Breaking the Bundle: Analyzing Duke’s Journal Subscriptions

Project Leads: Angela Zoss, Jeff Kosokoff
Undergraduates: Cathy Lee, Jennifer Zheng

Introduction

Academic libraries purchase journal subscriptions from publishers such as Elsevier that provide costly bundled deals by grouping a few thousand journals as a package. However, libraries handpicking a few hundred journals can cost even more. Libraries are therefore forced to subscribe to journals it may not want or need.

How can academic libraries optimize journal selection given their budget?

1. Generate a model to improve the data collection and analysis pipeline
2. Create a dashboard to help library strategists visualize and optimize journal selection
3. Provide academic libraries with new data and methods to aid negotiations with publishers
Data Collection

Web Scraping
Extract open access titles and editorial board members from Elsevier

API
Search journals Duke published in and journals Duke cited from Scopus API

Nodes
selector gadget
Base URL + Journal Title (e.g. “Spatial Statistics”)
Base URL + Affiliation ID (e.g. “60008724”) + Title (e.g. “The Lancet”)

Data Merging

Uniform Entry
Text Matching:
- Special characters behave differently in different operating systems
  - Windows:
  - Mac OS:
- R cannot identify certain special characters

Title Modification:
- Important to use one file that has a list of titles as the base list for merging
- Matching with IDs such as ISSN generates cleanest merge, but ISSN is not always available for certain titles

Manual Checks
Checking the data manually leads to the discovery of hidden errors
- Journals lacking data for certain variables
- Variation in titles among identical journals
Identifying similar characteristics among problematic journals can make debugging more efficient
Dashboard

User Input Bar
Created horizontal bar instead of sidebar to eliminate scrolling and to display all subjects at once.

Plot Package
Used plotly instead of ggplot to show info on hover and to select plot regions.

Data Subsets
Began by visualizing selected journals only, then added visualizations for non-selected journals, then created a panel for selected open access journal visualizations.

User-Friendly Tools
Relabeled names of input criteria to be more accurate and added tooltips to define and explain how each criteria was calculated.

Acknowledgments
Special thanks to: David Hansen, Virginia Martin, Emma Heet, Virginia Carden, Ryan Denniston, Sarah Park, and Bethany Greene. We would also like to thank Paul Bendich and the Rhodes Information Initiative at Duke for the opportunity to work on this summer Data+ project.