Diversity in Computer Science

Michael McRoskey - mmcrosk1@nd.edu

Friday March 18th, 2016

Overview

Having raw data on the ethnic makeup of computer science students at Notre Dame was really helpful because it allowed me to visualize how the campus climate is changing (or not) with respect to gender and ethnicity. To process this data, I used awk to:

- Create column arrays of the different years

- Use the size of those arrays to calculate class size

- Manually print calculated percentages for each group

I used `data.sh` for the awk commands and `plot.plt` to generate line graphs representing the percentage change over time.

Methodology

Processing Demographics

`data.sh` pipes `demographics.csv` into the awk command and counts the data in each column with the format `s2013[$1]++` where `s` represents sex and `r` represents race/ethnicity for each year.

The `END` block adds the M and F numbers to get a total class size for each year. I had to make sure that blank parts of the CSV file evaluated as `false`. Finally, I printed results in a terribly inefficient yet effective manner:

```
# 2013 Results
print "2013,", s2013["M"]*100/tot2013, ",", s2013["F"]*100/tot2013, ",", r2013["C"]*100/tot2013, ",", r2
```

All of this printing gets redirected to `data.csv`, which is used in `plot.plt`.

I think the most frustrating part was figuring out how to restructure the data. Because I was dealing with so many years, I wanted to nest my statements in for loops going from 2013 to 2018. However, I also wanted to utilize some of the abstraction of associative arrays. Ultimately, my solution wasn't pretty and required a lot of copy and paste, but it worked.

GNU Plot

`plot.plt` formats GNU plots `race.png` and `sex.png` by outputting a PNG line graph with the proper labels. I had to make sure that the percentage values in `data.csv` were multiplied by 100 so that the plot could handle the axes appropriately.
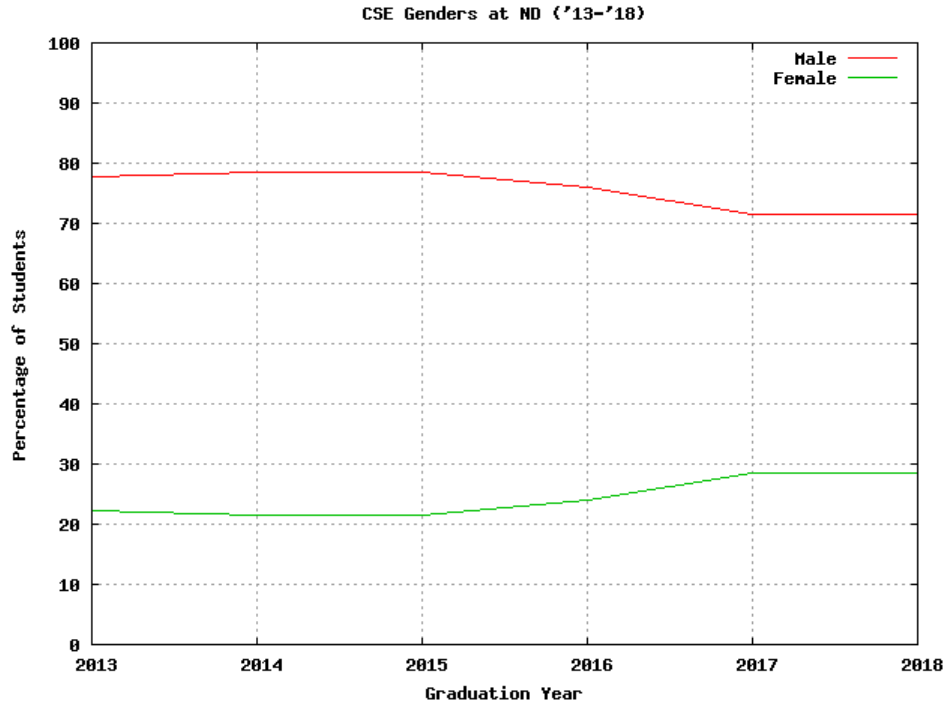
Figure 1: Gender Disparity in Notre Dame Computer Science

Analysis

I was surprised at the statistical disparity between men and women students in Notre Dame Computer Science. The lines were not very close to an equitable 50% and are fairly flat. However, there was a slight trend toward convergence in recent years. It was a good decision to use a percentage because that reflects proportion rather than pure numbers, which can be difficult to judge.

Figure 2 shows a no definable trend in recent years but a significant proportion of Caucasian students in Notre Dame Computer Science. Although there isn't a great line of best fit for other ethnicities, it's for certain that they remain well under 20%, even 10%.

The results of the outputted `data.csv` file:

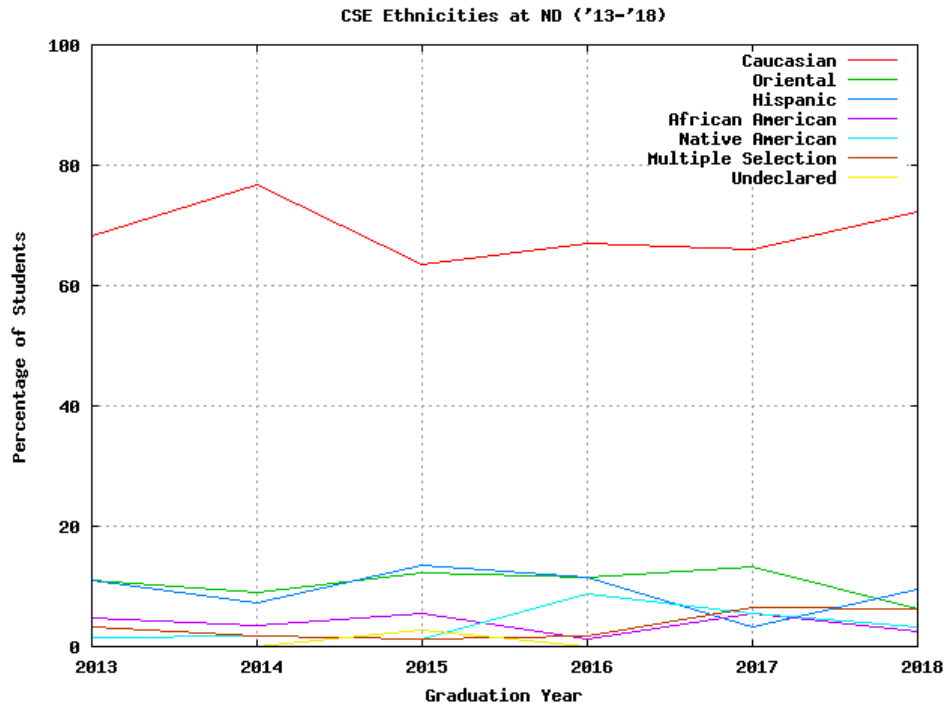| Year	Male	Female	Caucasian	Oriental	Hispanic	Black	Native-Am	Mult.	Undeclared
2013	78	22	68	11	11	5	2	3	0
2014	79	21	77	9	7	4	2	2	0
2015	78	22	64	12	14	5	1	1	3
2016	75	25	67	11	11	1	9	2	0
2017	71	29	66	13	3	5	5	7	0
2018	71	29	72	6	10	2	3	6	0

Table 1: Percentages of Groups

Figure 2: Ethnic Diversity Disparities

Discussion

Notre Dame's Ethnic Makeup

I think it's really important to pay attention to our ethnic makeup. I saw "Decoding the Gender Gap" this year, a really compelling documentary that explores these demographic disparities in computer science. I understand the need to promote diversity so that there can be a broader range of ideas for a broad range of humanity, especially in a profession where lines of code affect millions of users. I think it's also important to understand Notre Dame's core ethnic makeup, since we're predominantly Caucasian. I would love to find out statistics from our institution on ethnicity overall, because I think it will likely be the same type of results: vast disparities. That would lead me to conclude that this is not necessarily a departmental concern, but a University-wide one, although departmental efforts to broaden exposure never hurts.

Ways to Improve Notre Dame's CSE Diversity

Computer Science seems to be one of the fastest growing majors across the country. I think we need to tap into that and lure more talented students into the program, which could certainly help our diversity. I think there should be an Intro to Computer Science Course as offered at some of our peer institutions open to all first year students across all majors.

Although I have never felt unwelcome, intimidated, or outnumbered in computer science, I understand students who could feel this way. I think it's important for us all to "be in it together" and I appreciate the effect group projects have on this front. I've always felt a sense of belonging and want to make sure other students feel this way too.