

Ephmetrics

Ben Morton

Abstract The **ephmetrics** package analyzes Williams College Archived Course Catalogs in PDF form and reports specific data. This data includes information about the given year's tuition price, geographical distribution, and size of the undergraduate classes

Introduction

Data

The information in this package retrieves data from the Registrar Office of Williams College Archived Course Catalogs. These catalogs are found in PDF form, and the `pdftools` package was used to read and convert the PDF's into text. Using the 2009-2010 catalog as a test case, functions were developed to capture the relevant information from the given year.

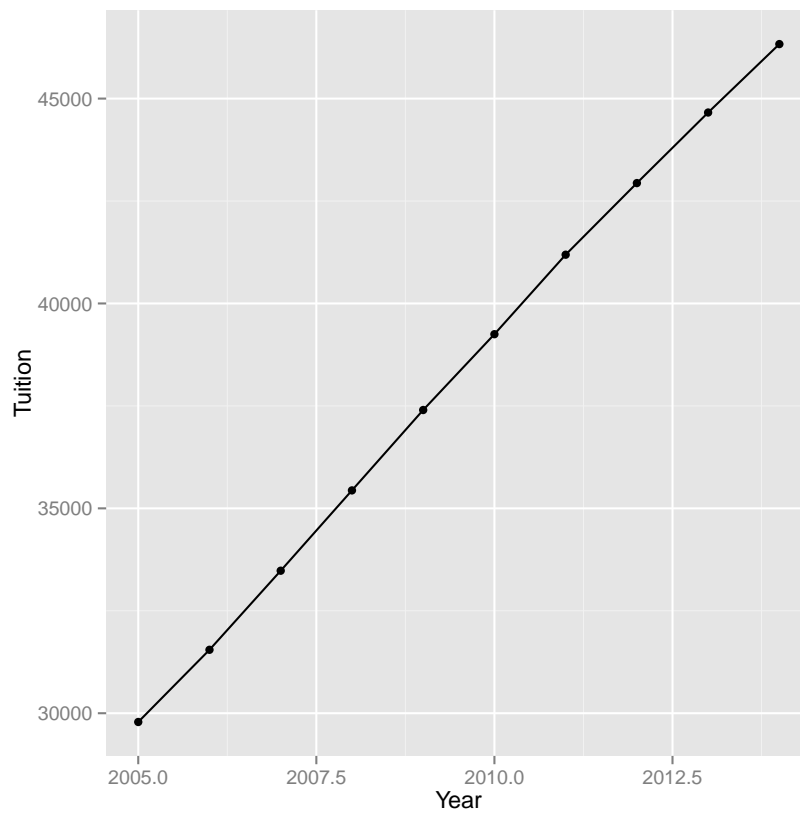
The format of the catalog from each year is not exactly uniform, so the functions require updating based on the assumptions that were made. Examples of this are column layouts, what specific data follows what, and the overall formatting of each year's information.

Examples of Using the Data

```
> price_data <- ephmetrics::get_price_data()
> price_data
```

	Tuition	Year
1	46330	2014
2	44660	2013
3	42938	2012
4	41190	2011
5	39250	2010
6	37400	2009
7	35438	2008
8	33478	2007
9	31548	2006
10	29786	2005

```
> print(ggplot(price_data, aes(x = Year, y = Tuition)) +
+       geom_line() +
+       geom_point())
```



Conclusion

Ben Morton
Mathematics and Computer Science
Williams College
Williamstown, MA, USA
bmm5@williams.edu