Findings in the Ages of the Williams College Faculty

Johnson Nei

2016-1-30

Abstract

This R package serves to answer the main question of what is the average age of the Williams College faculty using the 2013 through 2014 data of Williams faculty provided by Williams College. The main functions of this package are to clean and manipulate the data so that it may be more pliable to further functions, and to provide a general summary of the Williams College Faculty's age, including the Average, maximum, minimum, range of the ages, and so forth. Moreover, this package explores some small extensions such as sorting by the department the faculty are a part of through cleaning the data.

Background

School Background

Williams College is a private institution founded in 1793 with a general total undergraduate enrollment of around 2,000. Its setting is rural with a campus size of 450 acres. The college utilizes a 4-1-4 based academic calendar. Along with its 3 academic branches, Williams is also known for having close-knit small classes with a student-teacher ratio of 7-to-1. Williams College ranks number 1 in the 2016 Best National Liberal Arts Colleges.

Interest Background

Having preserved within the Top National Liberal Arts Colleges, much of which are due to the faculty that the school is so proud of, we are interested in knowing the demographic and general age at which a college's faculty should be.

Structure of the package

This R package requires the use of 3 other packages (devtools, dplyr, stringr) whose libraries are recommended to be opened for this package to perform to its fullest.

Upload

The upload function requires that you know the location of where your data, the data of Williams college 2013 2014 is in. You may then input this location in qutations into the upload function, which will then create the variable Faculty, to become a dataframe containing this data.
e.g. upload("/Users/jnjohn1995/Downloads/faculty/Faculty/faculty13.txt")

Cleaning

The clean function is the backbone of this package, it allows the other functions to be more easily accessed and utilized by first creating a easy to use dataset. The clean function performs in 3 parts, and most preferably in the following order: cleanyr1(dataframe), cleanyr2(dataframe), cleanyr3(dataframe), cleanname(dataframe), cleansubject(dataframe).
We assume that the faculty received their degree in the 1900s or 2000s. cleanyr1 first checks to see whether the 3rd column, which should contain the year that the faculty received his or her degree, starts with a 1 or 2. If it doesn't delete the frame and shift the frame on the right left by 1 while filling the whole that this creates with an NA at the end of the row of the dataframe. This function keeps shifting until the condition that the string starts with a 1 or 2 is reached or is NA.
After the 3rd column has been cleaned so that the only objects that may exist in the 3rd column are the year, NA, or a blank, cleanyr2 removes all rows that contain an NA in the 3rd column.
Similarly, after cleanyr2 has been processed, cleanyr3 removes those rows that are empty for column 3.
Now we are left with a data frame that only has the year at which the faculty received their degree in column 3.
Keep in mind one should assign a variable to the function since each function returns an unassigned dataframe. e.g. Faculty<- cleanyr1(Faculty)

The next part of the data cleaning is cleaning the names using cleanname. Cleanname essentially deletes all the extraneous symbols in column 1 where the name of the Faculty member is suppsoed to be.

The final part of the data cleaning, and also the most tricky, is changing the faculty's position, i.e. Associate Professor in Mathematics, to just their department, i.e. Mathematics. This allows one to more easily access the department to which the faculty is a part of and can therefore discover more about Williams College Faculty through each department and also create other histograms, plots, and so forth.

Summary

The summary function includes many other small functions that help one discover more about the age of the Williams Faculty. Funcions include summary, MaxAge, MinAge, MedianAge, and AgeRange, all of which are fairly self explanatory.

Histograms

The histogram functions merely provide functions to help create the histograms for one to more clearly view the data available.

Results

Through testing our data, I have received the results that the Average Age of the Williams College Faculty is about 48.93, with the Minimum age being 25, maximum at 78, and a total range of 53years.