

See Also:

The UBC School of Library, Archival and Information Studies Student Journal
2015 - Spring

University Journal Subscription Costs in the UK: An Exploratory Study

Amanda Wanner and Robert Hamaker

amandathinks@gmail.com - rdfhamaker@gmail.com

Keywords: librarians, academic libraries, journal subscription costs, scholarly communications, academic publishing, freedom of information, FOI

Abstract:

This poster project focuses on using visual analytics to understand complex data. For this project, we investigated a unique new dataset, "Journal subscription costs - FOIs to UK universities," published by Stuart Lawson and Ben Meghreblian. They collected data by sending Freedom of Information (FOI) requests to universities in the UK for information on the subscription costs to seven major publishers over a five year period, using the website whatdotheyknow.com. This work is ongoing, with data added as universities respond. The dataset was also chosen due to its CC-BY license, which allowed us the flexibility to remix and build upon the original work. We also supplemented this dataset with publically available UK enrolment data.

Database and journal subscription costs are becoming an increasingly hot issue for academic libraries. Many libraries sign non-disclosure agreements with publishers, which limits their ability to leverage or bargain over subscription prices. Although it is often reported that subscription costs have been on the rise and that libraries are under increasing financial pressure, it is unclear exactly how

**See Also:
2015 - Spring**

much libraries are paying and to whom. What is clear is that these increasing subscription prices have caused many to question the traditional publishing model, and whether the added value from traditional publishing is worth the cost. A recent Nature news feature, for example, discussed the benefits of an Open Access journals model that is able to operate profitably and more efficiently while free access to articles for the public.

Other recent research by Bergstrom et al. has also used FOI requests to find information about journal subscription prices in the United States. When comparing their data with factors such as total university enrolment, number of PhD students, and whether the school was a Research 1 or Research 2 institution, they concluded:

The contracts that we have seen show remarkable institution specific price variations that cannot be explained by university characteristics such as enrolment and PhD production. Some institutions have been quite successful in bargaining for lower prices, whereas others may not have been aware that better bargains can be reached. Perhaps this variation explains publishers' desire to keep contract terms confidential (Bergstrom, Courant, McAfee, & Williams, 2014).

By using a similar dataset, and merging the data with openly available UK university enrolment and geolocation data, this project builds upon research by Bergstrom et al. and, in the process, contributes to a new body of research creating transparency in the library purchasing field.

This research is preliminary in nature. At the time of data cleaning and analysis, the datasets were still incomplete. However, we investigate the following questions:

1. How have subscription costs generally changed over the 5-year period from 2010-2014?

2. How does university enrolment compare with the amount that universities pay major publishers?
3. Does the geospatial location of a university play a role in how much they pay in subscription costs to major publishers?

As the data is not very granular, it is difficult to answer any of these questions with certainty; however, we discovered several patterns that warrant further investigation. Why, for example, does it appear that institutions with average enrolment size pay the most for journal subscriptions, while those with the lowest and highest enrolment appear to pay less? Why are there so many drastic changes in price year to year at many universities? Future research should focus on acquiring more granular data and investigating these inconsistencies.

Copyright:

All authors in see also retain full copyright of their material.

All content in See Also is published under an [Attribution-NonCommercial-NoDerivatives 4.0 license](#).

University Journal Subscription Costs in the UK: An Exploratory Study

R.Hamaker & A.Wanner

Introduction

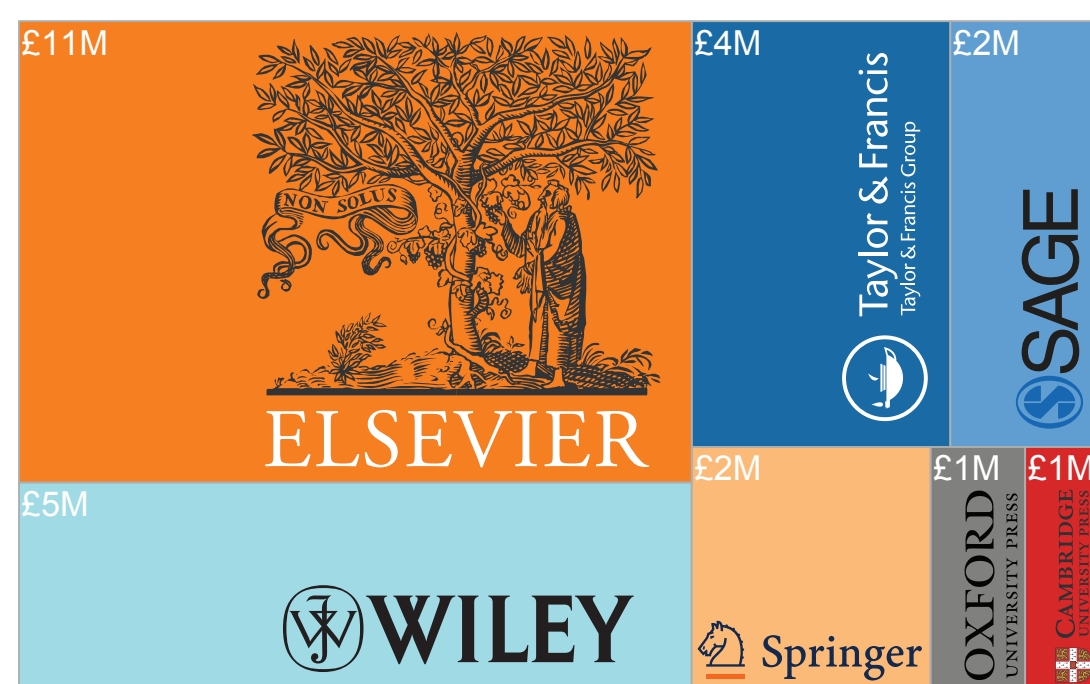
Database and journal subscription costs are a hot issue for academic libraries. Many libraries are forced to sign non-disclosure agreements with publishers, which limits their ability to leverage or bargain with the companies.

Although it is common knowledge that subscription costs have been on the rise and that libraries are under increasing pressure, it has been unclear exactly how much libraries are paying and to whom.

Hypotheses

- H1) Subscription costs scale linearly with other factors, such as the number of enrolled students, tuition prices, or the university's operating budget.
- H2) Subscription costs increase at a rate higher than inflation.
- H3) Universities located in wealthier areas (e.g. London) pay more for subscriptions than those in less wealthy or rural areas.

Visualizing the costs should make it possible to discern patterns with greater ease for large numbers of institutions with higher incomes.



Elsevier Dominates: This treemap aggregates the relative amount paid to the seven publishers in 2014 by 43 UK higher education institutions.

Dataset

For this project, we investigated a unique new dataset, "Journal subscription costs - FOIs to UK universities" by Stuart Lawson and Ben Meghreblian (2014). Lawson and Meghreblian aggregated the data from responses to Freedom of Information (FOI) requests to the universities.

Data consists of 134 universities and up to five years of reported subscription costs for academic journals paid to the publishers Elsevier, Wiley, Springer, Taylor & Francis, Sage, Oxford University Press, and Cambridge University Press. FOI requests are still ongoing, but as of November 2014, 43 universities have complete records.

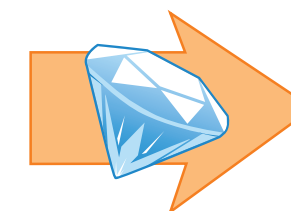
We merged the dataset with openly available UK university enrolment, university income, and geolocation data to contribute to a new body of research creating transparency of subscriptions costs for libraries.

Process

We used Google Refine, Reconcile-csv, and import.io to clean and process the data. Although the datasets were not very large, we encountered several issues. Visualizing was key at each step to confirm the abstract data manipulations were having the desired outcome.

ISSUE

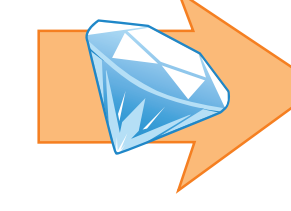
The original dataset is human-readable, but difficult to parse with software.



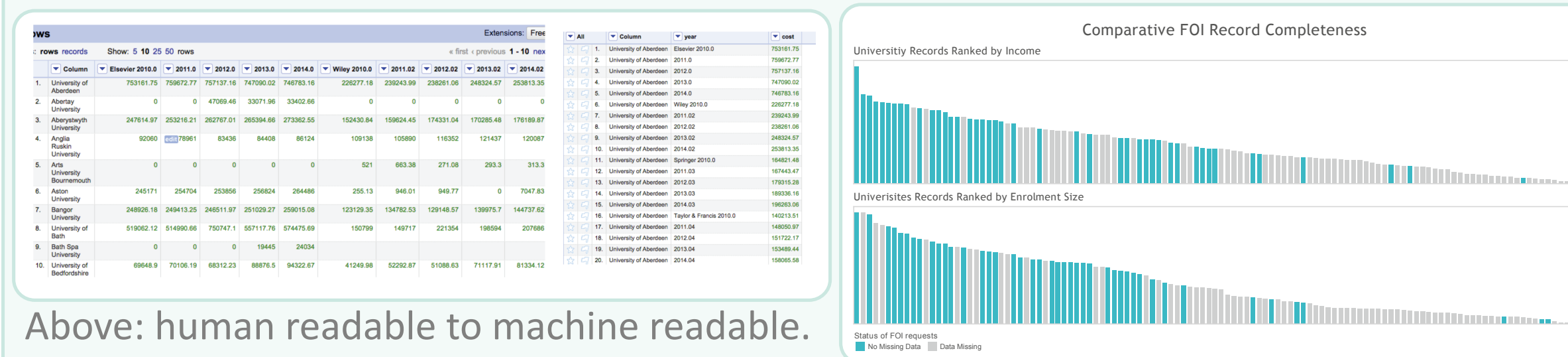
SOLUTION

Simple transformations of the data with Google Refine data cleaning software.

Primary dataset used different identifiers from the demographic data.

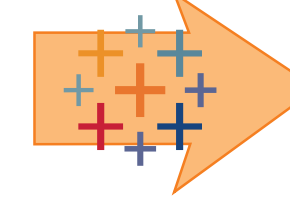


Add unique, common identifiers to the datasets with Reconcile-csv.



Above: human readable to machine readable.

Database still in development with high levels of missing data



Visualize the distribution of missing fields in the dataset to ensure roughly even representation of data.

Complete datasets of UK-targeted geodata were not available, and Tableau could not interpret them.



Web scraping of postcodes from the UK's Higher Education Statistics Agency "contacts" page, using an API based on UK survey data to add latitude and longitude.

Further Research

This research is still at the exploratory stage.

The FOI requests from the primary dataset were sent in August 2014, and the majority of the universities have yet to respond with complete records. Because colleges typically have non-disclosure agreements with major publishers, a few have outright refused to comply. A full statistical analysis is required to ensure that missing data are not biasing our results.

Further contextual information is needed to properly interpret the data.

Insights

While this project is exploratory, there are still several basic take-aways from our visualizations.

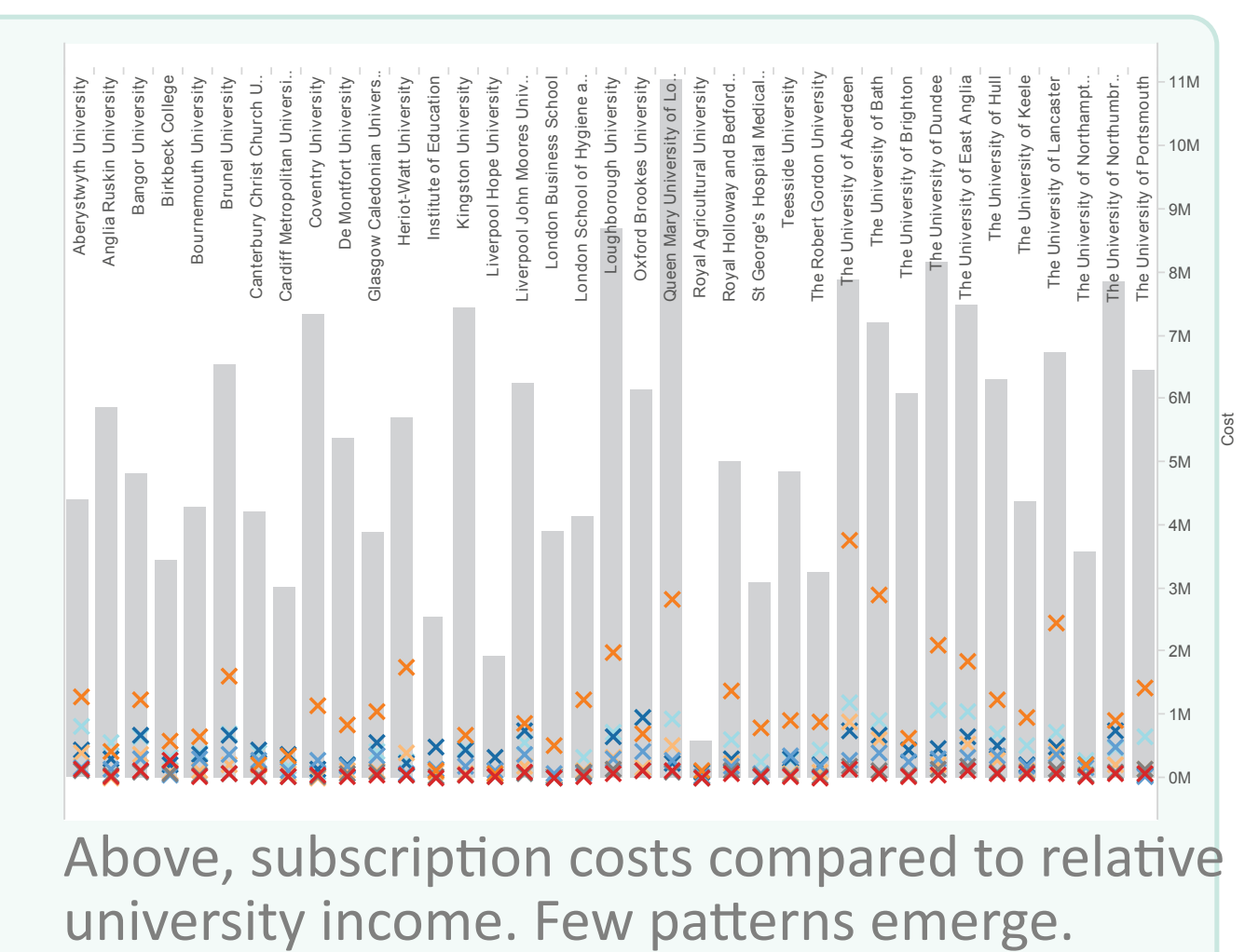
Elsevier's subscription prices are significantly higher than the other major publishers.

When broken down to the university level, there are inconsistencies in the data, such as several unexplained spikes and dips in subscription costs over the five year period, visible in the long chart that flows along the bottom of this poster.

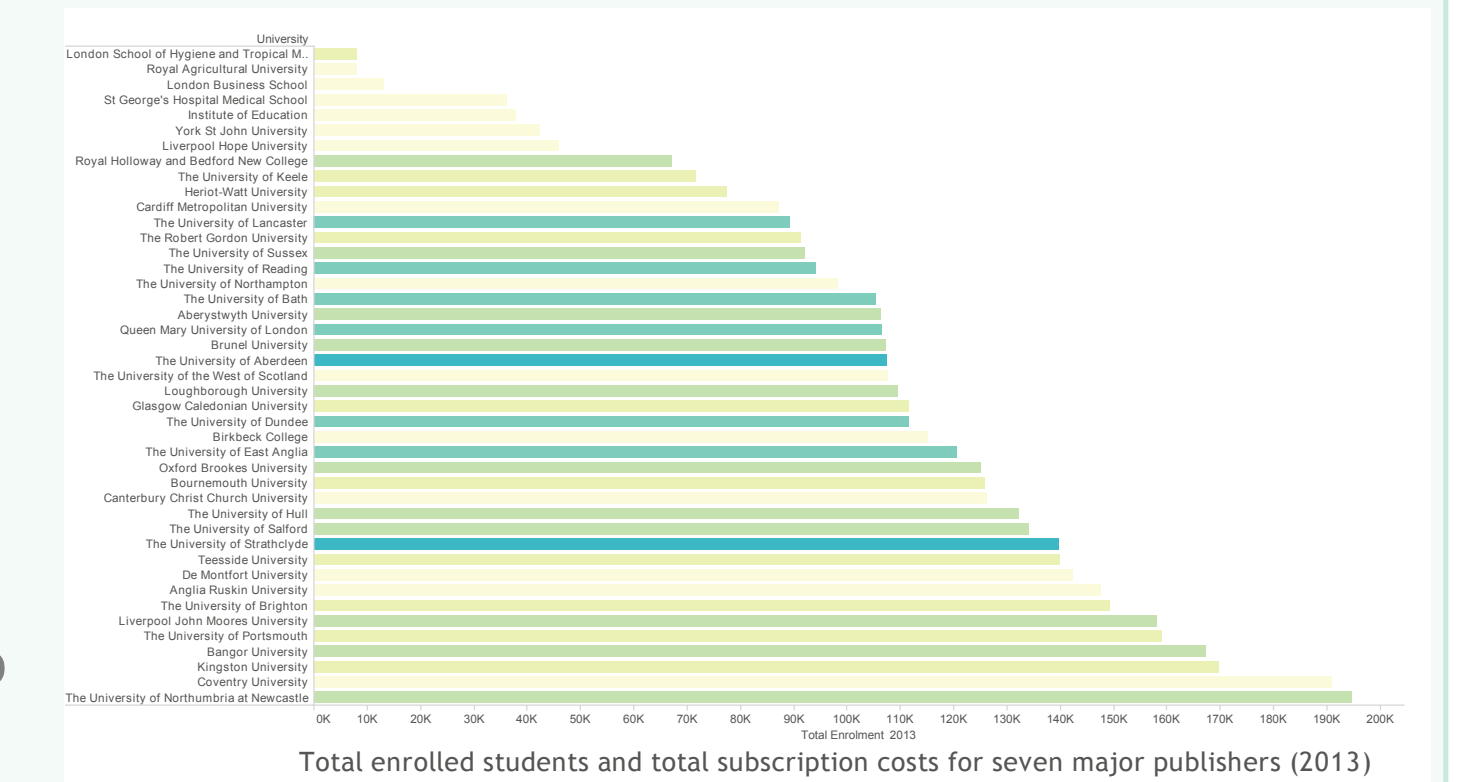
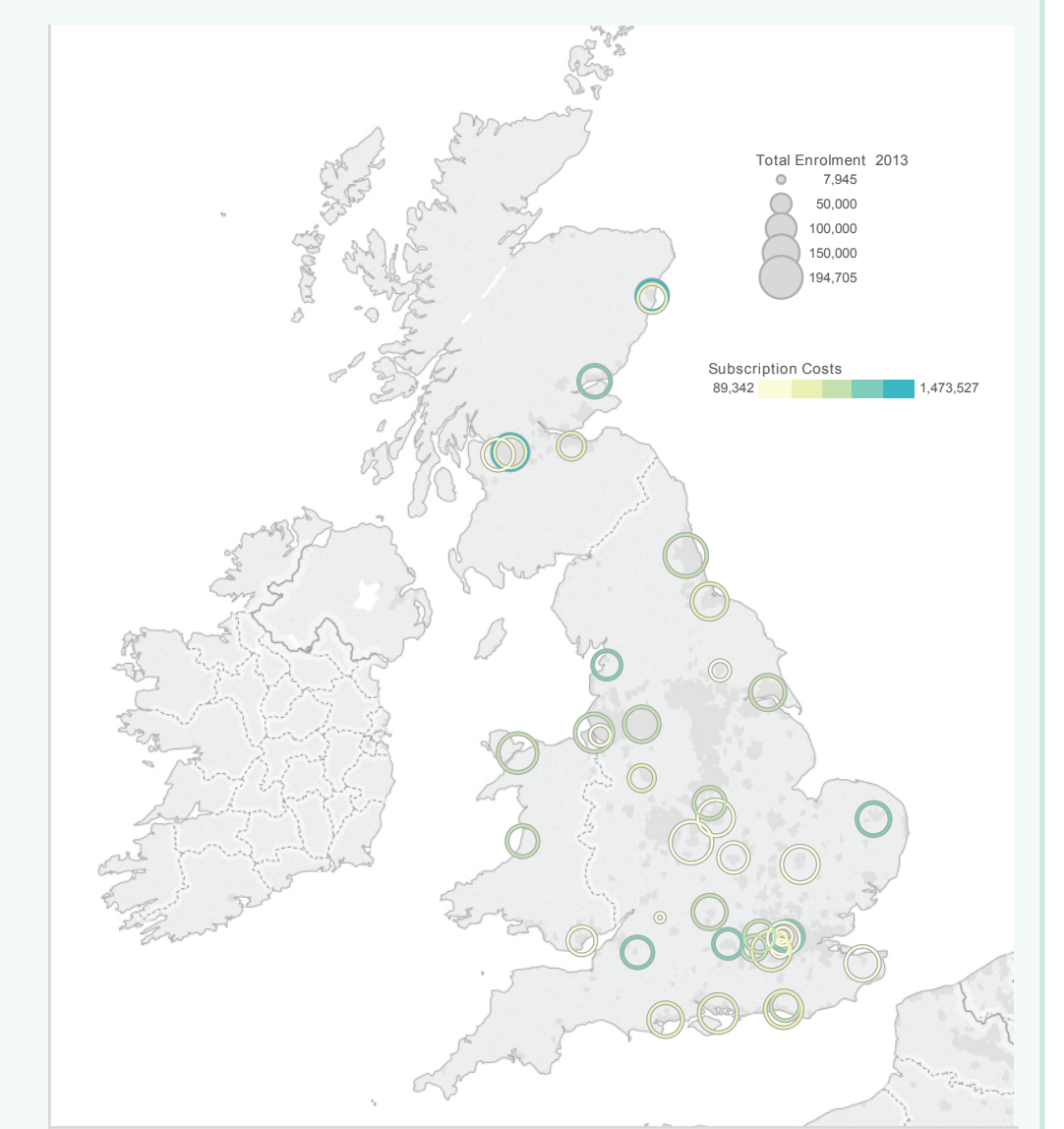
There are also inconsistencies when comparing the university-level data with enrolment information, geographic location, and university income.

Areas shaded darker on the map and bar chart to the right are higher subscription costs.

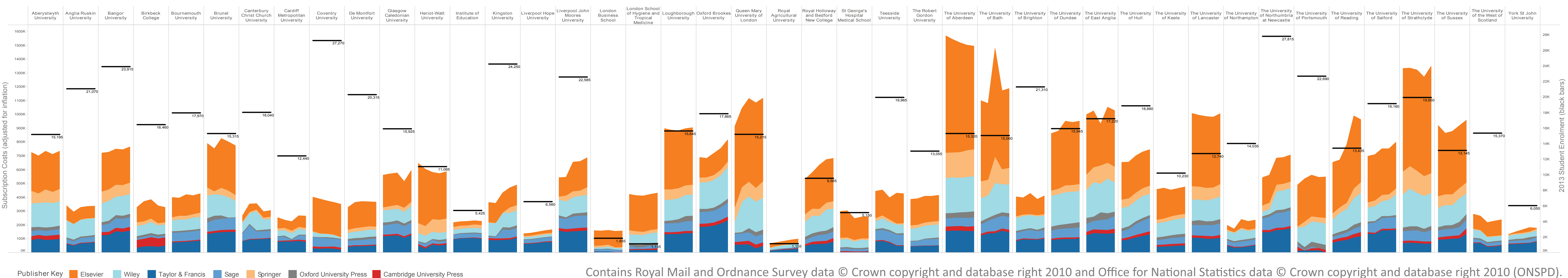
Overall, the data suggests that the highest subscription costs are paid by the Institutions with average enrolment, and that Institutions in close proximity pay widely divergent subscriptions.



Above, subscription costs compared to relative university income. Few patterns emerge.



Below, the subscription costs from 2010-2014 for the 43 institutions included in our analysis, compared to their enrolment in 2013



Publisher Key: Elsevier, Wiley, Taylor & Francis, Sage, Springer, Oxford University Press, Cambridge University Press