names

Jian Lu '19

2/12/2017

Abstract The **names** package uses the graduation data of the senior class each year published in the Williams College Course Catalogs and reformats it into a data frame, so that the data can be easily manipulated. This is done for thirteen years from the Class of 2003 to the Class of 2015. The secondary component of this package creates different line graphs that display the proportions of Jewish people, and the level of merit received upon graduation.

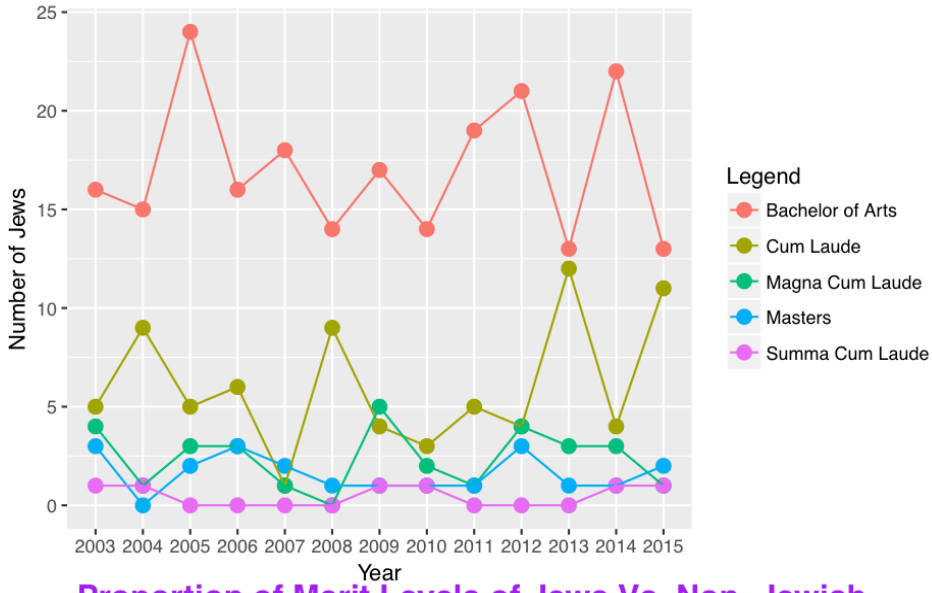
Intrduction

The data on graduating students of Williams College is an interesting data set that shows the merit each individual student has received at Williams. The data frame constructed from this data contains information from all students from the Class of 2003 to the Class of 2015; however, only students who had graduated with honors had their majors mentioned, so the data set does not provide information on the major of every single student. It includes students who have graduated from the undergraduate class, as well as the economic policy graduate school students. The variables in the data set include graduation year, major (if available), type of honors (if available), Phi Beta Kappa, Sigma XI, and ethnicity (specifically is the student of Jewish descent). With the tools available in this package, the success rate of Jewish people at Williams College can be measured.

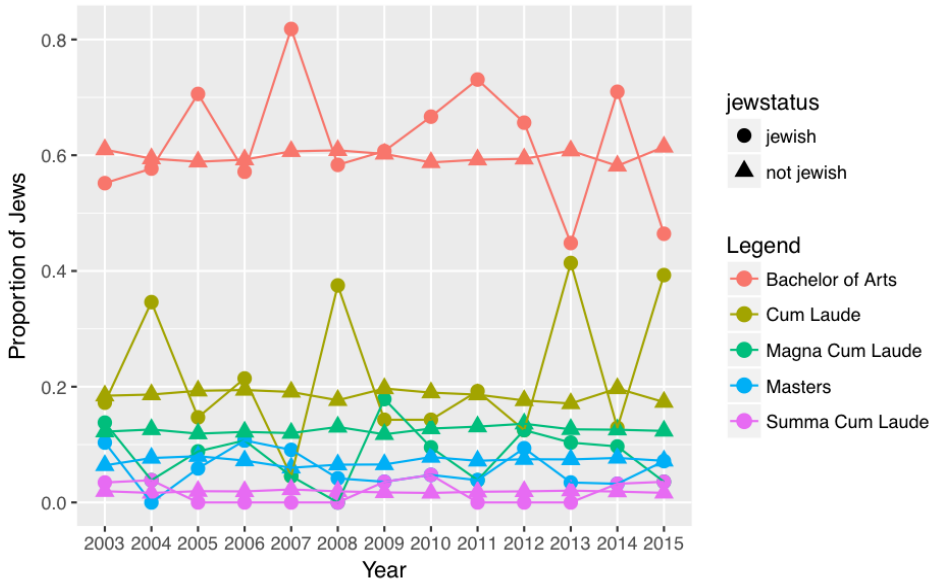
Data

The data used to construct the overall data frame in this package was taken from the Williams College Course Catalogs, from the website of the Office of The Registrar of Williams College, from years 2003 through 2015 by copy pasting the text directly into text files. Since there was some slight variation in the method that the information was presented each year, some modifications were made to the text files. The data files were changed so that every single line contained the information of only one student, as well any page numbers, and irrelevant information was removed. Eventually, all of the data files were organized in the same homogeneous fashion, and allowed for the function **readnames** to easily organized the information into a data frame. Each row created includes the full name of each person, and is followed by information personal information in the same row. The last variable of ethnicity is not provided by the college. Thus, a list of around 3000 Jewish last names was used to check if the student was of Jewish descent, or not.

Merit Levels of Graduating Jews



Proportion of Merit Levels of Jews Vs. Non−Jewish



Conclusion

The summary of the data shows that the mean number of Jews that attend Williams College each class year are 27.53 students compared to the mean of 530.61 non-Jewish students that are in each class year. The range of Jewish students in each class year was 13 with the minimum of 21 students in the Class of 2010, and a maximum of 34 students in the Class of 2005. It can be seen from the timeline graphic that the count of Jewish people at Williams College have stayed relatively the same over the course of 13 years at Williams College. We can see that the is not much deviation from the mean of 27 Jewish students attending Williams throughout the years. This may point toward the fact that Williams College is fulfilling quotas of certain ethnic groups. Although significant conclusions can only be drawn after receiving data, and testing not just Jewish people, but African Americans, Asians and Caucasians. The comparison graphic show that Jewish people at Williams perform on average the same as the rest of the non-Jewish population. Within the top of the class receiving, a Master's Degree, Magna Cum Laude and Summa Cum Laude, the Jewish population typically performs on par with the rest of the student population. Although there are some dips in performance in the classes of 2007 and 2015, and some out performance in the classes of 2008, 2013 and 2015. From this data, we can infer that the Jewish ethnicity is independent of successfulness at Williams, as the proportions of merit levels of the Jewish population and the non-Jewish population are relatively similar.

The **names** package includes functions that easily create graphics of on the proportion of Jewish students at Williams College and the level of merit they have received upon graduation. The data set created in the package is organized so that it can be easily subsetted and tested for other correlations, or interests of the user. For example, one can view all the types of majors at Williams college, and subset the majors with ethnicity to view which majors are most popular with the Jewish population. However, this package can certainly have many improvements. Currently, it only tests for students of Jewish descent based on last name, but certainly other ethnicity tests can be added in the future to make the research a lot interesting. With the current method of labeling students of Jewish descent, there are some Type I and Type II errors, so a huge development for the package would be to get the ethnic data on each student from Williams and implement it into the data frame. It would be ambitious to conclude that ethnicity is completely independent of success at Williams, but this poses a very interesting question and with further manipulation of the data on students it is completely possible to look into this hypothesis. This package sets the foundation for further observational tests of Williams College graduates and improvements in the data frame could yield very interesting results.