Final Visualization Report

INFX 561

Stephanie Ballard, Namrata Deshpande & Shane Miller

Data Source

For our final visualization report, we used data collected by the US Census Bureau through the American Community Survey (ACS). The ACS was developed to collect demographic data as a series of small monthly samples used to create annual estimates. This data was previously collected by the US Census Bureau as part of the decennial census long form. Beginning in 2010, the ACS replaced the decennial census long form.

The data set comes from the ACS table titled: " Occupation by Sex and Median Earnings in the Past 12 Months (In 2005 Inflation-Adjusted Dollars) for the Civilian Employed Population 16 Years and Over", which can be accessed through the link listed in the references. The data set has one table for each year 2005-2014 that includes values for twelve variables for each detailed occupation. For this visualization report, we focused on one occupation subset titled "Computer, Engineering, and Science occupations". This subset included three more specific occupation categories: "Computer and Mathematical Occupations", "Architecture and Engineering occupations" and "Life, Physical, and Social Science Occupations". Each occupation had a value for twelve variables, but we focused on the following three:

Variable	Data Type
Total Estimate	Discrete
Male Estimate (as a percentage of the total)	Continuous
Female Estimate (as a percentage of the total)	Continuous

We used the female and male estimates given as a percentage of the total estimate to calculate a discrete value for the number of male and female employees in each occupation group.

Limitations of the Data Set

This data set has a few limitations we would like to call attention to:

First, the ACS only provides for respondents to claim one sex out of two options (male and female). This does not accurately reflect the variety of gender presentations possible in the workplace, and creates tension for people who do not identify as male or female.

Second, prior to 2010 the three categories “Computer and Mathematical Occupations”, “Architecture and Engineering occupations” and “Life, Physical, and Social Science Occupations” were calculated as part of a larger group of “Professional Occupations”. To correct for this, we calculated the missing occupation category “Computer, Engineering, and Science occupations” for the years 2005-2009.

Third, the subcategory “Life, Physical, and Social Science Occupations” seems to present a gender balanced subcategory, but there has traditionally been an imbalance between male and female employees in the life and physical sciences (predominately male) and the social sciences (predominately female). We were not able to further divide this sub-category using the data available, but we think it is possible that this grouping of the data may present an inaccurate perception of gender balance.

Additionally, the lack of workplace diversity in STEM occupations is a rich and complex issue. Attempting to represent this issue fully through one data set is not meant to be a comprehensive overview of the status of the STEM industry, but to bring attention to one of the many ways the problem of diversity in STEM occupations can be represented through data visualizations.

References

American Community Survey: http://www.census.gov/programs-surveys/acs/
Data Table:
http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_05_EST_S2401&prodType=table

Target Audience

Organizations and the people working in STEM fields

Our first target audience is STEM employers, managers, recruiters, and HR personnel. This might include the public or private organizations working and employing people, mainly in STEM fields. This is irrespective of the size of the organization or the number of people currently employed at the organization. By organizations, we mostly mean the people in these organizations who play an important role in impacting the issue, like those who are in charge of recruitment, such as hiring managers or team lead, and those responsible for supervision and evaluation of employees such as project managers.

Reasons for inclusion and expectations

According to the employment projections by the Bureau of Labor Statistics, STEM industry is one of the fields who will show a rapid growth in employment generating more than 1 million jobs in total till 2024. If the decline in the number of women employed in STEM continues, these industries might face serious issues in filling these jobs and also in removing huge gender disparity. The organizations and people who are a part of this field are the ones who could and should understand the seriousness of this issue and the factors causing it. By using this visualization they could understand the chronic nature of the issue and some potential steps that they could take to remedy this scenario.

Assumptions made:

· The people and the organization as a whole, are interested in diversity inclusion and removing the gender disparity in their industry.

· They can understand the terms used in the visualization like performance evaluation and also the categories and subcategories of the occupations.

Organizations working for women empowerment and diversity inclusion:

This target audience might include the nonprofit, government or private organizations who are working for women empowerment and to remove the disparity in the workplace. Some examples might include National Center for Women & Information Technology (NCWIT) and the Office of Global Women's Issues (US Department of State).

Reasons for inclusion and expectations

These organizations play an important part in spreading awareness about major issues faced by women along with encouraging people to fight against such issues. Using the visualization, they can understand where we currently stand with respect to this issue in the STEM industry and some potential solutions that may reduce this disparity.

Assumptions made:

· These organizations are working against gender disparity and are interested in understanding the issues in STEM industry that are causing this disparity.

· They can encourage people to be aware of the issue and to work against the factors causing it.

Students enrolled or planning to enroll in STEM fields

This group includes people currently studying in STEM fields who are likely to join the STEM industry after graduation. Hence, it is important for them to be aware about the current scenario of such industries.

Reasons for inclusion and expectations

As mentioned earlier, the students who are planning to join the STEM industry in the near future and those who are planning to enter the field for their career, should be aware about the current statistics about the people who are currently employed in the industry. Using this visualization they can understand the proportion of people currently employed in different STEM fields and some major issues that they could face. The visualization could also act as motivation for female students to pursue STEM careers to remove this disparity.

Assumptions made:

· These students are interested in STEM fields, the current statistics of the people working in this industry and the major issues they may face in the STEM industry

· They could understand the distribution of the occupations and the terminology used.

Tasks to be accomplished

The major tasks to be accomplished by the visualization and the questions that should be answered include:

- To visualize the number of people employed in the different STEM occupations and their distribution between male and female workers over a period of nine years (2005 to 2014) to answer questions like:
    - For a given year what is the approximate number of male and female workers employed in all the Computer, Science and Engineering occupations?
    - For a given year what is the distribution of these workers among the three subcategories of the occupations, namely, Life, Physical and Social Sciences, Architecture & Engineering and Computer & Mathematics?
    - Is there an increase or decrease in the number of male or female workers employed in any of the field over time?
    - What percentage of women is employed in different STEM occupations and whether that percentage is increasing or declining with time?
    - What is the data pattern for women over a period of nine years and whether it is similar to that of men or not?
    - How big is the gap between the number of male and female employees in the STEM occupations?

- To understand some of the steps that could be taken to overcome this disparity and to encourage women to apply for STEM jobs.

Iterations

Sketch A: Stacked Area Graphs

DESCRIPTION OF THE GRAPH:

The first iteration of the visualization that was generated while conducting exploratory analysis of the data, makes use of stacked area representation. Four graphs were generated, one encompassing all the occupations under Computer, Science and Engineering field, while the other three represent the three sub-fields namely Architecture & Engineering, Computer & Mathematics, and Life, Physical & Social Sciences. In the visualization, in order to visualize the change in number of people over time, the x-axis was assigned to the dimension time (years ranging from 2005 to 2014) and the y-axis was assigned to the count of people employed in the respective field. Preemptive visual encoding of colors was used to represent the sex of people.

REASONS FOR THE ENCODINGS:

Area distribution was used to denote the number of people employed in the fields because by visualizing the area the user could get understanding of the density of people along with the distribution between the sexes. As the areas are stacked over each other, comparing the difference in number of male vs number of female is easy. The choice of colors for the visualization was based on the popularity of colors generally used for representing the two sexes.

TOOLS USED:

Tableau

VISUALIZATION:

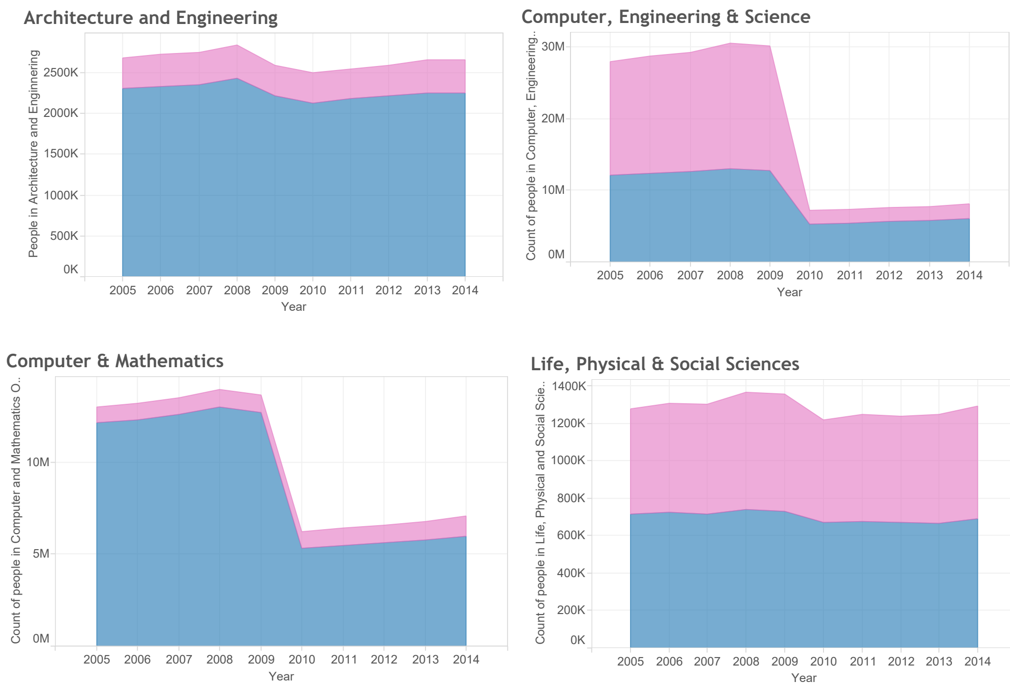


Figure 1: Stacked Area Visualization

Sketch B: Double line graphs

DESCRIPTION OF THE GRAPH:

The second sketch is the set of line graphs. As with the first sketch, four graphs were generated (one that includes all computer, science and engineering occupations, one for Architecture & Engineering, one for Computer & Mathematics, and one for Life, Physical & Social Sciences). Each graph has one line representing the number of males employed in the occupation and one line representing the number of females employed in the occupation. The x-axis represents the years in the data set (ranging from 2005-2014) and the y axis is the total number of employees. One advantage to this visualization is that the lines for each of the two sexes represented are the actual total of employees (as opposed to the cumulative nature of the stacked area graph and stacked bar charts). This reduces the potential for misinterpreting the data.

REASONS FOR THE ENCODINGS:

In this sketch, lines were used to illustrate the change in data over time. We chose to visualize the data set using the line graph because it is a fairly common type of visualization and is relatively easy to interpret. In this visualization, we diverge from the typical colors used to represent sexes to challenge gender stereotyping in visual representations.

TOOLS USED:

R and Illustrator

VISUALIZATION:

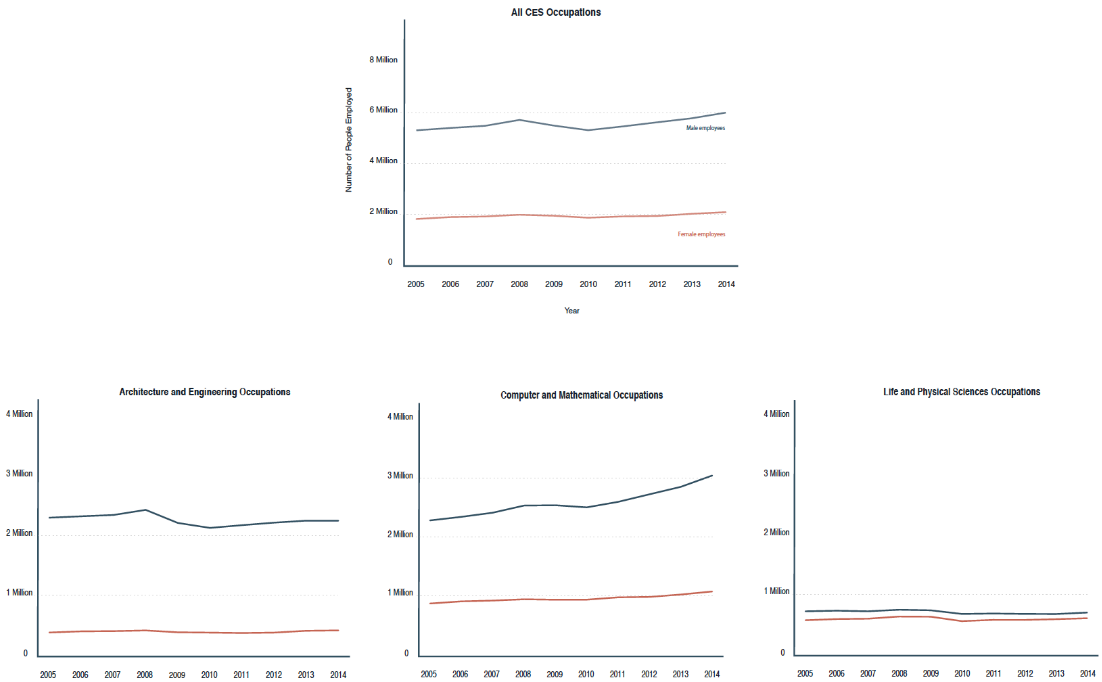


Figure 2: Line graphs

Sketch C: Stacked Bar charts

DESCRIPTION OF THE GRAPH:

The third sketch is the set of stacked bar charts. As with the previous sketches, four graphs were generated to represent the four categories in the data set. Each graph has one bar for each year along the x-axis. The bars are divided by sex of the employee. The x-axis represents the years in the data set (ranging from 2005-2014) and the y axis is the total number of employees. Like the stacked area graph, each bar is the cumulative total of both sexes employed in each year.

REASONS FOR THE ENCODINGS:

We chose to visualize the data set using the stacked bar graph because it is clearly represents the change over time and displays the number of employees of each sex as a proportion of the total. As with the second sketch, we diverge from the typical colors used to represent sexes to challenge gender stereotyping in visual representations.

TOOLS USED:

R and Illustrator

VISUALIZATION:

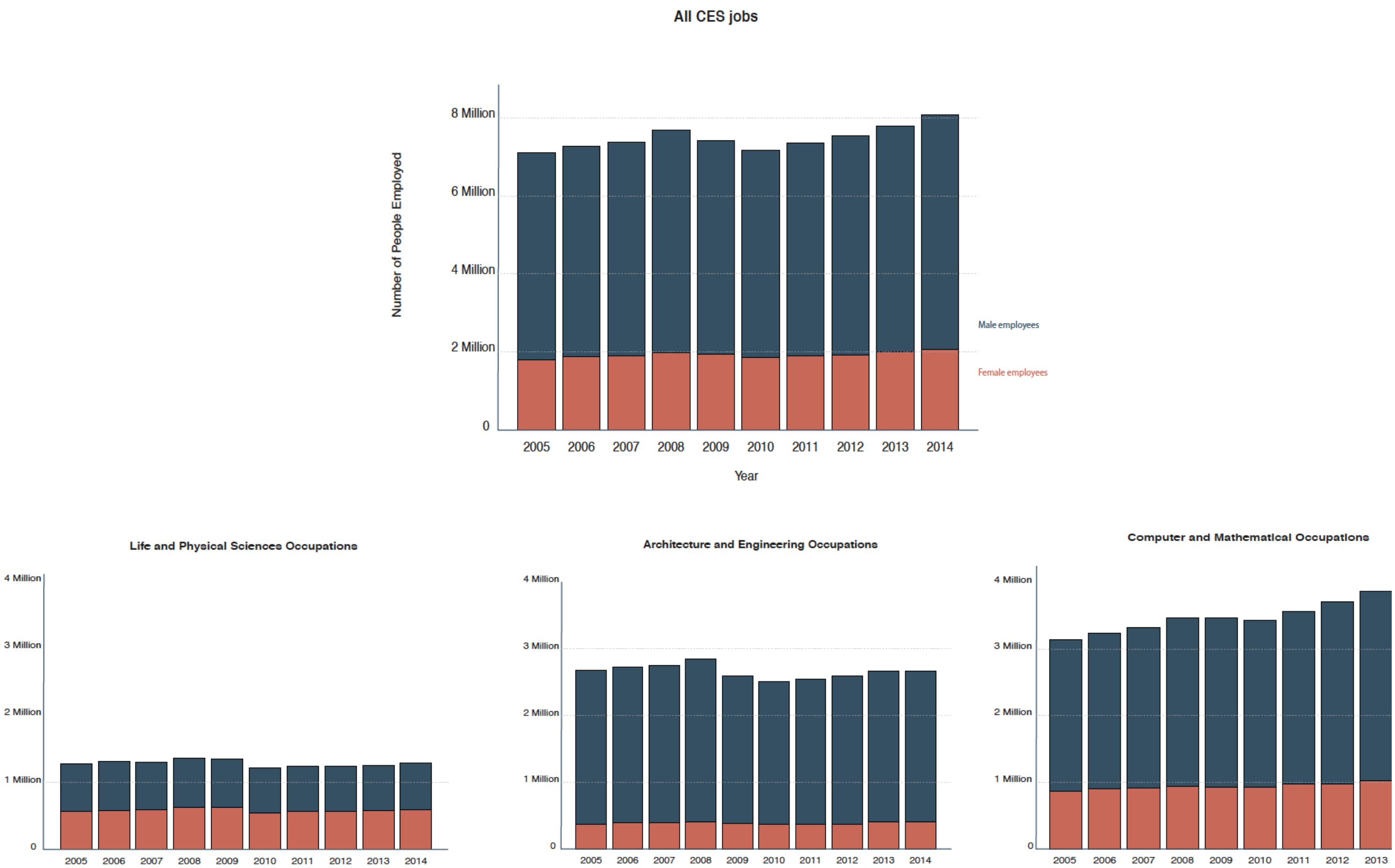


Figure 3: Stacked Bar Charts

User Study Details

Protocol

For each participant, we showed them one graph at a time between our line graph and stacked bar graph. We presented each graph to the test participant and asked us to explain their thought process as they read and analyzed the graphs. After a minute or two, we asked them three questions the graph they were viewing:

1. On a scale of 1-5, how easy is it to understand the scales on the graph?
2. On a scale of 1-5, how easy is it to recognize any data patterns in the graph?
3. On a scale of 1-5, how effective is the graph in presenting its data?

After completing this process for both graphs, we then asked the test participant the following questions:

1. What purpose do you think these graphs would be used for?
2. Which graph was your favorite of the two? Why?
3. Was there an aspect from either graph that stood out as being very helpful? If so, what was it?
4. Was there an aspect from either graph that stood out as being very distracting or confusing? If so, what was it?
5. Do you have any other comments or things you would like to point out?

Participants

PARTICIPANTS	AGE	SEX	SIMILARITIES WITH TARGET AUDIENCE	PREFERED VIS
Participant 1	29	Male	Employed for the past five years as Technology Analyst (one of the STEM field showing significant decline in women employees) in IT industry and hence have an understanding of the working environment and changing statistics in the industry.	Line
Participant 2	35	Female	Employed as Project Manager in an IT industry and supervises more than one technical team. As a woman at leadership position in a technical firm,	Line

			she has the exposure to the current scenario of issue and has the authority to make an impact.	
Participant 3	20	Female	A current undergraduate student at Portland State University studying Communications. Interested in Spanish and Culture and aspires to work for a study abroad program.	Line
Participant 4	22	Male	A current undergraduate student at the University of Washington studying Engineering. Currently works as a tutor in Math and Physics for kids with Dyslexia.	Stacked Bar
Participant 5	30	Male	Previously employed as a recruiter in the IT industry. This participant has an understanding of HR and company needs for diversity in hiring.	Line
Participant 6	30	Female	Previously employed as a recruiter, though not in the tech industry. This participant has insight into general employment trends in various industries and an understanding of how sex discrimination affects the workplace.	Line

Results

The results were similar across all tests, only having a few discrepancies. Regarding scales, the majority of participants preferred the stacked bar graph over the line graph due to the horizontal scale lines across the stacked bar graph that made reading individual data points easier. Regarding data pattern recognition and overall effectiveness though, the line graph performed significantly better. The pattern was able to be seen better because “the line graph directly contrasts how guys have higher rates than girls, whereas the stacked bar graph focuses more on the total amount of jobs - not the difference between male and females.” Another common suggestion was to reorganize how the graphs are presented on the screen (some were confused about how the three lower graphs were subsections of the first graph). Some asked for more information about the data such as the data collection region, citation/source, adding a legend for the line graph, and defining what “CSE” meant. A visual suggestion was also made that the line graph has a lot of white space, almost too much.

Actions Taken

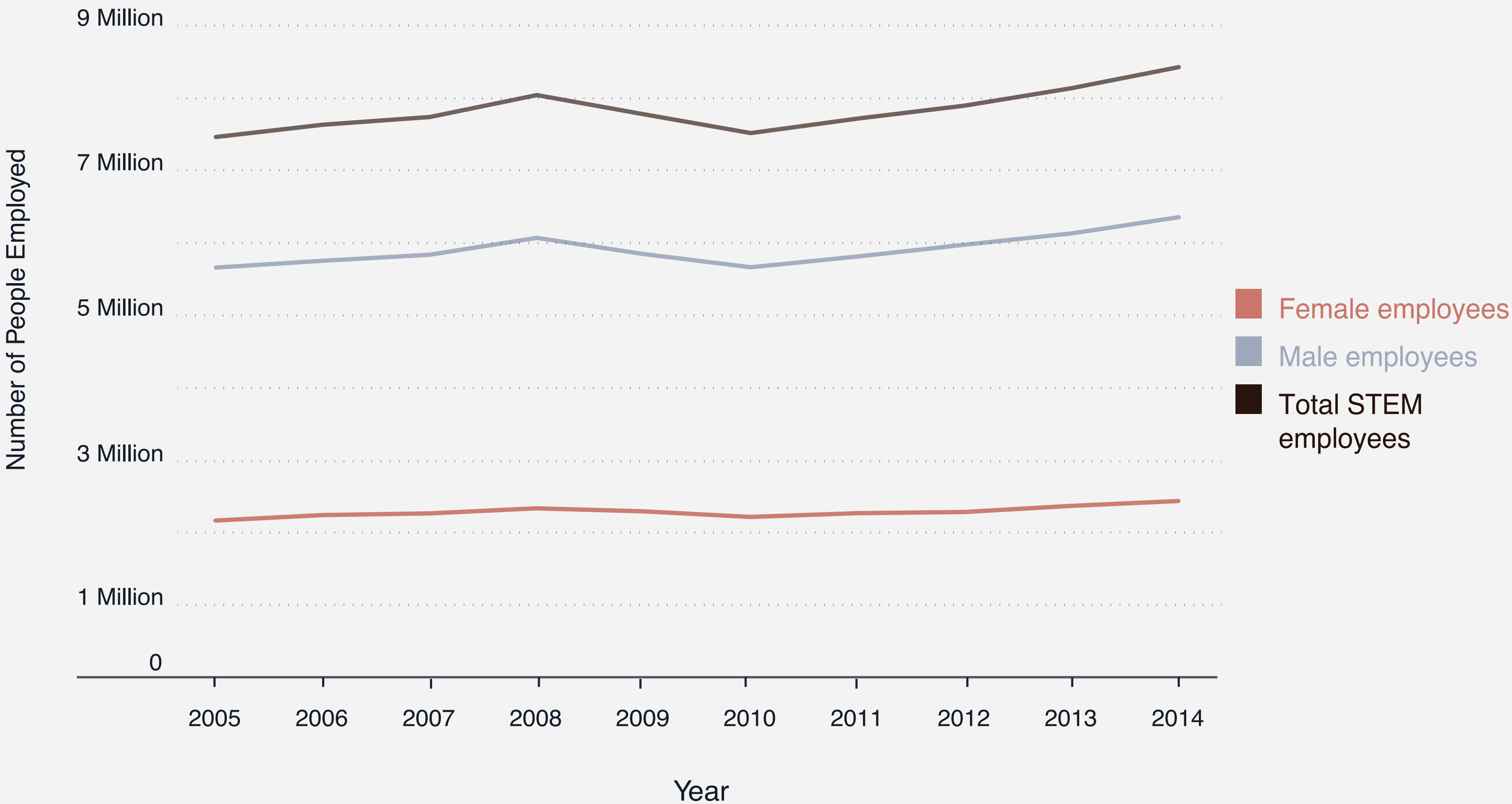
We decided to move forward with the line graph and iterate upon it based upon its favorable user feedback. We implemented the following modifications:

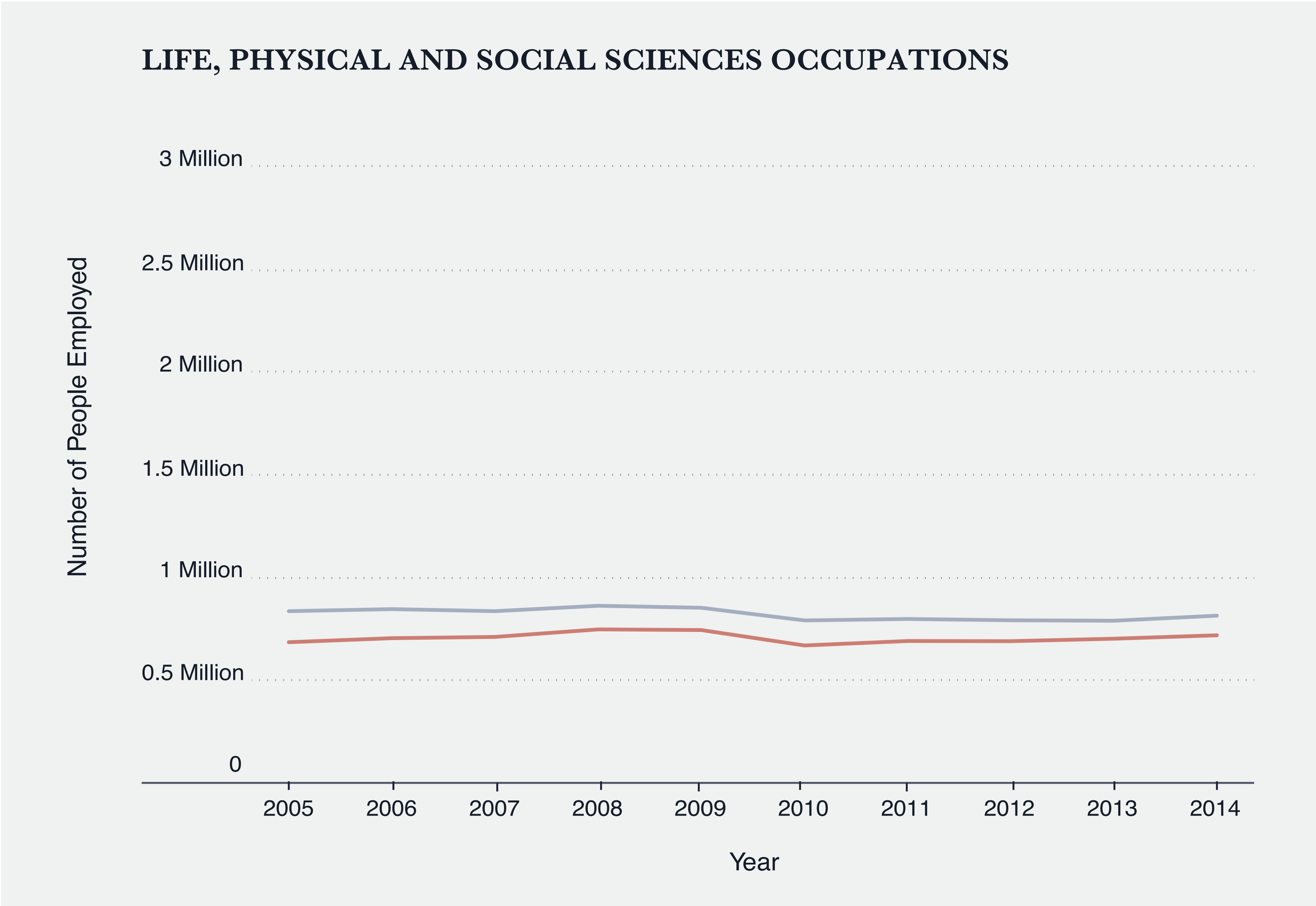
- Place the Industry Overall graph on a separate line
- Specify the geographic area (mention they are about the US)
- Add citation/source
- Add horizontal scale marks to the line graph
- Spell out “Computer Science & Engineering” instead of saying “CSE”
- Slightly zoom in the scale
- Thicken the graph’s data lines
- Add a legend for what the data lines represent

Final Visualizations (following pages)

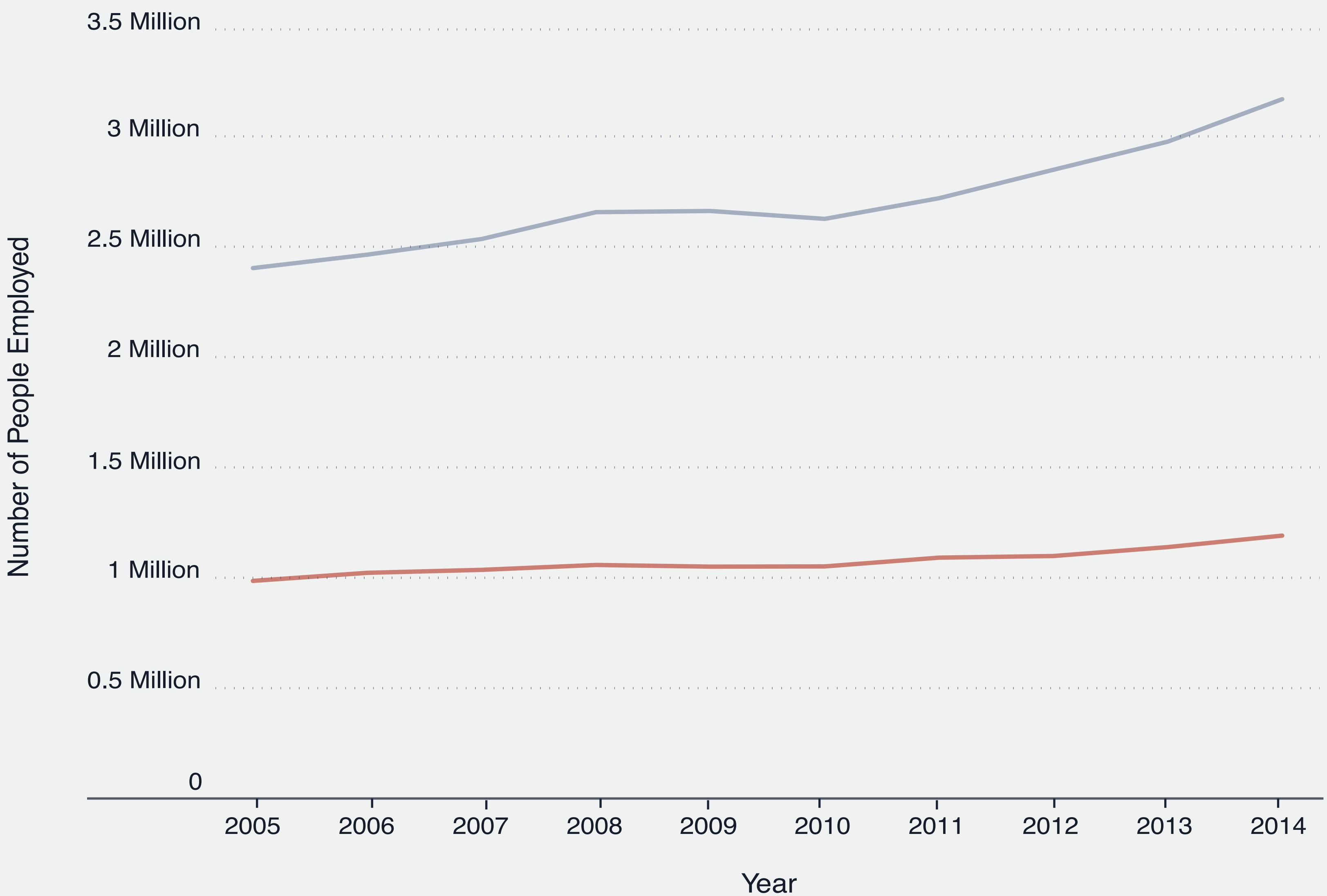
ALL COMPUTER, SCIENCE, AND ENGINEERING (STEM) OCCUPATIONS

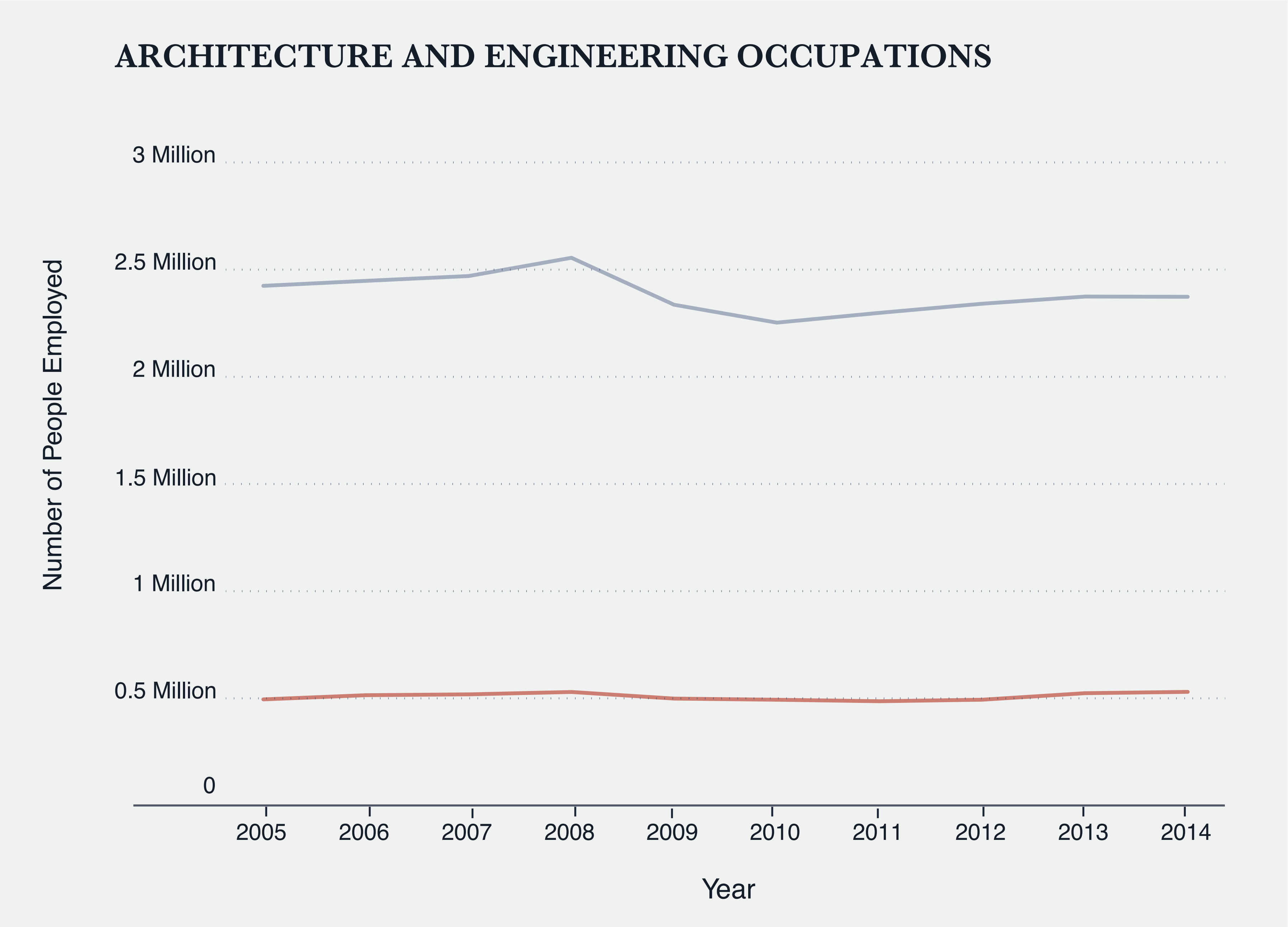
This data, taken from the US Census Bureau's American Community Survey, measures the number of male and female employees in the United States. From 2005 to 2014, the number of women employed in the STEM industy has increased only slightly, despite an overall increase in the number of STEM jobs.



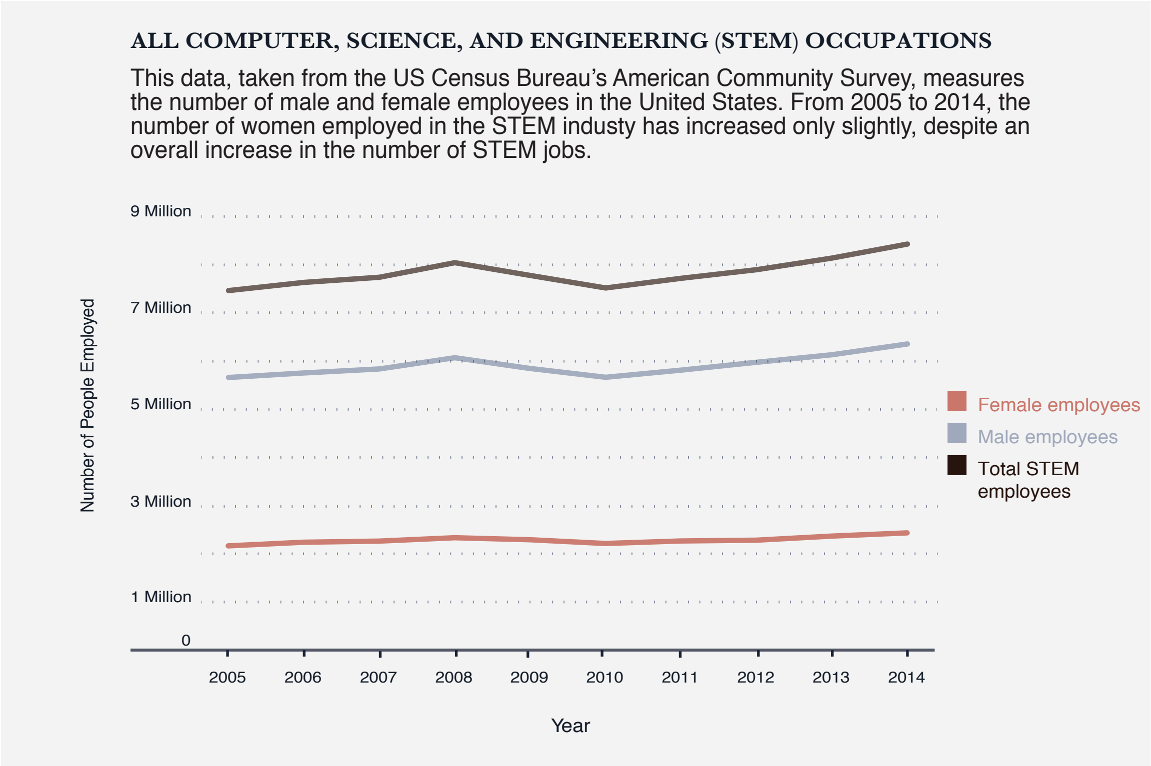

LIFE, PHYSICAL AND SOCIAL SCIENCES OCCUPATIONS
Number of People Employed
3 Million
2.5 Million
2 Million
1.5 Million
1 Million
0.5 Million
0
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
Year

COMPUTER AND MATHEMATICAL OCCUPATIONS




ARCHITECTURE AND ENGINEERING OCCUPATIONS
3 Million
2.5 Million
2 Million
1.5 Million
1 Million
0.5 Million
0
Number of People Employed
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
Year

Where are all the women in STEM?



WOMEN IN STEM

Although women constitute about half of the US labor force (47%), only 25% of them are currently employed in STEM [1]. In some fields like Computer, Mathematics and Engineering this gender gap is increasing significantly as while these occupations haven't shown any significant growth in the number of women entering the industry in the recent years. Increasingly, women employed in STEM are leaving their jobs after some time, thereby causing a decline in the overall percentages of women in these industry. The graphs on the next page show specific sub-categories of the STEM industry. As you can see, the difference between male and female employees is more severe in some sub-categories.

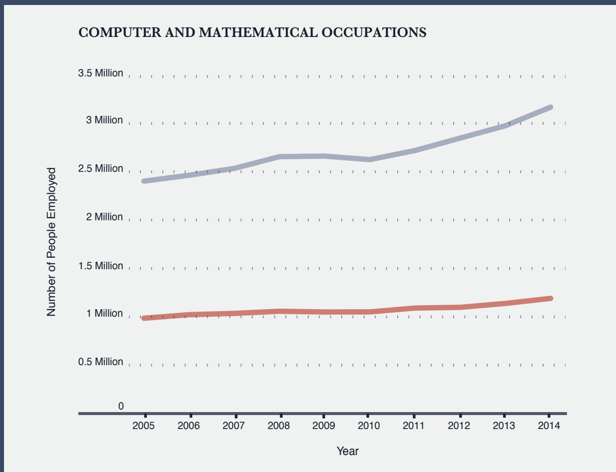
TOP THREE REASONS WOMEN IN STEM LEAVE THEIR JOBS

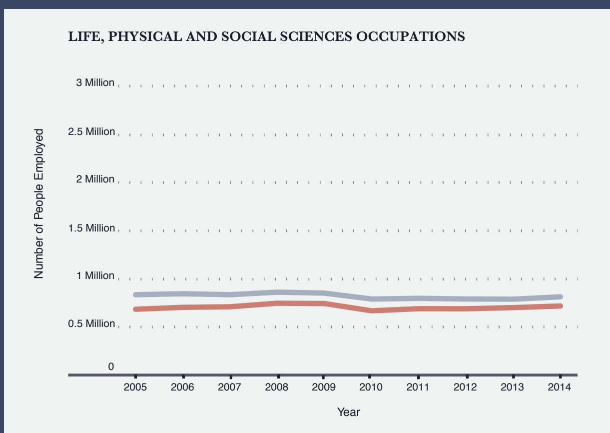
According to the Athena Factor, a research report by HBR [2], about 52% of highly qualified women working in STEM fields quit their jobs due to certain hostile factors like the ones mentioned below:

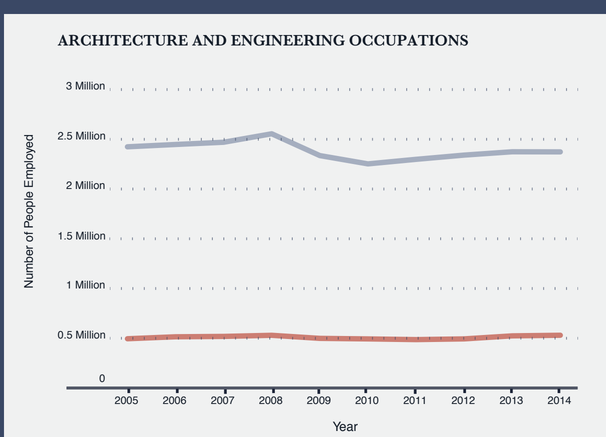
1. Hostile working environment
   Women in STEM have been facing hostile, exlusive male dominated cultures that may even predatory behaviors. 63% of women in the study have reported experiencing sexual harassment [2].

2. Unconscious stereotyping
   Women have faced instances where preexisting beliefs and attitudes towards women as a whole have affected their confidence and performance in the job [3].

3. Isolation
   Women in STEM face lack of mentors and role models to look up to. Also, most of them have faced difficulties acquiring sponsors for their ideas.







WHAT CAN WE DO TO KEEP WOMEN IN STEM?

Leaving behind the stereotypes

Gone are the days when women were considered good only for household work. Now, about 41% of highly qualified scientists, engineers and technologists are women [2]. It's high time that we change the way we talk or even think about women and their capabilities.

Diversity inclusion in recruitment

Organizations should expand their recruiting policies to attract and encourage women to apply for STEM jobs. One such example is Google's recruitment policy for employing a diverse set of people [2].

Mentoring programs

Organizations should develop mentoring programs for every employee that could facilitate an overall professional development as well as a healthy working environment. One example is the Technical Leadership Pipeline Program for Women by Intel [2].

Flexibility in work routine

Inclusion of flexible systems like part-time or telecommuting schedules would help women to balance work and life responsibilities. Examples could be the Freedom to work reform by BT group and flexible leave of absence by IBM [2].

Data Source

US Census Bureau American Community Survey. Occupation by Sex and Median Earnings in the Past 12 Months (In 2005 Inflation-Adjusted Dollars) for the Civilian Employed Population 16 Years and Over (2005-2014): http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_05_EST_S2401&prodType=table

References

[1] D. Beede, T. Julian, D. Langdon, G. McKittrick, B. Khan and M. Doms, "Women in STEM: A Gender Gap to Innovation," U.S. Department of Commerce, 2011.
[2] S. A. Hewlett, C. B. Luce, L. J. Servon, L. Sherbin, P. Shiller, E. Sosnovich and K. Sumberg, "The Athena Factor: Reversing the Brain Drain in Science, Engineering,and Technology," Harvard Business Review, 2008.
[3] C. Ashcraft and S. Blithe, "Women in IT: The Facts," National Center for Women & Information Technology (NCWIT), 2010.
[4] Bureau of Labor Statistics, "Employment Projections: Occupational employment, job openings and worker characteristics," April 2016. [Online]. Available: www.bls.gov/emp/ep_table_107.htm.