

# Geographic Distribution of Williams College Student Body from 2010-2015

*Nikhil V. Palanki '20*

## Introduction

The **geodist** package analyzes changes in the geographic distribution of the Williams College student body from 2010-2011 to 2015-2016 academic years. This package specifically answers the following three questions: How does the international student body change from 2010 to 2015? Which states consistently have the highest number of current Williams College undergraduates? Which countries consistently have the highest number of current Williams College undergraduates? Given that geographic diversity plays a role in admissions decisions as well as the overall identity of the Williams College student body, examining the trends in diverse student enrollment is important to ultimately evaluating the effectiveness of the Williams College Admissions Departments' domestic and international outreach efforts.

## Dependencies & Code

For the **geodist.R** file, there are not any external packages or libraries required to run the code. For the report, the **knitr:kable** package was used to both generate the pdf and present the sample tables nicely. In order to run the geodist.R file, move the **geodist** package to your desktop and run the geodist.R file since the directory paths within the file are already configured to find the dependent .csv files from the desktop.

## Data

This package analyzes data provided by the Office of the Registrar of Williams College. The PDF catalog documents compiled by the registrar were converted to text files by an online pdf-to-text converter and the appropriate data for the student body were further truncated. The data were then converted manually to .csv files.

For each year, the data were split into two tables: a table for international students and a table for domestic students. The .csv files, which are labeled by "year" and either "intl" or "dom," are loaded in a block of

```
library(knitr)
# Load the .csv files
dom2015yr <- read.table("~/Desktop/geodist/csvs/2015dom.csv", header=T, sep=",")
intl2015yr <- read.table("~/Desktop/geodist/csvs/2015intl.csv", header=T, sep=",")

# Convert to Data Frames
intl2015df <- data.frame(intl2015yr)
dom2015df <- data.frame(dom2015yr)

# Reorder Data Frames
orddom2015 <- orderbystudents(dom2015df)
ordintl2015 <- orderbystudents(intl2015df)

# Calculate Perc. of Int'l students
intlsum2015 <- sumofstudents(ordintl2015)
intlperc2015 <- approxperc(intlsum2015)
```

The following are sample results. The first value is an approximate percentage (with the total student body estimated at ~2200) of international students at Williams College during the 2015-2016 academic year. The second table shows the top 10 states with highest student representation among the Williams College student body.

```
# Percentage of Int'l Students, Top States  
intlperc2015
```

```
## [1] 9.227273
```

```
ordkabdom2015 <- kable(orddom2015[1:10,])  
ordkabdom2015
```

	State	Students
33	New York	413
23	Massachusetts	280
6	California	267
31	New Jersey	127
8	Connecticut	120
11	Florida	83
39	Pennsylvania	72
15	Illinois	67
22	Maryland	64
44	Texas	62