Fixing the Flaws of Networking

An Alumni Directory of the Harvard Women's Lacrosse Program

Grace Rotondo

December 13, 2019

Abstract

Networking is one of the most important skills a person can develop in life. Grades are important; however your network of people is what is going to help you ultimately land a job and launch a career after college. In my own job search, I found that Harvard's alumni resources are out-of-date and unreliable. To solve this problem, I developed an accurate and functional alumni directory on a much for the Harvard Women's Lacrosse Program. I compiled data using several Harvard resources as well as LinkedIn. Through this process, I found that Harvard Women's Lacrosse almni hold many powerful positions.

Background and Significance

Networking is one of the most important skills a person can develop in life. Grades are important; however your network of people is what is going to help you ultimately land a job and launch a career after college. As I was beginning my job search, I wanted to learn more about different industries so that I could better focus my search before applying anywhere. The best way to go about this would be to reach out to individuals in certain fields; however, it is more difficult to get a reply from someone with a cold introduction rather than a warm introduction. So, I shifted my focus toward alumni of the Women's Lacrosse team because I am a current member and know that our alumni are very supportive of the program and those involved in it. However, I found that there was no way to find out what our alumni were doing or where they were; the only way to contact an alum was to ask our coach for a specific person's contact information. Before doing so, though, we would need to know who the alum was and her background, and there was no way to know this information except through word of mouth. Thus, I found the need to create an alumni directory with accurate information.

Methods

I began this project by reaching out to the Assistant Director of Communications, Jeff Weinstein, for the Harvard Varsity Club and the Assistant Director of the Harvard Varsity Club, Jess Perillo, to gain access to information about Harvard Women's Lacrosse alumni. Mr. Weinstein sent me the link to the Varsity Letterwinners data and Ms. Perillo gave me access to an excel spreadsheet that included names and contact information of alumni. I used the Letterwinners data to create my own spreadsheet of names, titled *WLAX_LETTERWINNER_DATA_complete*.

In this Letterwinners spreadsheet, I included several other variables that I felt would be useful when searching individuals in the database. These variables included: *Graduation Year, Last Year Earned Varsity Letter, House, Concentration, Secondary Concentration, Citations, Graduate School, Company, Role, Industry, Area, LinkedIn,* and *Website Other*. I found the information for the "Graduation Year" and "Last Year Earned Varsity Letter" from the Letterwinners website. The *Graduation Year* variable corresponds to individuals' year of graduation and the *Last Year Earned Varsity Letter* variable refers to the last year recorded of an individual playing lacrosse and earning a varsity letter. Though I initially found the information for these variables from the Letterwinners website, I discovered that some of the data in the Graduation Year variable was incorrect when I cross checked it with University records on the Harvard Alumni Directory website. Assuming the graduation records on the Harvard Alumni Directory are official and correct, I used the graduation years found on the Harvard Alumni Directory and documented them in the Letterwinners spreadsheet. Using the Harvard Alumni Directory again, I found the information for the *House* and *Concentration* variables. These variables were important to include because the housing system at Harvard plays a significant role in a student's Harvard experience, thus housing affiliation is a useful conenction point, and concentration information is valuable to know as another connection point. I also used the Harvard Alumni Website to find some individuals' information about *Graduate School*. For variables *Secondary Concentration, Citations, Graduate School, Company, Role, Industry, Area,* and *LinkedIn,*, I searched individual names on Google and LinkedIn and recorded the information I found from LinkedIn profiles. I found that alumni who graduated in the 1980's tended not to have profiles on LinkedIn, so I found their information through websites on Google. I thus input these websites in the vairable, *Website Other*.

After compiling the Letterwinners data, I uploaded both datasets to RStudio and cleaned them. I then merged them using the command full_join() so that I would not lose any of the information; I named the new dataset *data*. From there, I needed to further clean the variables in the dataset to create a new dataset, *data_1*. The next step in preparing the Shiny web application was to find a dataset with coordinate information of cities in the United States. I chose to only find information regarding cities in the United States because the majority of the alumni was from cities in the United States. I found this dataset from the United States Cities Database. From this website, I downloaded the Excel file and imported it into R. I then selected the two variables I needed to correspond to the city and state variables in my greater dataset: *city_ascii, state_id, lat,* and *lng*. I then renamed *city_ascii* and *state_id city* and *state*, respectively. This

preparation allowed me to then join the coordinate information with the greater dataset to create *full_data*, joining by the *home_city* variable with the *city* variable and the *home_state* variable with the *state* variable so that the coordinates would join by the correct cities and states.

After creating the full dataset, I shifted my attention to the map creation. I used the Leaflet package and found code online that created an interactive map of the United States. This was in preparation for what I would input into the Shiny server. The final step in my Shiny preparation was making the *full_data* dataset into an RDS file so that Shiny could use it.