

Web Scraping Tufts Professors' RateMyProfessors.com Reviews to Assess Student Bias

San Akdag
CS 116/PS 118

Abstract: This project hones skills learned in both Cyber Security and Information politics to learn something new about the study body at Tufts and to engage them at the same time. I present a paper and a set of Python programs that were used to save and analyze all of the RateMyProfessors.com reviews for Tufts professors. The goal was to find a pattern of bias in review score and content. I discuss the skills learned in Cyber Security used to develop a web scraping tool to compile all of the information available on RateMyProfessors.com about Tufts. Then, using skills learned in both classes, I discuss the method by which I collated and evaluated these reviews in order to assess any underlying biases of student body using publicly accessible data about how they review their professors. As elaborated upon in the results section, the findings indicate that Tufts students rate their male professors more highly, and often describe their female professors more superficially based on the frequency of certain words used.

Introduction

Web scraping is the process of extracting data and content from websites. The goal of this project is to scrape and analyze RateMyProfessors.com (RMP) entries for Tufts professors to try and find any patterns of bias in student reviews. The project was inspired by an existing project¹, which analyzes the frequency of specific words in all reviews on the RMP site. This project is different because it is Tufts-specific, which might allow it to highlight trends that exist at Tufts, but not among all students of higher education. Although many may harbor doubts about the validity of the information posted on the site, studies indicate that in lieu of actual course review availability, RMP serves as a practical alternative for students. The data and programming part of the project had three discrete phases: data collection, data collation/analysis, and data synthesis. Each of these phases required skills learned in both classes. The first phase details the underlying computer science concepts and methods used to collect the information, and explains the

¹ <http://benschmidt.org/profGender/>

political nature of the data collected. The second details the methods and programs used to organize and analyze the data. The third discusses the results of the data analysis and presents some of the key findings, including figures. After the results is a discussion, which includes shortcomings of the project and provides an intellectual underpinning through existing studies. This is followed by a brief application section, outlining some goals for the future of this project, and a conclusion.

To the Community

My goal is that this project does not end when I turn it in. The program and paper, as they stand, are useful, but that is limited to me and the people who grade it. This project is meant to explore the relationship between technology and politics in order to engage my community. The project, in its current form, can be found on GitHub², but is still a command line tool. Therefore, over the winter break I plan to make this project accessible to the community by making a web application in order to see how the Tufts community might use it. The goal is for the community to be able to better see biases that it aware of or discover new ones.

Data Collection

The first phase of this project was data collection. The data, gathered from RMP, includes name, department, and review data for every professor listed on Tufts University's "professors" page. Although this data might not seem like the basis for a project in Information Politics or Cyber Security, the method by which it was gathered and the purpose it serves demonstrate that it is. As mentioned, web scraping is the process by which information and content is stripped from sites. Web scraping poses a security vulnerability for a site in the sense that information can be co-opted and used for purposes other than the intended, like theft of copyrighted content,

² <https://github.com/sanakdag/TuftsRMPScrape>

reselling of information, or undercutting prices. Although this project is not illegal, the data is not being used for its intended purpose.

I began the web scraping process by determining what information to scrape. Each professor page contains the name, department, and individual reviews for every professor. This information is encoded in the HTML of the webpage. The HTML file is what the browser uses to render a webpage and can be viewed using the developer tools in Google Chrome (see Figure 1). Each professor is given a “TID” by the site, which is used as the token in the URL for a JavaScript program called ShowRatings.jsp (see Figure 2). These are the numbers at the end of any professor page URL. So to access the pages for all Tufts professors, all I needed was their TID.

I compiled a list of all of the TID numbers by using another Python script. This script went through a saved copy of the HTML of the Tufts page where all of the professors have been loaded and picks out the element that stores the TID. Once all of the professor IDs were saved to a file, I could start scraping each of the individual professor’s pages. Using a tool called Selenium, I automated a Google Chrome window to load each professor’s page one by one and scrape the results. The first step in the automation process was to install AdBlock, a common Chrome extension that stops ads from loading on a webpage, to ensure the only information being rendered by the browser was relevant. Although this seemed like a relatively straightforward task, a new problem quickly presented itself. On each individual professor’s page, only the 20-most recent reviews loaded when the page was visited; accessing all of the reviews required clicking a “Load More” button which would execute a JavaScript command and send a request to the server which would then send back the next 20 ratings for the browser to render (See Figure 3). The website executes a JavaScript program that loads more content without

refreshing the page. The solution was a simple loop that uses the automation to find, scroll to, and click the “Load More” button until all of the reviews have been loaded. Then the program does what traditional web scraping programs do and copies the reviews from the HTML into individual .txt files. The program organizes files into one of two formats:

FirstName_LastName.txt

FirstName_LastName_reviewX.txt

which hold biographical info about a professor and one their reviews respectively. The fact that I was able to scrape all of this data so quickly is an example of a flaw on the part of the RMP website. Many sites that post information on the web are vulnerable for exploitation. There are many shady services that web scrape public information and resell it for a sizeable fee. Many sites that are targets for web scraping check to see the rate at which individual IPs make requests. At the speed at which these pages were loaded, I was anticipating some sort of reaction on the part of the service, however, none ever made itself present.

Data Collation/Analysis

The data from RMP provided name, department, and all reviews for each professor, but it did not supply gender information. To determine if there was a gender bias, I had to first determine the gender of the professor. One approach would be to look up every professor, find their page on Tufts’ website, and determine gender by the information available there (picture, name, or explicit gender status). However, this seemed too difficult and prone to error. Instead of guessing and running the risk of misgendering professors I used the information available to me, the reviews. Rather than personally make that inference, I relied on the student’s use of gendered pronouns in their reviews. The solution was a Python script that went through all of the reviews for an individual professor, and whichever set of gender pronouns (he/him/his, she/her/hers) was

found with the most frequency, the corresponding gender would be appended to the professor's information file. This took the onus off of me to correctly identify a professor's gender, and put it on the students who reviewed them. Although not always mentioned in technical writing, it is important to stay aware of the fact that the gender binary is not only forcefully coded into how we conduct our lives, but also in how we write code.

With this information in hand, the next step was to find trends in the reviews. The information available to me, subjective professor reviews and rating scores, each provided a valuable resource, one qualitative, one quantitative. In order to come to a reasoned conclusion, I decided to first find the average review for a professor in every department by gender. This was accomplished using a Python script that took the score from every review, and saved that information along with gender and department. Then it simply tallied up the number of reviews of every department and gender combination and found the average. The average professor rating by department is found and presented in 3 arrangements (see Figures 4-6).

The other source of data, the actual reviews, had a more complicated means of analysis. This culminated in a more complicated Python program that, for any given keyword, provides a lot of information about reviews containing that word. First, the program, given a keyword with the '-s' option, can print the reviews containing that keyword using the '-p' option. If one is curious to see if some fairly obscure words are present in the reviews at all, this is a good option. For more common words, like "happy", "smart", "mean", "sweet", etc. which all appear far more times than "verbose" does, we are more curious about who these words are used to describe in what frequency. Using the '-f' option, one can filter by either 'gender' or 'dept' to print the number of times a review occurs for a professors of different genders, or more specifically, departments and genders. The result is a percentage-wise breakdown which is the ratio of the

number of reviews that contain the keyword for a certain department/gender combination and the total number of reviews that contain the keyword. The ‘-n’ option provides a way to normalize the results, comparing it to the overall number of reviews, not just ones that contain the keyword. The ‘-v’ option prints additional information, like notes and average review length for a keyword specific keyword filtered by gender. Using the gender filter with the ‘-n’ option will print in CSV format, which can be opened in Microsoft Excel and Google Sheets. Running the program with the ‘-h’ option, or without any options, will print a helpful help menu.

Results

Results came in two forms: ratings and reviews. The ratings were collected and averaged by gender and department. The number of reviews and ratings an individual professor received was not included in the program, but of the 13 professors with more than 50 ratings, 12 were men. Of the 48 departments where there were average ratings for both male and female professors in that department, men rated higher in 28 of them. The average ratings were ranked in Figure 4-6. The data in Figure 4 is sorted by average male professors rating in a department. The data in Figure 5 is sorted by average female professor rating in a department. And Figure 6 is sorted by the greatest difference between female and male average rating in a department.

The results were interesting not only because of the results of searching certain keywords, but also the kinds of words that yielded the most difference. Using the ‘-l’ and ‘-n’ options in the keyword.py program, I inputted a series of keywords that I came across most often in some of the other studies. These words included some like “nice” and “sweet”, that are used to describe the personality and likeability of the professor, and words like “smart” and “tough” which are more indicative of evaluation, rather than personal character traits. The data shows that Tufts students tend to “evaluate” their male professors only slightly more often, but more interestingly,

use words to describe personality traits like “sweet” and “caring” with much more frequency when reviewing female professors. The bias, in this case, turned out to be subtler than first thought. The reviews indicate that Tufts students do think their female professors are “tough”, “boring”, or “strict”, but are more quick to factor in personality traits that are not as relevant to the course. Although these results do not suggest that there is a large gender bias in terms of professional evaluation, female faculty reviews are still subject to culturally-conditioned gender stereotypes. This view is supported by an older study on formal student reviews³ which found that while students do not show bias in description of course-relevant information, there are observable difference how students perceive their instructors as warm, potent, or self-assured, which are indicative of gender stereotyping.

Discussion

Given this limited data set and limited result, one could argue that these results are insignificant. But as similar studies suggest^{4 5} RMP reviews reflect what might be expected of traditional measures of student learning⁶. There seems to be an academic consensus on the validity of these reviews and the purpose they serve when formal student evaluations are not accessible, but there is a rift in terms of the characterization of the goals of the user. According to one study⁷, the two arguments are two sides of the same coin. One the one side is the participatory agency perspective, which values user-generated content. The author attributes this view to Yochai Benkler in *The Wealth of Networks* (2006) and to another study⁸, which explores the role of RMP as a means of civic exchange. This view values user-generated content and

(note: see sources cited for full citations)

³ Bennett 1982

⁴ Bleske-Rechek & Michels 2010

⁵ Brown, Baillie, Fraser 2009

⁶ Otto, Sanford, & Ross 2008

⁷ Yoon 2015

⁸ Ritter 2008

information sharing on the Internet. According to Ritter, someone who posts on RMP is participating in a public platform which gives students agency to produce and share valuable pedagogical information. The studies suggest that when formal student evaluations are not accessible, RMP allows students to challenge institutional power hierarchies (Yoon 2015) and to control the means and the dissemination of comments (Ritter 2008). The other side sees RMP as a threat to the value of higher education. In this view, RMP positions the student as a consumer rather than a learner. This argument, known as the commodified agency perspective, sees RMP as an example of neoliberalism in higher education. I tend to support the view of RMP as an example in neoliberalism in higher education, but also understand the way in which one works with what one has. If RMP is the only data available to help a student make a more-informed decision, so be it.

With this in mind, we can address some of the lingering issues with the data. Although the results were consistent with existing literature on the subject, some of the numbers might have had more to do with the demographic makeup of certain departments than any bias on the part of the student. The fact that certain keywords appear much more often to describe female professors in the foreign language department might be because a majority of the professors listed on the department website are female. Although this might explain the number of reviews concentrated in this department, it does take away from the findings. These types of findings are more relevant when the makeup of the department is more even, or when calculated across department. Another issue with the findings was the concentration of reviews for specific professors. Only 13 professors received more than 50 ratings, and of them 12 were men. This would push the results towards the opinions students have of those professors. However, this concentration of reviews was present for other schools in Tufts' cohort. This poses an issue

because professors and departments with the most reviews. The resources I had used to theorize this project both either had much larger data sets, or backed up the reviews on the site to reviews submitted through course evaluations, which are usually much more substantive and accurate. Considering that the trend in the concentration of reviews is consistent across colleges and universities in Tufts' cohort while the makeup of the respective department at each these schools is different, another approach would be aggregate reviews over groups of schools like the NESCAC, the Ivy League, or public vs. private colleges of a specific region.

Nonetheless, this project demonstrates how computer science and political science are increasingly being used hand-in-hand to poll the web. According to Monroe et al.,⁹ most studies of this nature do not use random polling or other methods in political science that are traditionally used to collect data. The article suggest that data science and political science can be used together to make better comparisons between populations of interest and observe relevant political behavior that is difficult to detect. It highlights studies that utilize a mass quantity of subtle observations to understand the behaviors of a small, specific population. This requires firm knowledge of both political science research methods and data science.

One point that I would be remiss to not discuss is the role of standard evaluations of teachers. Many of the publications I found in my initial research also advocated that colleges and universities make public their formal anonymous course reviews¹⁰. They argue that the privacy implications of releasing internal teacher evaluations are erased by the prevalence of user-generated sites like RMP. Like others, they argue that students will use what is available to them, whether or not what is available is a proper characterization of the professor, and conclude by advocating that these reviews be made publicly accessible.

⁹ Monroe et al., 2015

¹⁰ Coladarci and Kornfield 2007

Application

The purpose of this paper was to use what I learned in my classes this year to try and use a macro-level lens to learn something specific about my community, but my analysis and results were not exhaustive. This is a shortcoming only if the goal of community engagement is not realized. In order to do that, I would like to build a web application around the keyword program that would be able to graphically represent the results of the command line program. This would allow the public, not just my peers in the computer science department, to access this tool. If all goes to plan, I will be sharing the link to the website to the Tufts community. The website will also record keyword searches. That data could serve as the basis for another study. Hopefully this will engage the community and lead to more interesting results and conversations about the topic. If there is a tangible response to the release of the website, the next step would be to approach the school to develop a similar program to privately analyze the student course evaluation results, which according to the literature cited above are more accurate and ought to be made public.

Conclusion

The programming process that underpins all of the data was an excellent learning tool. To apply what I learned in Cyber Security to Information Politics seemed only logical, but the shape it took was not what I would have expected. The programming portion, scraping reviews and analyzing content provided some tangible results. And despite the limitations of the data set, the reviews, which have been shown to be reliable in lieu of standard teacher evaluations, tend to follow patterns of gender stereotyping on the part of the students to their female professors. I hope to make this paper and tool available to the Tufts community in the hopes that Tufts might be compelled to release their own internal course and professor reviews.

Figure 1: Chrome Developer Tools to Find TIDs on RMP

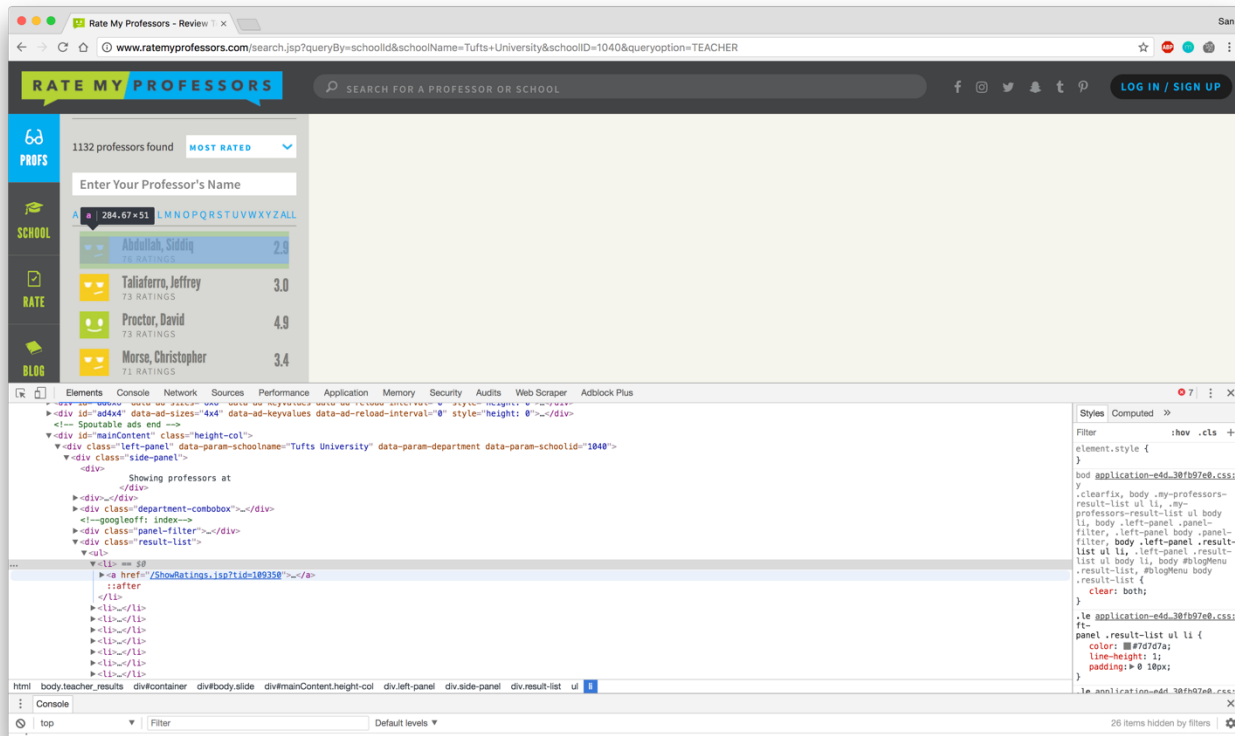


Figure 2: Using Chrome Developer Tools on a RMP Professor Page

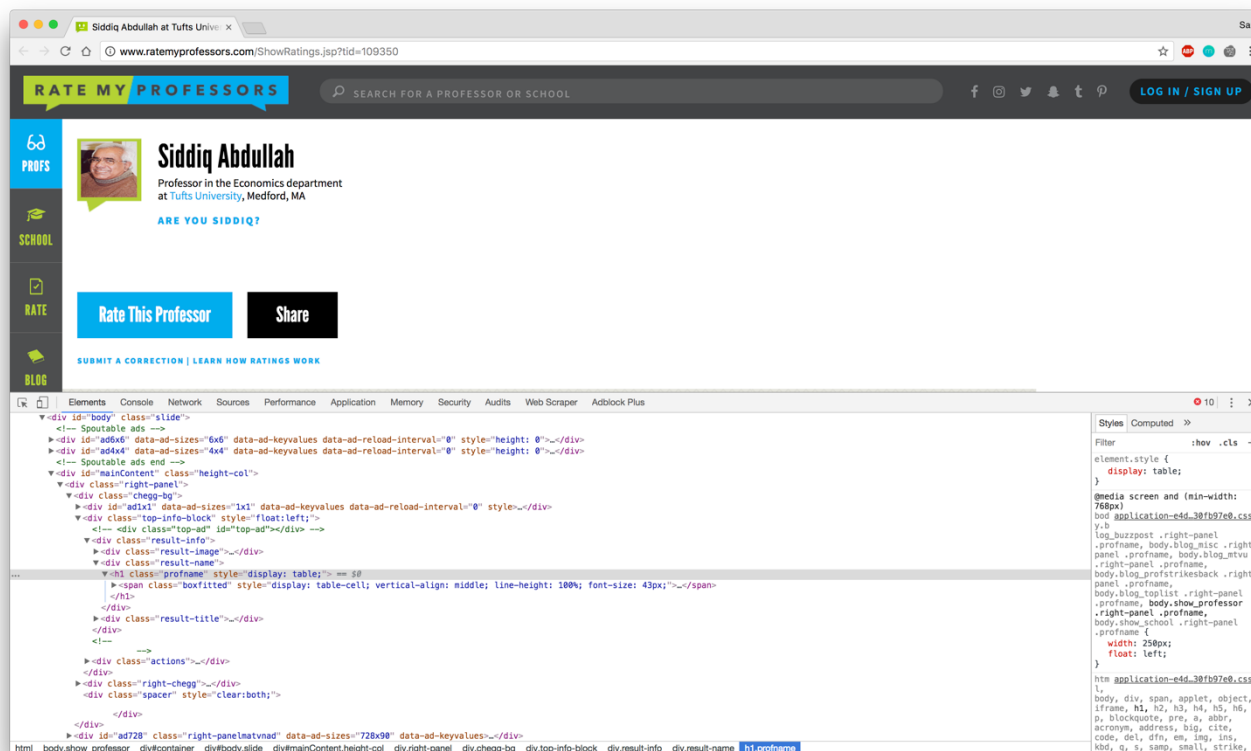


Figure 3: Load More Button

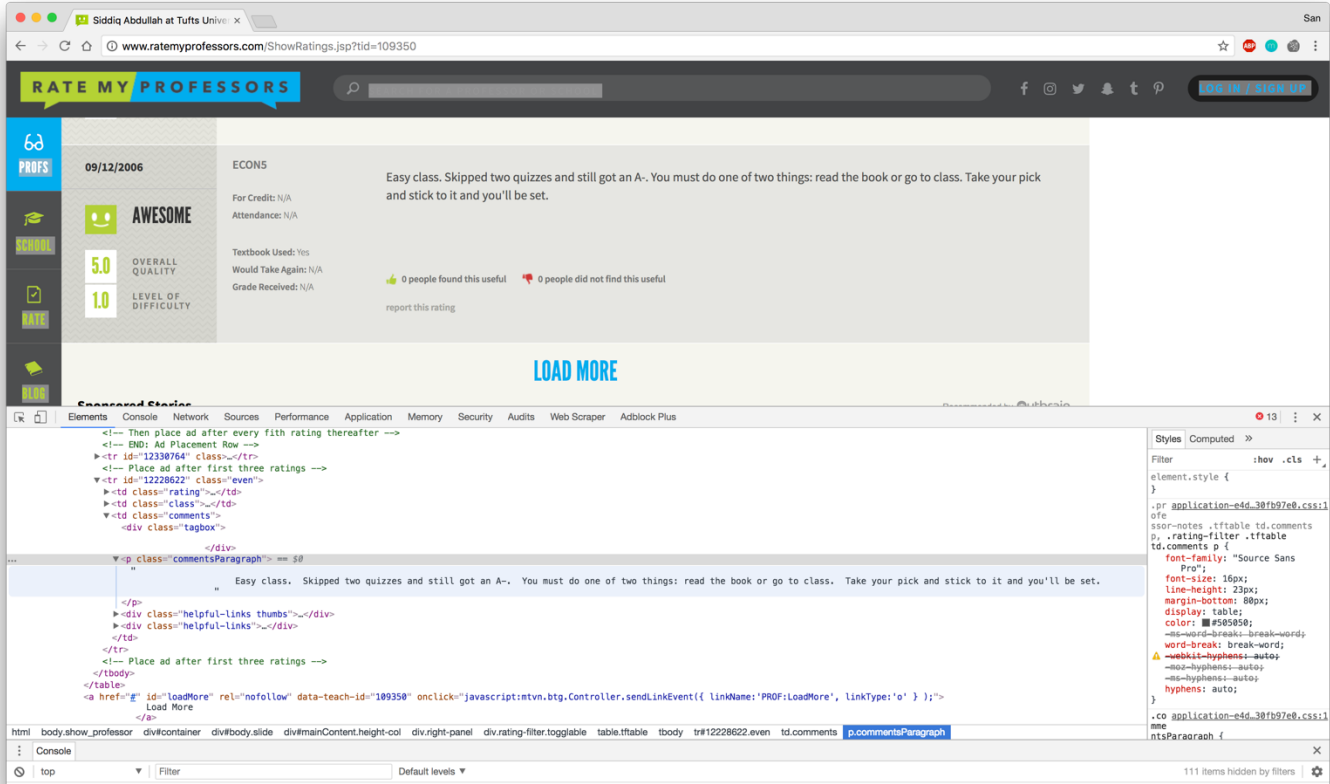


Figure 4: Average Rating by Department (Sorted by Male Rating Score)

Avg. Prof. Rating by Dept.
(sorted by male score)

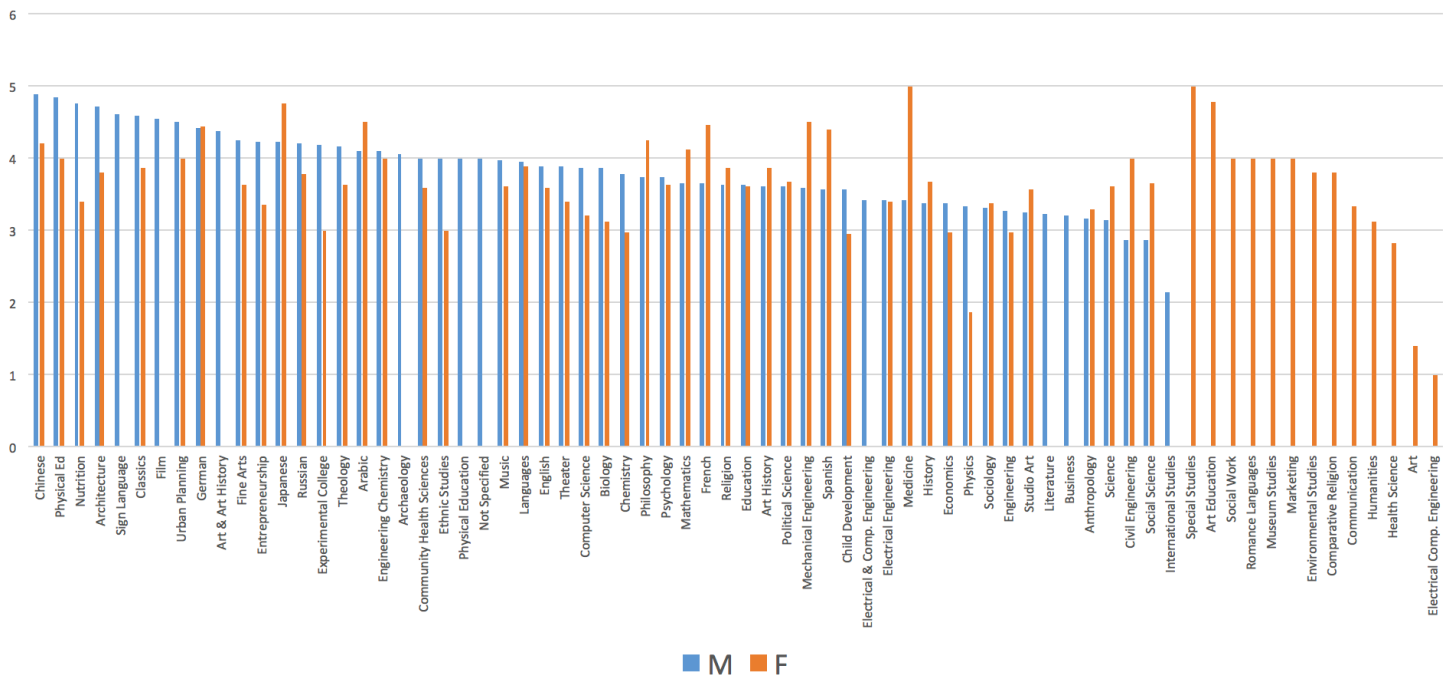


Figure 5: Average Rating by Department (Sorted by Female Rating Score)

Avg. Prof. Rating by Dept.
(sorted by female score)

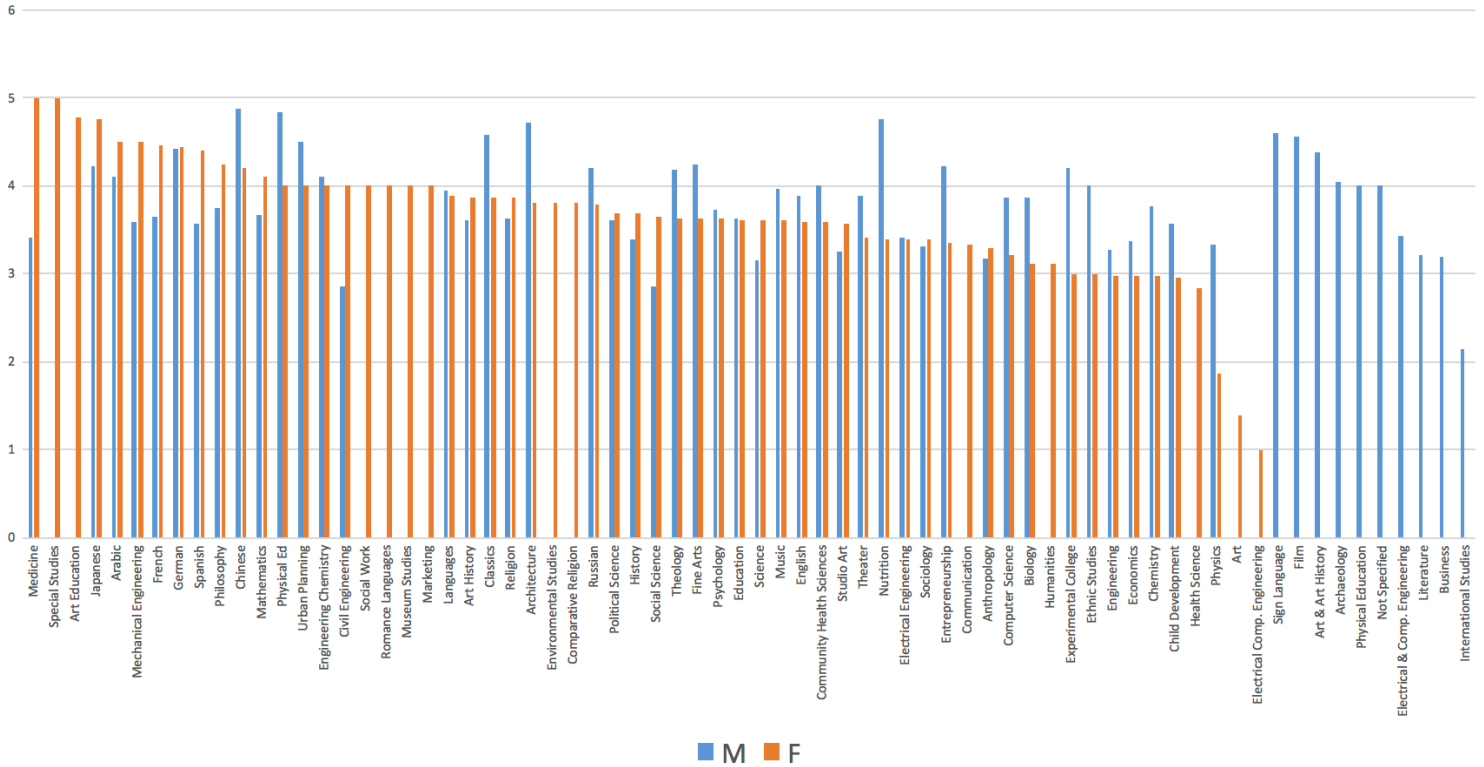


Figure 6: Average Rating by Department (Sorted by Difference in Rating Score)

Avg. Prof. Rating by Dept.
(sorted by greatest difference of f-m)

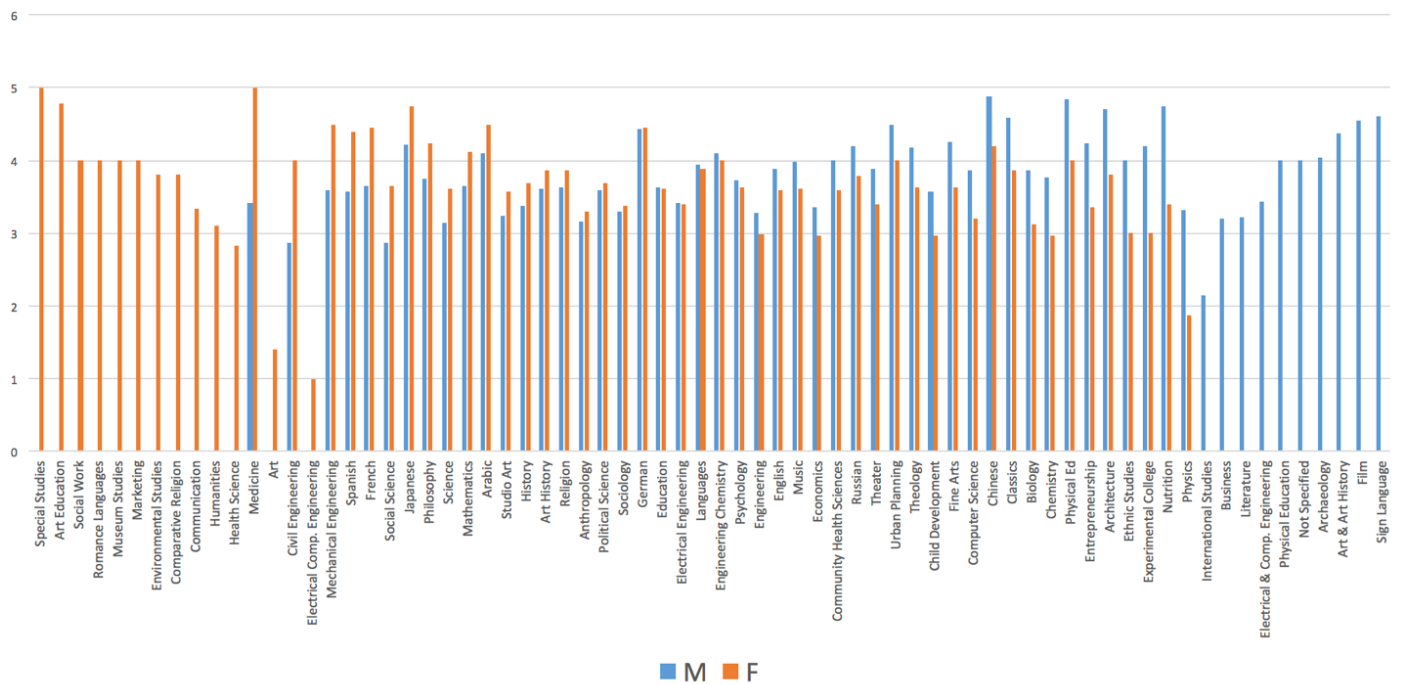
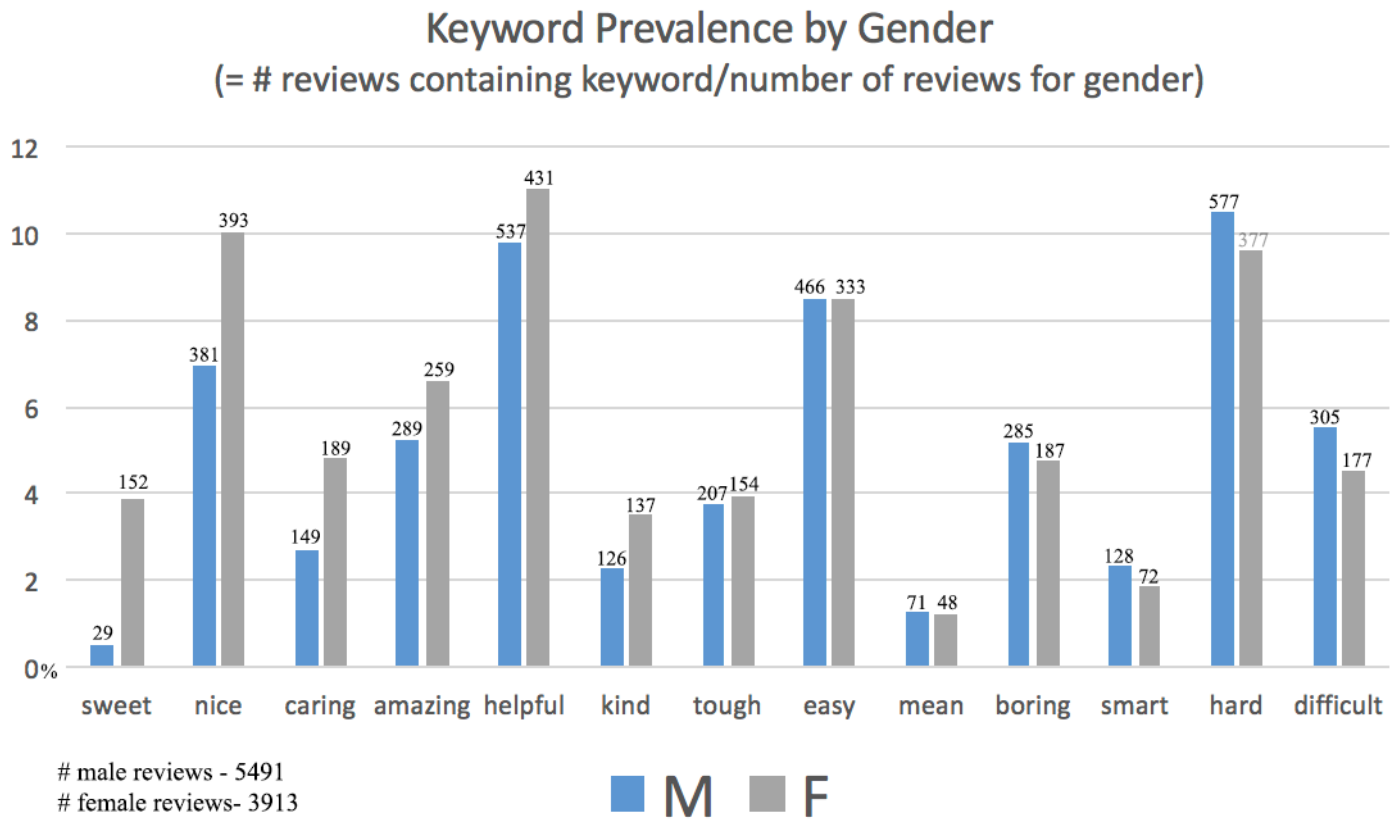


Figure 7:



Sources Cited

- Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, 74(2), 170. <http://psycnet.apa.org/record/1982-24396-001>
- Bleske-Rechek, A., & Michels, K. (2010). RateMyProfessors. com: Testing assumptions about student use and misuse. *Practical Assessment, Research & Evaluation*, 15(5), 1-12. <http://www.bleske-rechek.com/April%20Website%20Files/Bleske-Rechek%20and%20Michels%202010.%20RMP.com.pdf>
- Brown, M. J., Baillie, M., & Fraser, S. (2009). Rating RateMyProfessors. com: A comparison of online and official student evaluations of teaching. *College Teaching*, 57(2), 89-92. <http://www.tandfonline.com/doi/abs/10.3200/CTCH.57.2.89-92>
- Coladarci, T., & Kornfield, I. (2007). RateMyProfessors. com versus formal in-class student evaluations of teaching. *Practical Assessment, Research & Evaluation*, 12(6), 1-15. <http://pareonline.net/getvn.asp?v=12&n=6>
- Monroe, B. L., Pan, J., Roberts, M. E., Sen, M., & Sinclair, B. (2015). No! Formal theory, causal inference, and big data are not contradictory trends in political science. *PS: Political Science & Politics*, 48(1), 71-74. <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/S1049096514001760>
- Otto, J., Sanford Jr, D. A., & Ross, D. N. (2008). Does ratemyprofessor. com really rate my professor?. *Assessment & Evaluation in Higher Education*, 33(4), 355-368. <http://www.tandfonline.com/doi/pdf/10.1080/02602930701293405>
- Ritter, K. (2008). E-valuating learning: Rate my professors and public rhetorics of pedagogy. *Rhetoric Review*, 27(3), 259-280. <http://www.tandfonline.com/doi/abs/10.1080/07350190802126177>
- Yoon, K. (2015). Affordances and negotiations of the digital reputation society: a case study of RateMyProfessors. com. *Continuum*, 29(1), 109-120. <http://www.tandfonline.com/doi/abs/10.1080/10304312.2014.968525>