Student and Instructor Gender Division by Area of Study in Upper Division Computer Science Technical Electives at University of Illinois Urbana-Champaign, 2011-2021

By Hannela Jane

Final Report for DREU Summer Internship 2021

Mentor:

Colleen Lewis, University of Illinois Urbana-Champaign

Introduction

Computer Science (CS) as a field suffers from known gender disparities on all fronts, academic and otherwise. To combat inequality and increase female representation and participation, it's important to research the specifics of gender inequality in hopes of eventually changing the culture and framework of CS as a whole to create a fully representative community.

In this report, I analyze my research into gender representation in specific areas of upper division CS electives at the University of Illinois Urbana-Champaign (UIUC). I look at female representation of both students and instructors across eight subfields, and how female representation of instructors relates to that of students.

I found in my research that the subfields of Big Data, Media, and Human Impact have a higher percent of both female students and female instructors compared to other subfields, and that there is a significant correlation between the proportion of female instructors and that of female students in each subfield.

- **RQ1:** How does female enrollment in upper division CS electives compare to total enrollment by concentration at UIUC between 2011 and 2021?
- **RQ2:** How does the percentage of female instructors differ by concentration at UIUC between 2011 and 2021?
- **RQ3:** What is the relationship between female enrollment and female instructors by concentration at UIUC between 2011 and 2021?

Previous Research

Across industry and academia, institutions are trying to increase the flagging representation of women in Computer Science. General outreach and recruitment are decent short term fixes to Computer Science's diversity problem, but long term retention can only be achieved through a cultural shift—and the minutia of that necessary shift must be informed by the minutia of gender disparity in CS. To that end, I wanted to look at gender representation among CS students under a metaphorical microscope, to look at students' preferences for different areas of upper division CS study based on their own gender as well as the gender of their instructor. Gendered enrollment patterns will help inform future efforts to diversify the field, and assist educators in supporting female CS students.

Looking into gendered patterns in specific areas of study can also inform future research into salaries in different CS subfields. Specialization in school influences later salary, and gender disparities in CS subfields could exacerbate the wage gap.

In 2016, Wade Fagen and Cinda Heeren, UIUC professors, and Corly Leung, a UIUC CS undergraduate student [1], looked at data on CS classes at UIUC from 2011-2015 with the purpose of looking for patterns in gender representation among students and faculty, and whether those patterns informed the success of female students. They found some trends that suggested that female students have fairly consistent preferences in the specific areas of CS they decide to study.

At UIUC, studies have been done on gender disparities and biases in CS, their effects, and how to address them [1, 2, 3], but more work needs to be done. I want to take Fagen's research a step further by collecting data from the past decade from multiple top CS universities. At the time of this report, I've only received data from UIUC, but even so it is twice as much information as Wade et al. used. I hope that a specific analysis of disparities within concentrations can contribute to the efforts to improve UIUC's CS community, and those of CS Departments across the country.

Dataset and Methods

University of Illinois Urbana-Champaign (UIUC) Division of Management Information. The Division of Management Information provided a list of every upper division elective offered in the CS department from Fall 2011 to Spring 2021. Upper division is defined based on the UIUC CS department's definition—as courses with numbers above 400, plus CS 397 which is Individual Study and counts as an upper division elective according to the CS Department's major requirements.

Dataset includes, per course:

- Semester/Year
- Course number
- Course title
- Concentration(s) satisfied
- Instructor name
- Instructor gender (M/F)
- Total enrollment
- Number of female-identified students
- Number of male-identified students
- Number of students of unspecified gender

I categorized each course into one or more of the 8 concentrations offered by UIUC—Software Foundations, Models of Computation, Big Data, Human Impact, Media, High Performance Computing, Distributed Systems, and Machines. All UIUC CS

students are required to pick a concentration and take at least 3 of their 8 required technical electives from that area [4]. To categorize the courses, I removed all courses that either didn't fall under a concentration or were too vague to categorize, then categorized them based on UIUC's CS major requirements. Specifically, I removed Individual Study, Seminars, Thesis Research, CS Team Projects, Senior Thesis, and Senior Projects. I also removed all Special Topics classes, because although those courses often satisfy concentration requirements, the course titles aren't listed in the data, so there is no way to tell which concentration a Special Topics course satisfies in any given semester.

The eight concentrations defined by UIUC are a good way of breaking up CS courses into categories, which can give insight into specific gender distributions within CS. Fagen et al. [1] use the same categories in their analysis of CS gender distributions at UIUC. I believe the concentrations are more widely applicable categories that can be used in analysis of the diversity of CS Departments across universities as well as analysis of industry jobs for women in CS. If the categories can be applied to industry jobs, as I believe they can, then the salaries of each category can be compared to gender distributions of undergraduate CS students, providing valuable insight about equity in the CS community.

Furthermore, the concentrations can be used to look at the gender distribution of instructors, and how it relates to the gender distribution of students, which will pave the way for further CS Education (CSEd) research into how instructor's gender affects students, and which specific areas inspire/encourage women to go into research and why.

Results

> RQ1: Female enrollment in upper division CS electives compared to total enrollment by concentration at UIUC between 2011 and 2021

Women are an underrepresented group in the field of computer science, and UIUC's CS department is no different—although women make up more than half of the world's population and between 45% and 47% of UIUC's student body over the last two decades [5], they make up only around 25% of degrees granted in CS at UIUC. Given this disparity, more detailed research is necessary to investigate the disparities within concentrations—and how they can be overcome.

Figure 1 is a bar chart showing the total enrollment by gender in each concentration, displaying the notable gender disparity in orange (males) and blue (females), and in what areas that disparity is most extreme. Figure 2 shows the percentage of female students taking classes in each category. Media, Human Impact, and Big Data have the highest relative female enrollment of any category over the past 10 years, while Machines and Models of Computation have the lowest.

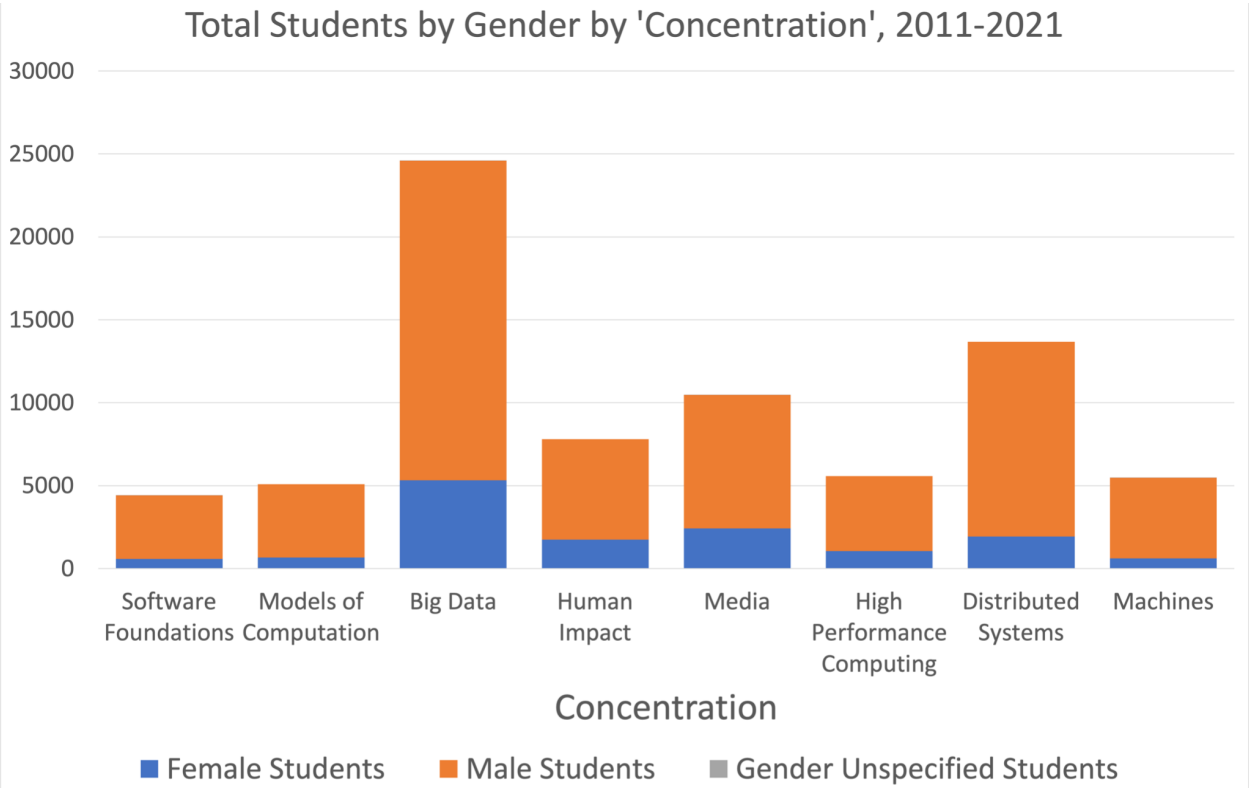

Figure 1: The total number of students by gender from Fall 2011 to Spring 2021 who enrolled in classes that fulfill requirements in each concentration (there are too few students in the Unspecified Gender group to show up visibly on the graph).

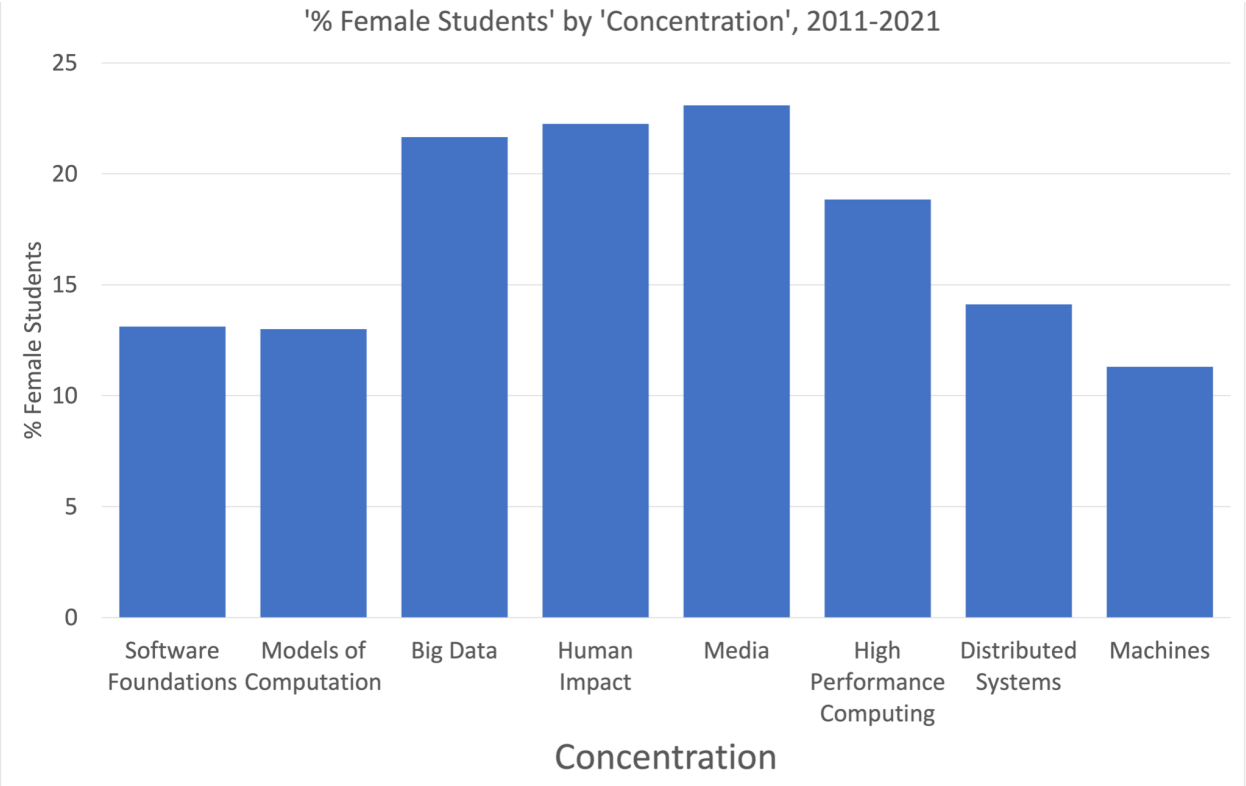

Figure 2: The percent of female enrollment in each category, out of total enrollment from Fall 2011 to Spring 2021.

> RQ2: How does the percentage of female instructors differ by concentration at UIUC between 2011 and 2021?

The percentage of female instructors reflects the gender disparities in CS as well, and shows some dramatic differences in female instruction in the different concentrations. Figure 3 shows the percentage of female instructors in each category. Big Data, Media, and Human Impact have the highest percentage of female instructors, while High Performance Computing and Software Foundations have the lowest. High Performance Computing is an outlier, with only 1 female instructor over the last 10 years.

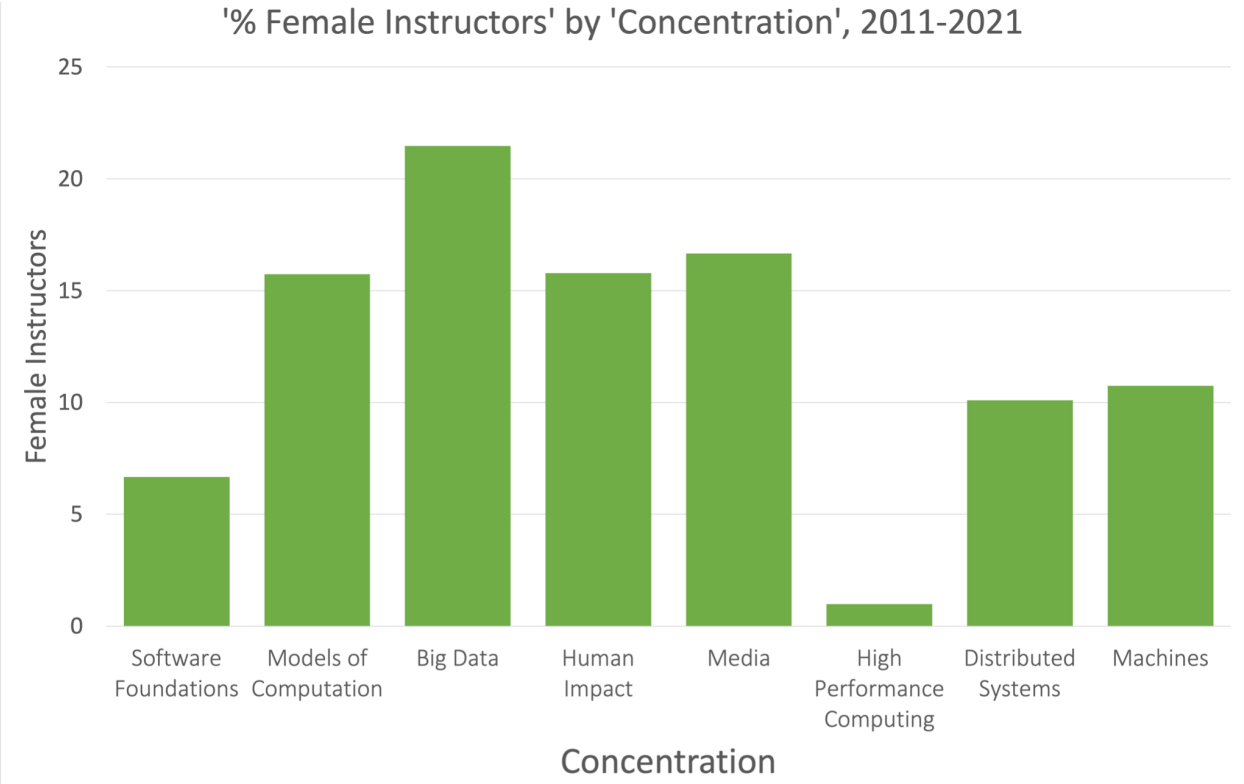

Figure 3: The percent of female instructors in each concentration over the past 10 years.

> RQ3: What is the relationship between female enrollment and female instructors by concentration at UIUC between 2011 and 2021?

Seeing representation in mentors and teachers increases interest in a particular subject area can offer students important support [3], perhaps explaining the correlation between female instructors and female students in the different categories. The correlation likely also relates to the graduate school pipeline; more female students in a given category means more potential future instructors in that category.

Figure 4 is a bar chart showing percentages of female representation in instructors and students in the different concentrations. With the exception of Models of Computation, every other concentration has a higher percentage of female students than female faculty, perhaps suggesting a disparity in female CS representation in graduate school. The top three concentrations for percentage of female instructors are the same as the top three for female students—Media, Human Impact, and Big Data.

Figure 5 plots the relationship between female instructors and female student enrollment in each concentration. I chose to exclude the High Performance Computing (HPC) concentration from the graph because it is an extreme outlier, with only 1 female

instructor in the last decade. The correlation coefficient is .728, showing a very strong positive correlation between the percentage of female instructors and the percentage of female students in each concentration except HPC. Including HPC, the correlation coefficient is .414, which still indicates a strong positive relationship. It is unclear whether a high percentage of female instructors in one area encourages a higher percentage of female students in that area, a higher percentage of female students in one are predicts a higher percentage of female graduate students who go on to become professors in that area, women have increased interest in certain areas for other reasons (perceived social impact, etc), a combination of all three, or something else entirely drives this correlation.

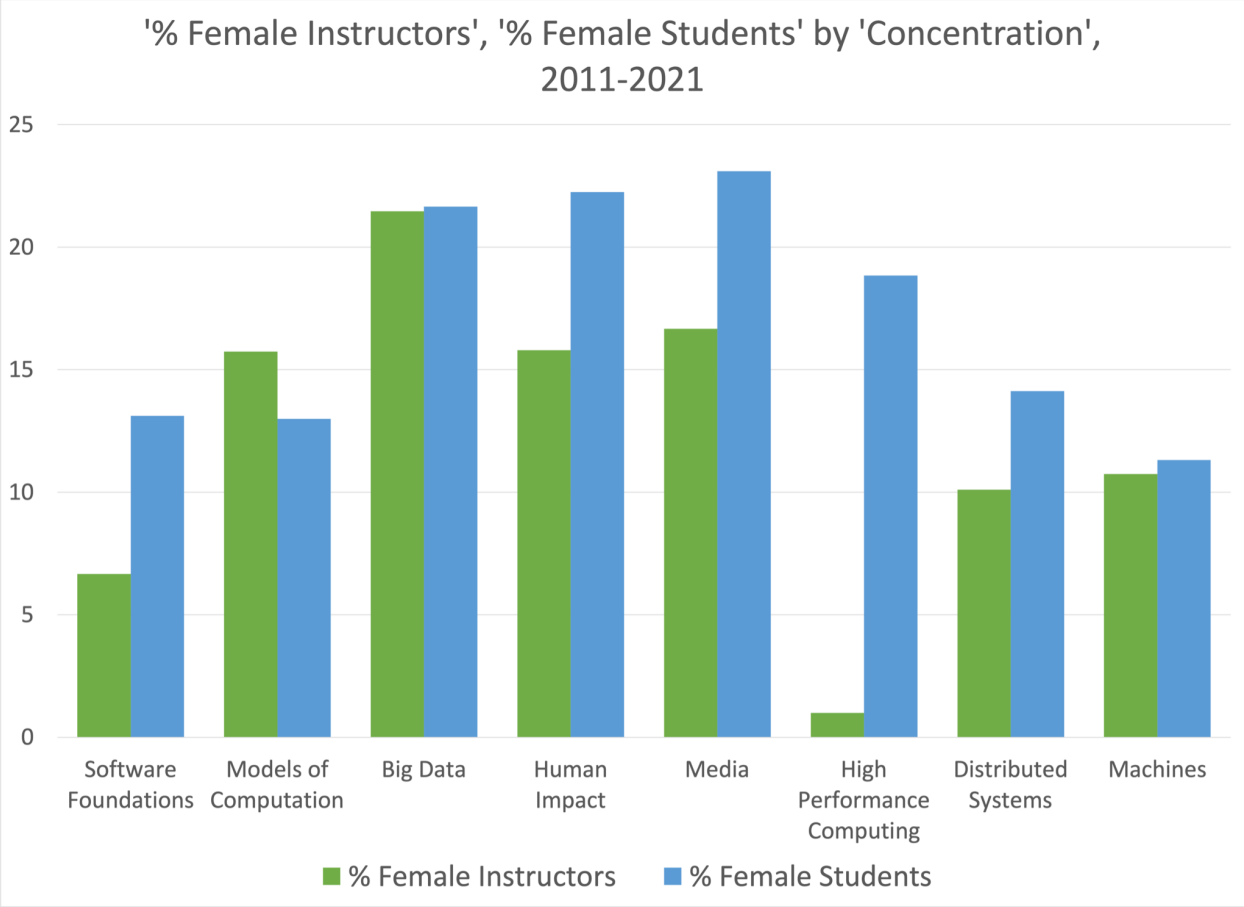

Figure 4: Comparison of the data from Figures 2 and 3, showing the percentages of female students and instructors in each concentration over the last 10 years side by side.

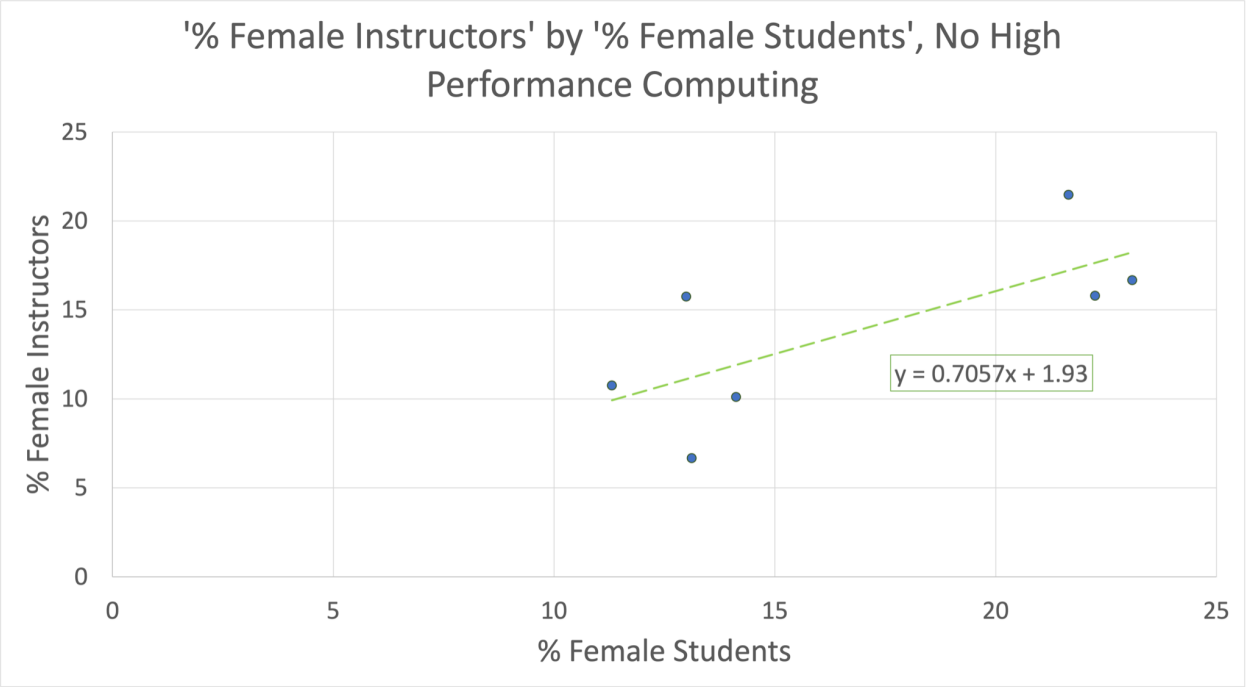

Figure 5: The relationship between the percentage of female students and instructors in each concentration, excluding High Performance Computing (HPC).

Considerations and Takeaways

I think it's worth noting that there are probably students listed as male or female that do not actually identify as such; in all likelihood a third option was only recently introduced. The first "Gender Unspecified" student appears in Fall 2018.

The data doesn't include information about what concentration each student has actually picked, only enrollment per class, meaning that many students are certainly counted multiple times per semester. This could skew the information based on how many upper division CS electives students choose to take, and how much they choose to concentrate or branch out. Since a concentration is defined as a category in which students take at least 3 out of 8 technical electives, there can be students who specialize by taking all 8 electives in one category, and students who prefer to branch out and take only 3 electives in one category and the remaining 5 in five different categories. It's very possible that female students tend towards one extreme or the other, which will skew the data. If female students tend to specialize, they will be over represented in their preferred categories, but if they tend to branch out the trend in preferred concentration will be harder to track. It is also possible that female students tend to either branch out and take non-CS courses with any extra time, or tend to take extra CS electives. Because

of this, I look at the data as representing areas of general interest rather than actual concentrations.

The data I used also doesn't include Special Topics courses, because I wasn't provided with the names of the courses or what concentrations they satisfied. This could significantly change the data, because Special Topics courses often satisfy concentrations, but I think the data still provides a pretty good picture of the gender distribution landscape, since Special Topics courses make up a relatively small fraction of the upper division technical electives offered.

I think the next step, aside from continuing to collect data from other universities, is to look at the implications of the fields where women are most and least represented. Are the fields with the most women (Big Data, Human Impact, and Media) generally higher or lower paying? Do they tend towards research and grad school or towards industry jobs? Undergrads should be interviewed about why they've chosen their concentration—do they perceive their area of study as being more helpful/impactful to society?—and female undergraduates should be asked how the gender of their instructor influences their experience of a class/subject. This further research can help inform teaching strategies, outreach, and other equity efforts within the CSEd community.

Bibliography

[1] Fagen, Heeren, Leung. (2016). Gender in CS at UIUC. Senior Thesis of Corly Leung. https://courses.engr.illinois.edu/cs199205/sp2016/discovery/gender_in_cs_at_illinois/

[2] Tomkin, J. H., West, M., & Herman, G. L. (2018). An improved grade point average, with applications to CS undergraduate education analytics. ACM Transactions on Computing Education (TOCE), 18(4), 1-16.

[3] Crenshaw, T. L., Chambers, E. W., Heeren, C., & Metcalf, H. E. (2017). Ten years toward equity: Preliminary results from a follow-up case study of academic computing culture. *Frontiers in Psychology*, *8*, 816. https://doi.org/10.3389/fpsyg.2017.00816

[4] UIUC Grainger Engineering Office of Marketing and Communications. (n.d.). B.S. in Computer Science Major Requirements. https://cs.illinois.edu/academics/undergraduate/degree-program-options/bs-computer-science

[5] UIUC Division of Management Information. (n.d.). UIUC Student Enrollment. Retrieved from https://www.dmi.illinois.edu/stuenr/