# Abstract

## Abstract

The kane2017 package utilizes the data published in Williams Online Catalog to assess the number of students with Chinese/Cantonese/Japanese origins and compare their academic achievement with other students. This is done for 15 years from 2002 - 2016 by scanning the data directly from the text version of the webpages. A sequences of statistical comparisons will be utilized to examine the academic achievement of (partial) Asian International/American students.

## Introduction

College admission, espeically recently, has been suspected of raising a higher standard towards Asian students. Though it is unclear that the admission standard has been raised at Williams, looking at the Asian students ratio of attaining Latin honors will be a useful tool in assessing the overall academic performance of Asian American students and can thus further reflect on the bar set during admission. The package kane2017 employs string manipulation(splitting, pattern detection) to split each string into its last name.

The data set is further manipulated by comparing each of the last names to a pre-existing list of possible last names of Asian origin. But processing whether if these last names are of Asian origin, we can then figure out the ratio of Latin honors among Asian students.

## Data

The information used to construct the dataset is taken from the Registrar office. The text file is manually converted from the catalog through copying and pasting.

Though the input are quite concise, since last names and honors are the only things we need, we need to cut down the strings.

For example, for input: *+Kathleen Malone Palmer, with highest honors in Neuroscience, readHonor will use unlist(str_split(data[j],","))  and gsub(". ","",name)* to turn the original string into Palmer. After obtaining a list of last

names, we then compare the last names with the given list of last names. Now we have chineseLastName, japaneseLastName, cantoneseLastName, and totalVeri, which is the collection of the previous three. The name verifies and years are previously stored through RData and passed into readHonor(data, verifier)

readHonor will then return a data.frame as below

```
readHonor(Data2012,chineseLastName)
```

```
##      name honor SummaPer MagnaAbovePer CumAbovePer percentSigma percentPhi
## 1    Meng Summa 6.060606      24.24242    36.36364     9.433962   8.823529
## 2    Wang Summa 6.060606      24.24242    36.36364     9.433962   8.823529
## 3      Li Magna 6.060606      24.24242    36.36364     9.433962   8.823529
## 4      Li Magna 6.060606      24.24242    36.36364     9.433962   8.823529
## 5    Ting Magna 6.060606      24.24242    36.36364     9.433962   8.823529
## 6   Xiong Magna 6.060606      24.24242    36.36364     9.433962   8.823529
## 7      Yu Magna 6.060606      24.24242    36.36364     9.433962   8.823529
## 8   Zheng Magna 6.060606      24.24242    36.36364     9.433962   8.823529
## 9     Cha   Cum 6.060606      24.24242    36.36364     9.433962   8.823529
## 10  Huang   Cum 6.060606      24.24242    36.36364     9.433962   8.823529
## 11   Song   Cum 6.060606      24.24242    36.36364     9.433962   8.823529
## 12   Yong   Cum 6.060606      24.24242    36.36364     9.433962   8.823529
## 13     Bi       6.060606      24.24242    36.36364     9.433962   8.823529
## 14  Chang       6.060606      24.24242    36.36364     9.433962   8.823529
## 15   Chen       6.060606      24.24242    36.36364     9.433962   8.823529
## 16    Chu       6.060606      24.24242    36.36364     9.433962   8.823529
## 17    Cui       6.060606      24.24242    36.36364     9.433962   8.823529
## 18   Dong       6.060606      24.24242    36.36364     9.433962   8.823529
## 19  Huang       6.060606      24.24242    36.36364     9.433962   8.823529
## 20  Huang       6.060606      24.24242    36.36364     9.433962   8.823529
## 21   Kang       6.060606      24.24242    36.36364     9.433962   8.823529
## 22    Lin       6.060606      24.24242    36.36364     9.433962   8.823529
## 23    Lin       6.060606      24.24242    36.36364     9.433962   8.823529
## 24    Liu       6.060606      24.24242    36.36364     9.433962   8.823529
## 25    Liu       6.060606      24.24242    36.36364     9.433962   8.823529
## 26   Shen       6.060606      24.24242    36.36364     9.433962   8.823529
## 27  Sheng       6.060606      24.24242    36.36364     9.433962   8.823529
## 28   Wang       6.060606      24.24242    36.36364     9.433962   8.823529
## 29   Wang       6.060606      24.24242    36.36364     9.433962   8.823529
## 30   Wang       6.060606      24.24242    36.36364     9.433962   8.823529
## 31    Yun       6.060606      24.24242    36.36364     9.433962   8.823529
## 32   Zeng       6.060606      24.24242    36.36364     9.433962   8.823529
## 33  Zhang       6.060606      24.24242    36.36364     9.433962   8.823529
```

By using read(honor, verifier), we can then utilize the data frame and look at the ratio of students of Asian origins with different Latin honors in comparison to the average ratio of Latin honors among college students.
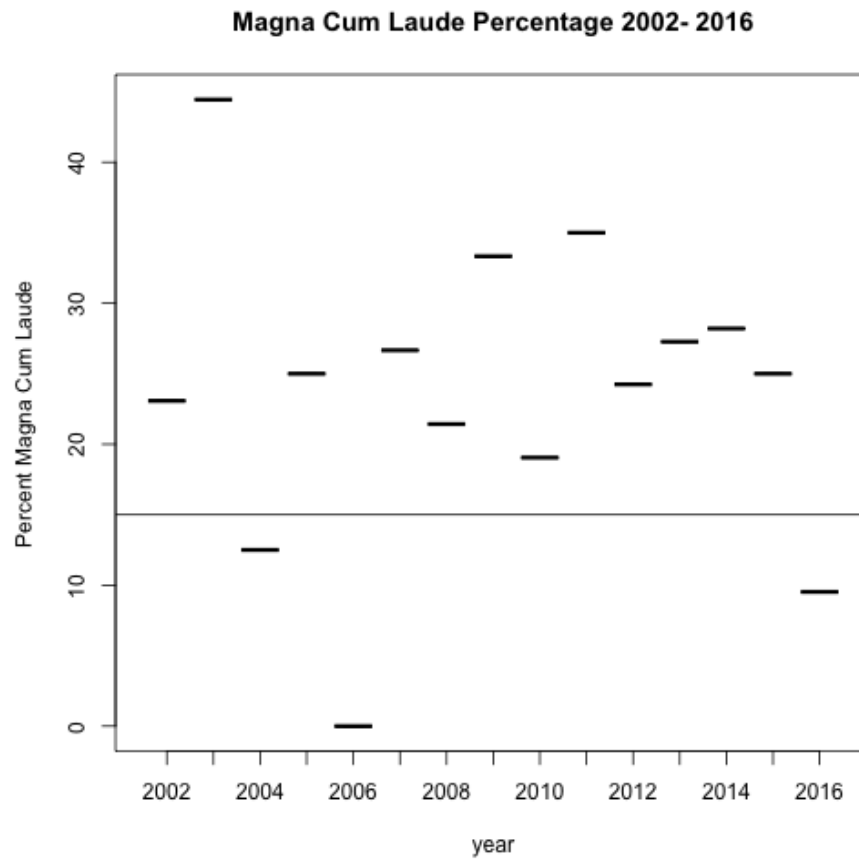
```
plot(read("Summa",chineseLastName),  ylab="Percent Summa Cum Laude")
```

```
## [1] "Summa"
```

```
title(main="Summa Cum Laude Percentage 2002- 2016")
abline(a=2, b=0)
```
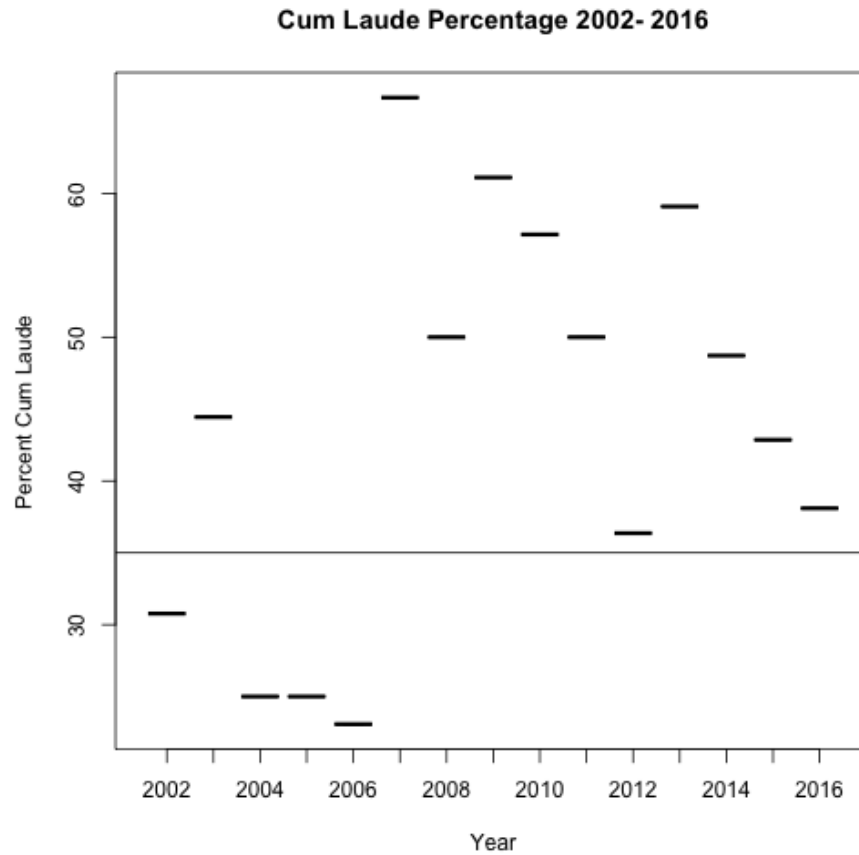
**Summa Cum Laude Percentage 2002- 2016**



```
## [1] "Magna"
```

**Magna Cum Laude Percentage 2002- 2016**
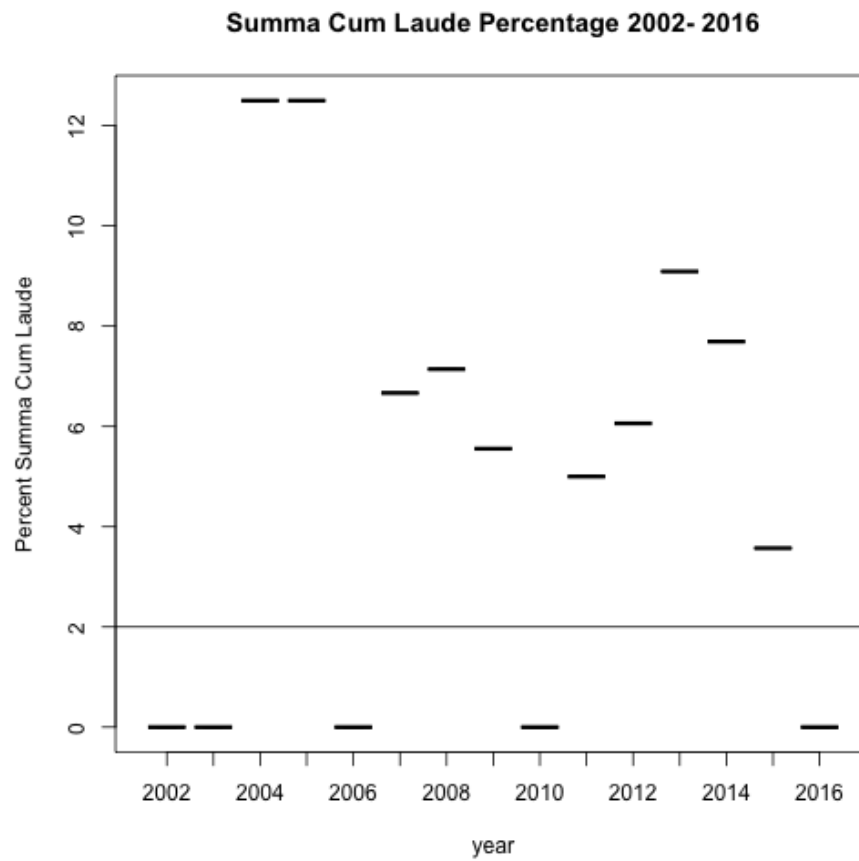


```
plot(read("Cum",chineseLastName), xlab="Year", ylab="Percent Cum Laude")

## [1] "Cum"

title(main="Cum Laude Percentage 2002- 2016")
abline(a=35, b=0)
```

4

**Cum Laude Percentage 2002- 2016**



By attaining the mean value of different origins over 16 years through originAverage(verifier) that takes in country of origin, one is able to compare the average Latin honors with that of students of Asian descend.

```
plot(originAverage("Cum"), ylab="Average Percent Cum Laude")
```

```
## [1] "Cum"
## [1] "Cum"
## [1] "Cum"
```

```
title(main="average Cum Laude 2002- 2016")
abline(a=35, b=0)
```

## average Cum Laude 2002- 2016
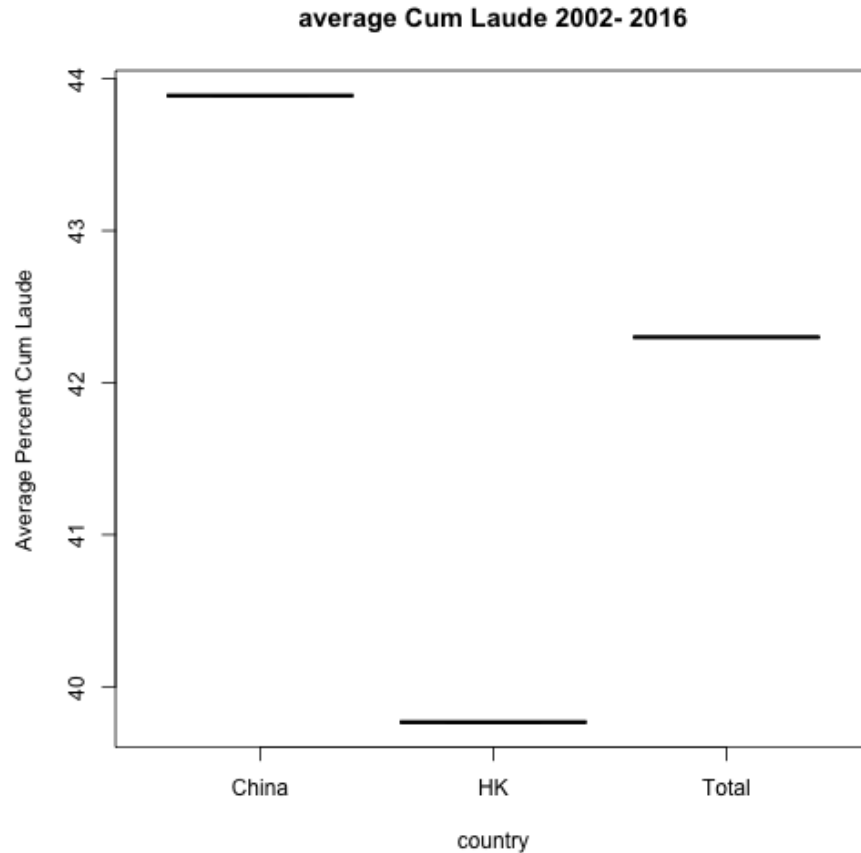


## Conclusion

The summary statstics indicates that according to the Asian last name reference sheets, students of Asian origins tend to be significantly overachieving, with an overall average Cum Laude percentage around 7.5% higher than the average 35%. Also, in the scatterplot graphs, the majority of all stats over years is higher than the average ratio of that particualr honor and above. Thus, it can be inferred that Asian students tend to be more academically achieving. Such conclusion, therefore, should serve as a powerful evidence towards some of the variations in admission standards Asian students might be facing.

However, this package is still incomplete in many ways. First, distinguishing last names from its linguistic composition can be extremely ambiguous. It doesn't take into account some of the identical last names different cultures can share. Also, it doesn't take into account students of mixed racial identity. Therefore,

6

the next step will be to incorporate more last name recognizing templates as well as trying to distinguish last names of ambiguous identities base on first names.