

Breaking the Bundle: Analyzing Duke's Journal Subscriptions



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.

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

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

Base URL +
Journal Title (e.g.
"Spatial Statistics")

selector gadget

Nodes

data cleaning

author	Title
J. Mateu	Spatial Statistics

API

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

Base URL +
Affiliation ID (e.g.
"60008724") +
Title (e.g. "The Lancet")

API request

JSON files

data cleaning

year_duke	af_id	journal_name
2018	60008724	The Lancet

Data Merging

Uniform Entry

Text Matching:

- Special characters behave differently in different operating systems

Windows:

Title	ISSN
Atención Primaria Práctica	2605-0730

Mac OS:

Title	ISSN
Atenci<f3>n Primaria Pr<e1>ctica	26050730

- R cannot identify certain special characters
- ```
> journal[49]
[1] "Borsa Istanbul Review"
> gsub("İ","i", journal[49], perl = TRUE)
[1] "Borsa Istanbul Review"
```

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

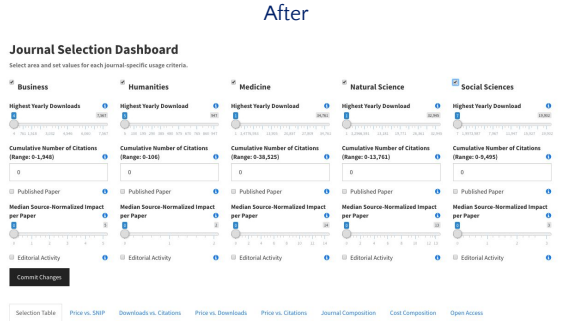
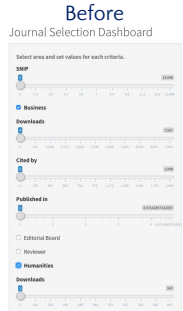
- Journals lacking data for certain variables
- Variation in titles among identical journals

Identifying similar characteristics among problematic journals can make debugging more efficient

# Dashboard

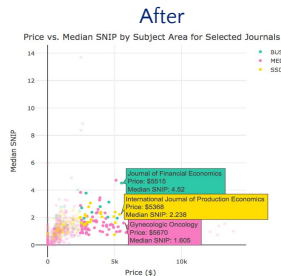
## User Input Bar

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



## User-Friendly Tools

Relabeled names of input criteria to be more accurate and added tooltips to define and explain how each criteria was calculated



## 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

Selected Journals

| Title                                                 | Price   | downloads.max | Cited.total | Published.mean | editors | reviewers | median.snip | subject | Type       |
|-------------------------------------------------------|---------|---------------|-------------|----------------|---------|-----------|-------------|---------|------------|
| 1 Journal of Voice                                    | \$1,080 | 947           | 60          | 1              | false   | false     | 1.3         | HUM     | Freedom    |
| 2 Poetics                                             | \$ 95   | 562           | 106         | 0.71           | false   | false     | 1.4         | HUM     | Subscribed |
| 3 Russian Literature                                  | \$1,672 | 343           | 4           | 0              | false   | false     | 0.1         | HUM     | Freedom    |
| 4 Studies in History and Philosophy of Science Part A | \$1,198 | 313           | 7           | 0.29           | false   | false     | 1.1         | HUM     | Subscribed |

Showing 1 to 4 of 4 entries

Non-Selected Journals

| Title                                   | Price    | downloads.max | Cited.total | Published.mean | editors | reviewers | median.snip | subject | Type       |
|-----------------------------------------|----------|---------------|-------------|----------------|---------|-----------|-------------|---------|------------|
| 1 Accounting, Organizations and Society | \$ 3,316 | 325           | 34          | 0.14           | false   | false     | 2.8         | BUS     | Subscribed |
| 2 Advances in Accounting                | \$ 612   | 43            | 8           | 0              | false   | false     | 0.7         | BUS     | Subscribed |
| 3 Australasian Marketing Journal        | \$ 600   | 57            | 0           | 0              | false   | false     |             | BUS     | Freedom    |
| 4 Business Horizons                     | \$ 704   | 904           | 12          | 0.14           | false   | false     | 2.2         | BUS     | Subscribed |
| 5 Business Research Quarterly           | \$ 0     | 0             | 0           | 0              | false   | false     |             | BUS     | Open       |
| 6 Case Studies on Transport Policy      | \$ 449   | 26            | 0           | 0              | false   | false     | 0.6         | BUS     | Freedom    |

## 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.