestige is likely to affect their gender ratio. With this study, we examine the top universities in the United States that offer engineering programs. The objective of doing this was to gain insight into these factors and draw conclusions about the trends in gender ratios of the university population, so that such data is available to complement similar data sets that pertain to the workforce.

Methods

The first step in performing this study was to select the universities that would be sampled from. The overall goal of this study was to determine if there was any sort of meaningful trend in the gender ratios of students in engineering, and more specifically computer science, as the reputation of the university increases. A major part of the study was thus sorting through prominent universities to determine which ones are the most reputable and which ones are most renowned in the field of engineering.

In the end, the US News ranking of undergraduate institutions’ engineering programs was selected for several reasons: easily accessible, easy to follow, and relatively predictable. For example, MIT and Stanford are near the top of the list, as are a large portion of well-known California engineering schools. A few other lists were consulted (the CollegeBoard list, as well as a few sites that are not particularly well-known or worthy of note,) but it was decided that the US News list was the best aggregate representation of the average perception of the reputation of universities in the United States. CU Boulder itself was also added to the list of researched universities as a point-of-reference school due to its locality and relevance to the location of the study being conducted.

As this study was a meta-study, most of the methods concerning the actual gathering of data revolved around unearthing previously performed research/surveys, aggregating the data from those surveys, and trying to isolate trends in the data. Thus, most of the time that was put into researching for this report was spent with writing Google searches and gathering links. This was also an iterative process, as the best way of finding the in-depth data required for this study (essentially, headcounts and/or gender ratios by major, which are not likely to be part of a university's primary fact sheet) was not necessarily a known quantity from the outset.

Queries started out as, for example, "<university name> gender ratios" or "<university name> data." However, such coarse queries are more likely to result in the university's pre-generated fact sheet, rather than any sort of meaningful data report. The data on such fact sheets is usually concerned with very broad statistics that will be popular with the general public and encourage enrollment. These statistics include the overall gender ratio in all programs at the college, total enrollment numbers, quantity of different programs offered, and the like. Unfortunately, none of these are specific enough to be useful to this particular study.

CU Boulder's own enrollment summary was helpful as a jumping-off point in this regard, as it gave way to search queries such as "<university name> data summary." It was also easier to find than many other universities' overall data summaries. This discovery led to the unearthing of in-depth fact sheets for several other universities, which in turn improved the formulated search queries that led to more data, and so forth. This process eventually led to the most useful search queries that the authors were able to come up with, such as "<university name> enrollment report," and "<university name> factbook", which yielded the best results overall. In some cases, a combination of queries or a custom Google search of the university's web site was necessary to overcome the difficulty of finding data.

One of the major issues with such a study is that there is no universal method of reporting in-depth statistics such as those which were necessary for this study, so each university's layout will be slightly different. While CollegeBoard specifies a "common data set" that many universities publish on their website, this data set is similar to many of the universities' generated fact sheets, which means that it is not terribly comprehensive or useful for research purposes. Thus, some universities have their useful data under the website for the Office of the Registrar; others, under the Office of Internal Reporting; others, the Office of Budget and Finance; and still others, hosted on an external data visualization website or cloud storage service, such as Tableau or Google Drive. There are also universities who do not make available the data that was necessary for such a study. Such a stratification of data storage and sorting is to be expected amongst major universities. However, it makes drawing any meaningful conclusions about the population of university attendees in North America as a whole significantly more difficult.

After gathering a variety of links, the meaningful data (in this case, the gender ratios) was brought into a Jupyter notebook (a web interface for the popular programming language Python) for analysis. A variety of Python modules were utilized for the storage, sorting, and exploratory analysis of data. Python was chosen for this purpose because it can essentially function as a lightweight, flexible framework for data analysis, as well as due to prior experience on the part of one of the authors with data analysis in this manner. This analysis consisted mostly of plotting the data and examining the resultant plots for any meaningful trends. Plots were made of the proportions of women in various schools with respect to time, the means of the data sets with respect to time, and the proportions of women with respect to the ranking of the school. A synthesis of the data, as well as a selection of the more interesting plots of the data, is provided below in the Results section.

Results

One of the most important parts of this study was seeing how CU Boulder itself stacks up against these larger and more well-known engineering schools. As a result, one of the comparisons that was made was the mean of the gender ratios in the top 25 engineering schools vs. the gender ratios at CU Boulder itself. While CU Boulder lags behind slightly in the case of engineering as a general field (by a difference of roughly 2 to 3 percentage points), it is on par with the rest of the engineering schools examined in this study for the more specific field of computer science. The plot of the gender ratios in engineering as a whole field is reproduced below as *Figure 1* so that the reader may see the difference. In all figures in this paper, a black line is drawn at the 50% mark as a point of reference; if a data point appears at this line, this indicates a 1:1 gender ratio in the relevant data set.

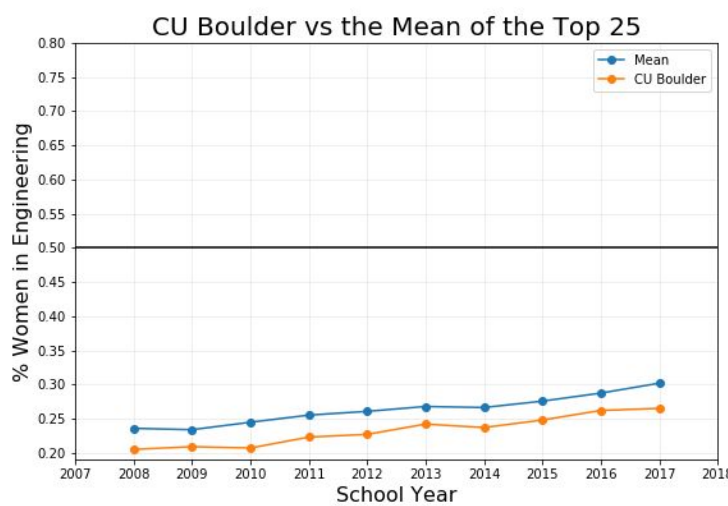
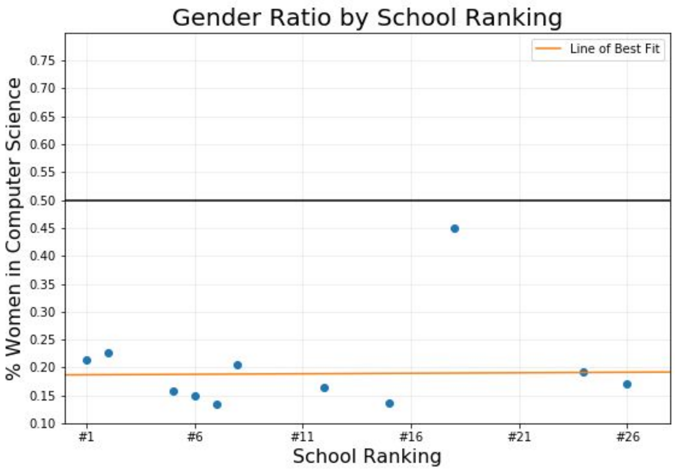
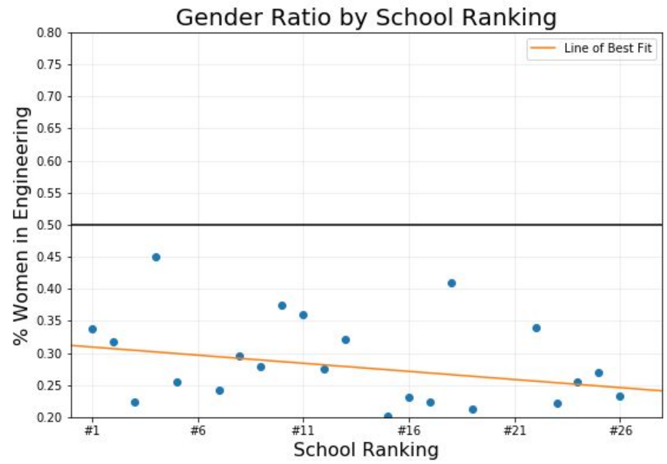


Figure 1: The mean of the proportion of women in engineering at the top 25 schools vs. at CU.

Another important trend that was examined was the trend of whether or not schools are more likely to have a higher proportion of women in engineering and/or CS if they are ranked better. A scatter plot was designed for this purpose, with the higher-ranked schools on the left and the lower-ranked schools on the right, along with a line of best fit drawn throughout the data. The slope of this line would then indicate some sort of trend, with a negative slope indicating a downward trend as rank decreases, and the inverse for a positive slope. The figures generated for this purpose are reproduced below as *Figure 2* and *Figure 3*. The line of best fit indicates a possible negative trend for engineering.





Figures 2 & 3: The trends of gender ratios with respect to school ranking, plus best-fit lines.

It is important to note that while the line of best fit indicates that there may be a negative trend in the engineering data, statistical analysis of the data does not indicate that the trend is meaningful or that it applies to the whole population. For the purpose of such analysis, a t-test was performed for the slope of the line to determine if it is likely that the true underlying system trends downwards. This t-test demonstrated that there is not likely to be a correlation ($p \gg 0.05$) in the overall population between the rank of the school and the proportion of female students, given current data.

It is also important to note the lack of data in Figure 2 specifically. Wherever data was unavailable for a given school, its respective point on the scatter plot was left out. This effect is especially well-illustrated in Figure 2, which contains only 11 of the desired 26 points of data. This lack of data is due to the fact that several universities do not make data on the headcounts of students by major readily available. While such data may be available by contacting f.e. the office of the registrar for a given university, it was not feasible to gather such data given the time scale, budget, and status of this research report.

The final trend that was examined was the difference between the means of proportions in engineering and computer science, which is reproduced below as *Figure 4*. It is interesting to note that while the mean proportion of women in computer science is lower than the mean proportion of women in engineering overall, both are trending upwards. In fact, the trend for the proportion of women in computer science appears to be trending upwards more quickly. This may indicate that the two are set to reach parity with each other in the foreseeable future. However, for the moment, there is a stark contrast between the means of 7 to 10 percentage points. In fact, when comparing the proportion of women in engineering to the proportion in computer science for a given university during a given school year, there were only 4 cases (out of roughly 200 valid data points) in which the latter proportion was larger. In addition, both means are still at least twenty percentage points below 50% in even the most recent data set, indicating a long ways to go before reaching the ideal 1:1 ratio.

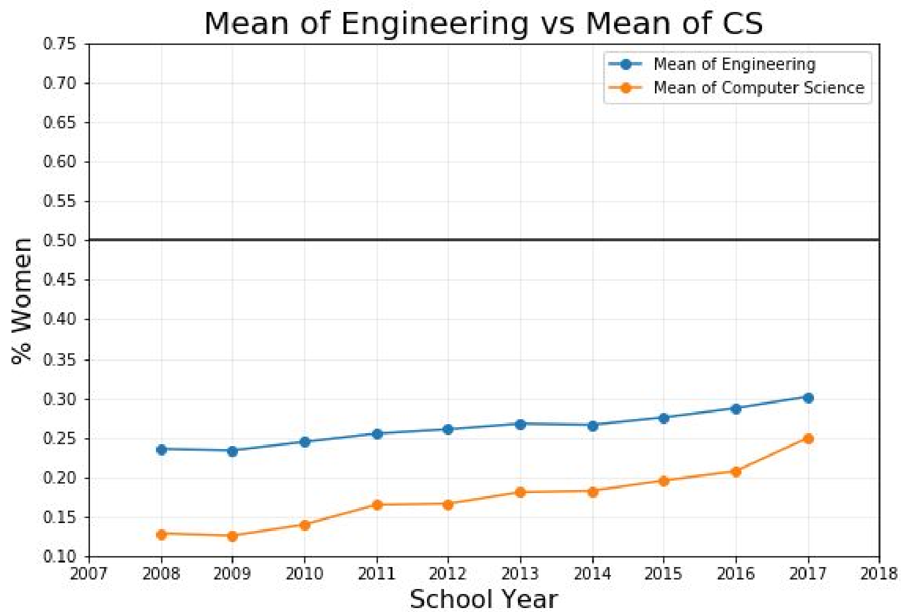


Figure 4: The means of gender proportions in both engineering as a general field and specifically Computer Science, across all schools sampled in the study.

Discussion and Future Directions

Our data shows that on average, the US News ranking of our sampled universities has a negative correlation with the percent of students in computer science and engineering overall who are women. If this trend is found to be statistically significant, it would raise a number of questions regarding barriers to entry for women into these fields.

Higher ranked universities tend to pull applicants from elite secondary schools that provide better academic support and programs for student enrichment. Could the cause for this trend lie not in the universities, but rather in secondary schools who do not place emphasis on supporting female students in science and math classes? In addition, elite universities are more likely to have affluent students. Could affluence be correlated with less entrenched gender roles? Students who live and go to school in less affluent areas may be less likely to have access to computer science and engineering programs, and therefore the potential for girls to be exposed to these concepts at a young age could be minimal.

We would like to expand this study to increase our sample size in order to verify the appearance of a downward trend in female to male ratios in computer science and engineering as school ranking decreases. The universities sampled are the most elite schools in the nation, and we would like to widen our sampling to institutions that are further down the US News list and are ranked around or below CU Boulder. This would allow for us to verify the validity of the trends that we have seen so far for the general population, as well as lay the groundwork for potentially examining the root causes of gender inequality at our own institution.

In addition, we would like the approach the sampled universities for their cooperation in obtaining comprehensive and up-to-date statistics. The publication of the relevant data is unstandardized, and often not easily accessible. Obtaining statistics from the universities directly would, in addition to providing a more consistent comparison, bring awareness to the fact that such information is not readily available. This would be the first step in a potential movement towards greater institutional accountability.

Upon gathering more comprehensive and diverse data, if the trend we observed in our preliminary study holds, then we would like to explore the driving causes behind the trend of lower ranked universities having poorer gender ratios in STEM. Possible causes for this may include less funding allocated toward diversity outreach and socioeconomic factors that may be correlated with institutional sexism.

If our observed trend holds, then the question that naturally follows is, what can we do on a societal scale to extend the benefits of social progress in STEM to those who are unlikely to have the opportunity to attend elite schools? As a society, what can we do to encourage women and girls toward fields where they are underrepresented? We hope that our study can be a starting point for a wider discussion about what society is doing to remove stereotypes and biases that have been held for generations about what types of people are suited for STEM studies and careers.

Bibliography

Stout, Jane G, et al. “STEMing the tide: Using ingroup experts to inoculate women's self-Concept in science, technology, engineering and mathematics (STEM).” *Journal of Personality and Social Psychology* , 2011, pp. 255–270.

Cheryan, Sapna, et al. “Do Female and Male Role Models Who Embody STEM Stereotypes Hinder Women’s Anticipated Success in STEM?” *Social Psychological and Personality Science*, vol. 2, no. 6, 2011, pp. 656–664., doi:10.1177/1948550611405218.

Diekman, Amanda B., et al. “Seeking Congruity Between Goals and Roles.” *Psychological Science*, vol. 21, no. 8, 2010, pp. 1051–1057., doi:10.1177/0956797610377342.

“Figure 2f from: Irimia R, Gottschling M (2016) Taxonomic revision of Rochefortia Sw. (Ehretiaceae, Boraginales). Biodiversity Data Journal 4: e7720. https://Doi.org/10.3897/BDJ.4.e7720.” doi:10.3897/bdj.4.e7720.figure2f.

Heilbronner, Nancy N. “The STEM Pathway for Women.” *Gifted Child Quarterly*, vol. 57, no. 1, Sept. 2012, pp. 39–55., doi:10.1177/0016986212460085.

Heilbronner, Nancy N. “The STEM Pathway for Women.” *Gifted Child Quarterly*, vol. 57, no. 1, Sept. 2012, pp. 39–55., doi:10.1177/0016986212460085.

Iyer, Lakshminarayan M., et al. “Polyvalent Proteins, a Pervasive Theme in the Intergenomic Biological Conflicts of Bacteriophages and Conjugative Elements.” *Journal of Bacteriology*, vol. 199, no. 15, 2017, doi:10.1128/jb.00245-17.

Klijn, J. “Ongoing 2b/3a inhibition In Myocardial infarction Evaluation.” *Http://Isrctn.org/*>, May 2012, doi:10.1186/isrctn06195297.

Prescott, David S., et al. “Introduction.” *Feedback-Informed treatment in clinical practice: Reaching for excellence.*, pp. 3–9., doi:10.1037/0000039-001.

Ramsey, Laura R., et al. “The effects of an academic environment intervention on science identification among women in STEM.” *Social Psychology of Education*, vol. 16, no. 3, 2013, pp. 377–397., doi:10.1007/s11218-013-9218-6.

Smith, Jessi L, et al. “When Trying Hard Isn't Natural.” *Sage journals*, 27 Nov. 2012.

Enrollment Statistics: MIT Office of the Registrar, web.mit.edu/registrar/stats/yrpts/index.html.

Spotlight on Data | Institutional Reporting, Research, and Information Systems (IRRIS) | The University of Texas at Austin, reports.utexas.edu/spotlight-data.

Office of Budget and Planning - University of California, Santa Barbara, bap.ucsb.edu/institutional.research/planning.data.book/tables/index.cfm.

Enrollment Report - FALL SEMESTER 2008, registrar.vanderbilt.edu/docs/tenth08f.html#page1.

Berkeley, UC. *OPA – University of California Berkeley*, pages.github.berkeley.edu/OPA/our-berkeley/student-enrollments.html.

“Enrollment.” *UCLA Academic Planning and Budget*, www.apb.ucla.edu/campus-statistics/enrollment.

“Enrollment.” *Enrollment - Institutional Planning and Research - University of Florida*, ir.aa.ufl.edu/uffacts/enrollment-1/.

“Enrollment.” *Enrollment | Institutional Assessment & Studies, U.Va.*, ias.virginia.edu/university-stats-facts/enrollment.

“Enrollment in Undergraduate Programs.” *Penn State College of Engineering*, www.engr.psu.edu/facts/undergrad-enrollment.aspx.

“Enrollment Reports.” *Enrollment Reports - Office of the Registrar*, ro.umich.edu/enrollment/enrollment.php?report=enrollmentcounts.

“Factbook – Enrollment.” *Cornell Institutional Research and Planning*, irp.dpb.cornell.edu/tableau_visual/factbook-enrollment.

“Fall Enrollment 2017-18.” *Fall Enrollment 2017-18 - Caltech Office of the Registrar*, www.registrar.caltech.edu/academics/enrollment.

“Fall enrollment headcounts.” *University of California*, 16 May 2017, www.universityofcalifornia.edu/infocenter/fall-enrollment-headcounts.

“Georgia Institute of Technology Institutional Research and Planning.” *Enrollment | Institutional Research and Planning*, irp.gatech.edu/enrollment.

“Loading...” *Data Digest*, www.purdue.edu/datadigest/.

“Mission, Facts, and Stats.” *UW College of Engineering*, 8 Nov. 2017, www.engr.washington.edu/about/mission.

“Preliminary Enrollment.” *Data and Research Services - Student Data and Reports*, dars.tamu.edu/Student/Preliminary-Enrollment-(1).

“School of Engineering Enrollment 2016-17.” *School of Engineering Enrollment 2016-17 | Registrar's Office*, registrar.stanford.edu/everyone/enrollment-statistics/enrollment-statistics-2016-17/school-engineering-enrollment-2016-17.

“The Numbers.” *Harvard John A. Paulson School of Engineering and Applied Sciences*, 11 Oct. 2012, www.seas.harvard.edu/about-seas/quick-facts/numbers.

“UIUC Student Enrollment.” *Student Enrollment*, www.dmi.illinois.edu/stuenr/#race.

“UIUC Student Enrollment.” *Student Enrollment*, www.dmi.illinois.edu/stuenr/#race.

University, Carnegie Mellon. “Institutional Research and Analysis.” *Factbook 2017-2018, Vol. 32 - Institutional Research and Analysis - Carnegie Mellon University*, www.cmu.edu/ira/factbook/facts2018.html.

"University of Colorado--Boulder | Best Engineering School | US News." *U.S. News & World Report*, U.S. News & World Report, www.usnews.com/best-graduate-schools/top-engineering-schools/university-of-colorado-boulder-02032.

" Office of Institutional Research." *First-Time Undergraduates : Rice University Office of Institutional Research*, oir.rice.edu/Factbook/Students/Admission/Undergraduate/First-Time_Undergraduates/.