Harnessing Alma Analytics & R/Shiny for Insights

Drew Fordham & David Lewis, Curtin University Library, ANZREG 2019

Curtin University

- Over 56,000 students (36,000 EFTSL); 25% of students are international
- 10th largest university in Australia (2017) & 9th largest for overseas students (2017)
- Almost 4,000 equivalent full-time staff; one of the largest single-site employers in WA
- Destination of choice for WA university applicants with 53% domestic market share
- Major campuses in Perth, Kalgoorlie, Singapore, Malaysia, Mauritius and Dubai

Introduction

- In 2017, a self selected group of staff interested in data formed a working group.
- **Aim** - to use data analytics and visualisation to have a measurable impact on library operations and planning.
- A lot of data available for the physical collection via Alma Analytics.
- For further reading IATUL paper presented and available here

Why use R/Shiny

- Functionality of Alma Analytics limited in terms of analysis and visualisation when compared to software such as R/RShiny (free/GNU)
  - *R – programming language*
  - *RStudio – development environment for R*
  - *RShiny – interactive visualisations for R*
  - *RMarkdown and Slidy Presentation*
- There is also Tableau (proprietary / mainly visualisation)
- **Interactive Visualisation tools useful as user determines analysis parameters used.**
- **Visualisation can allow for quick, intuitive insights and understanding of situations.**

Analytics to R

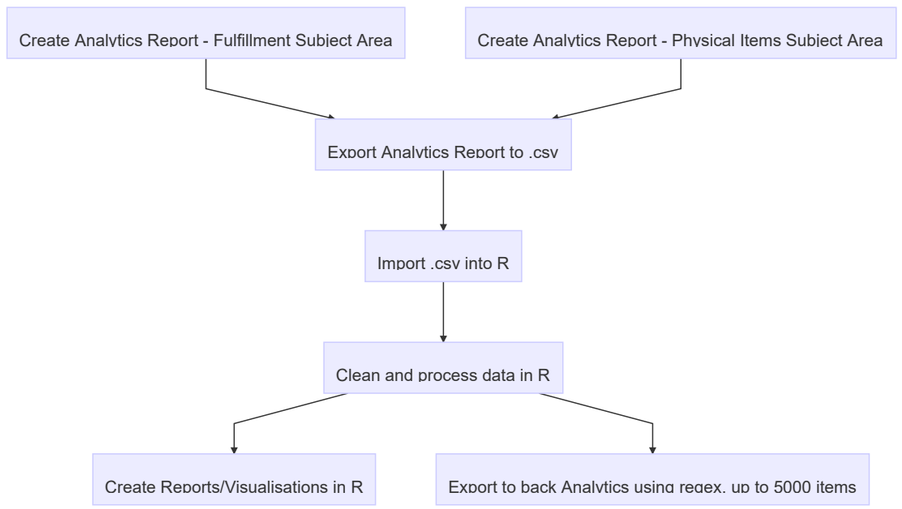

Figure 1. Basic Workflow Alma Analytics to R (and back again)

Early Success

- Existing physical collection de-selection task - multi-copy, zero use.
- Casual staff had been working from printouts of the **whole collection**, applying moderately complicated logic to make deselection decisions. Logic included low use in recent years, last copy available on shelf etc
- The task was
  - *highly manual & paper based*
  - ***lots of errors & seemingly endless***
- Imported Alma Analytics fulfillment and physical item data into R
  - *ran logic over entire collection*
  - *printed targeted pick lists - 15k items rather than 500k items*
  - *exported deselected items back to Analytics for printing official reports purposes*
  - *very few errors and finished in 5 months*

Impacts of Deselection Visualisation

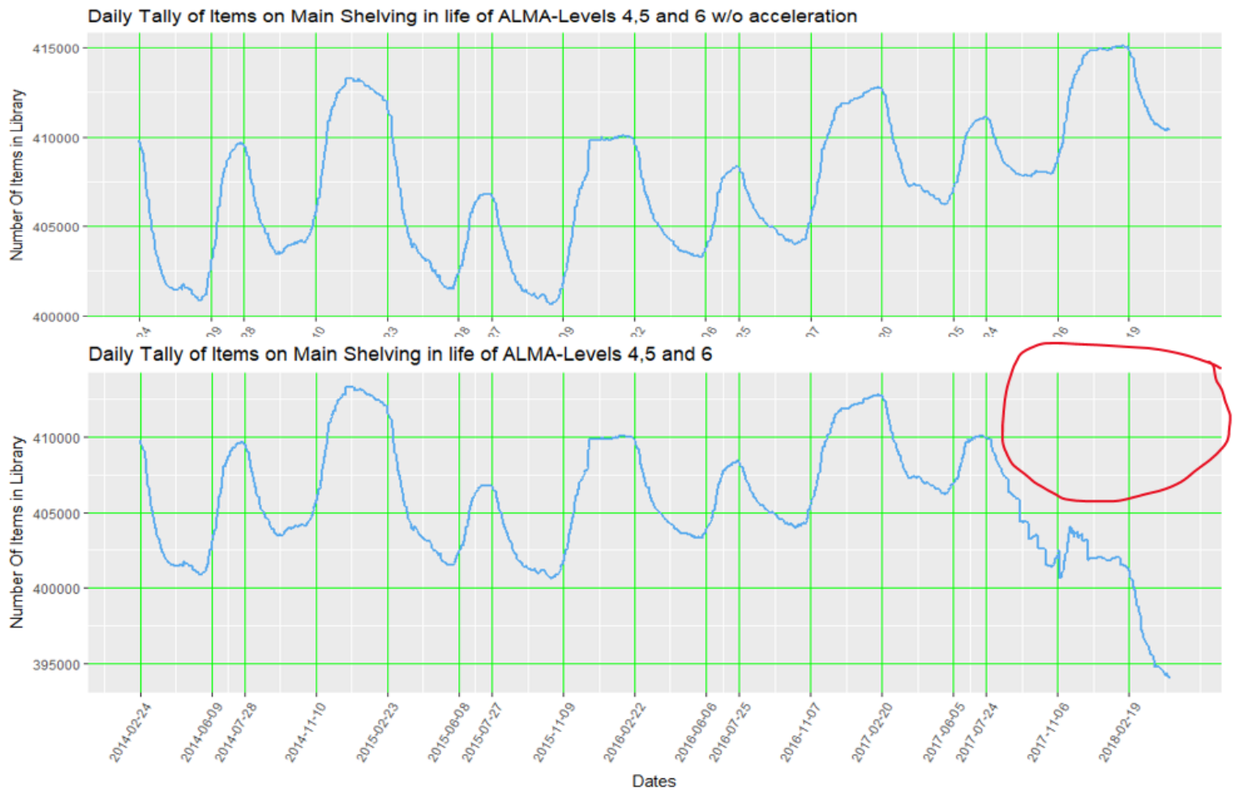

Figure 2. Comparing items on shelves before and after the acceleration of the multi copy, zero use weeding deselection project (Using Alma, R and ggplot).

Other Early Success

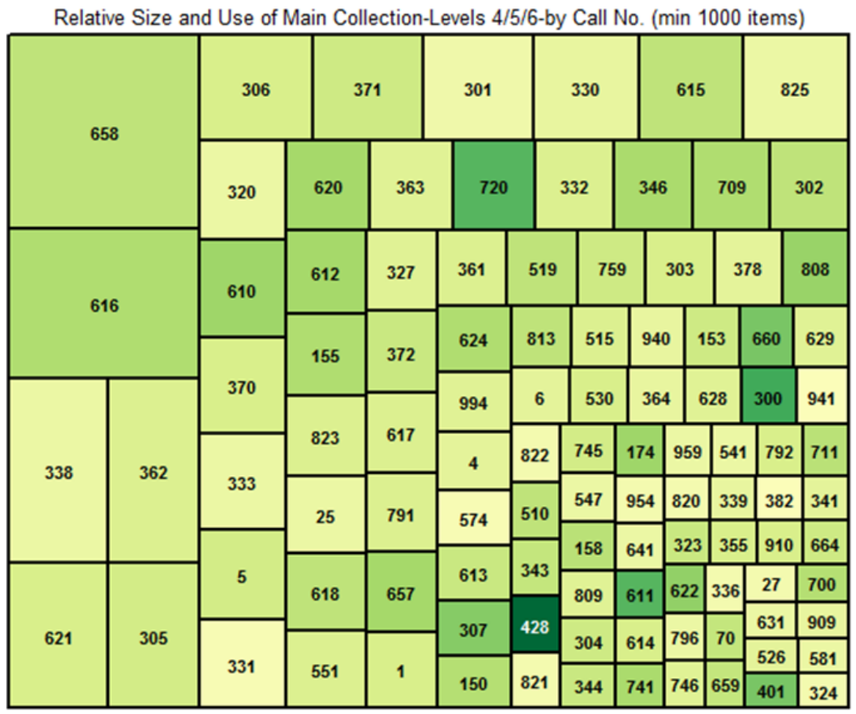

Figure 3. Relative size and use of Curtin Library's Physical Collection, Bentley Campus by Call or Dewey No. (min 1000 items, using Alma, R and treemap).

Refurbishment Modelling

- pressure on shelving space - reduction of physical collection footprint by one third
- moving to large compactus
- initial thoughts
  - *large collection, relatively low use except for small number of high use holdings,*
  - *therefore use closed access compactus with small open access collection*
- what would staff effort look like, moving items between closed and open collections?
- what could data analysis tell us?

The model

- Created Alma Analytics reports on collection use
- An interactive model using RShiny was created based on most recent semester's loans/holds/returns data - as if there was a large low use closed access compactus with small high use open access shelving
- several adjustable modelling parameters
- e.g. previous loan history
  - *steady bias to recent loans over those from many years ago*
  - *only consider loan activity from last two semesters only*
  - *flat, no bias across between loans recent or old*

Refurb Visualisation

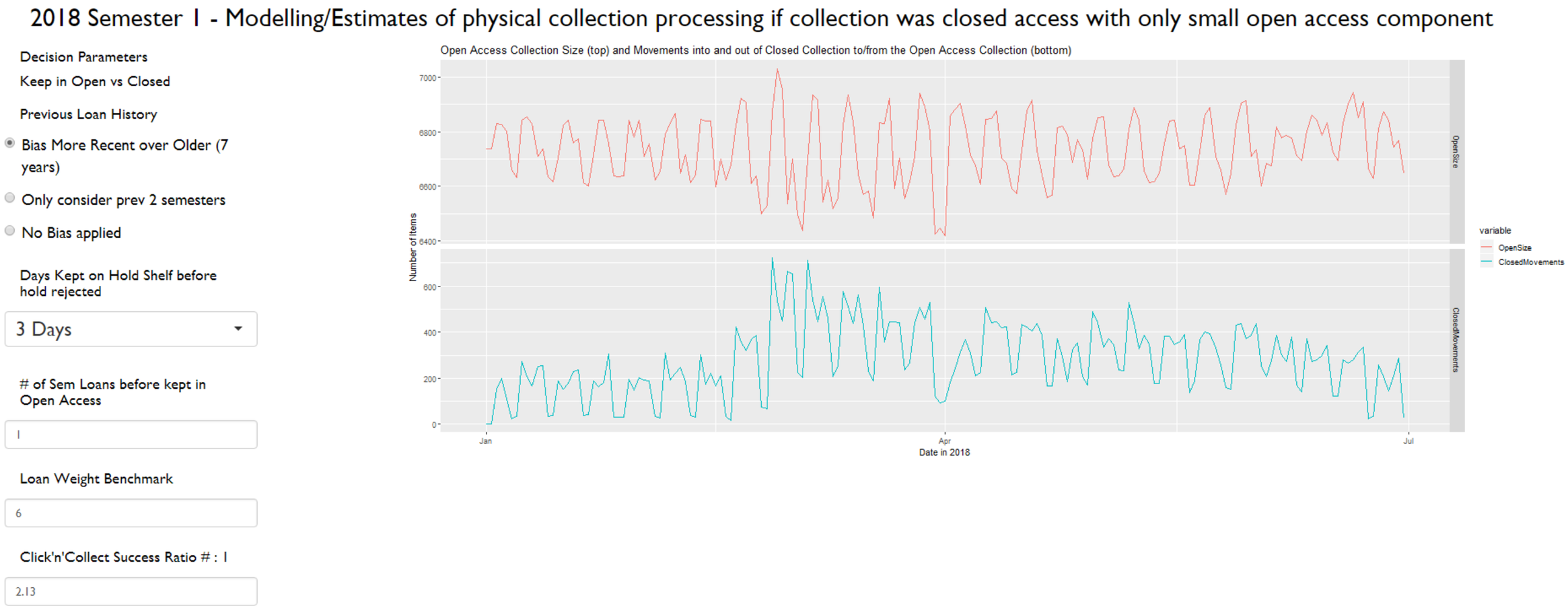

Figure 4. Collection modelling for Robertson Library refurbishment planning (Alma, RShiny).

Refurb modelling Conclusions

- what appears to be a low use collection still requires a lot of staff effort
- **Conclusion** – consider a larger open access collection, compactus or otherwise

Trove API

- Trove is a searchable discovery layer administered by the National Library of Australia and aggregates holdings across all Australian libraries
- Trove API (Application Programming Interface)
  - *publicly available*
  - *Returns data in JSON or XML format – quite different to table/spreadsheet style data*

JSON and XML formats

- Good packages in R to work with both, extracting data into regular table type format.
  - *rjson*
  - *jsonlite*
  - *xml2*
- Trove API JSON not ideally formed, inconsistent structure between single/multiple instances between holdings – adds to complexity and affects speed.
- Trove API XML easier to work with, simpler code and fast (Drew's opinion!)

Loading Data From Trove

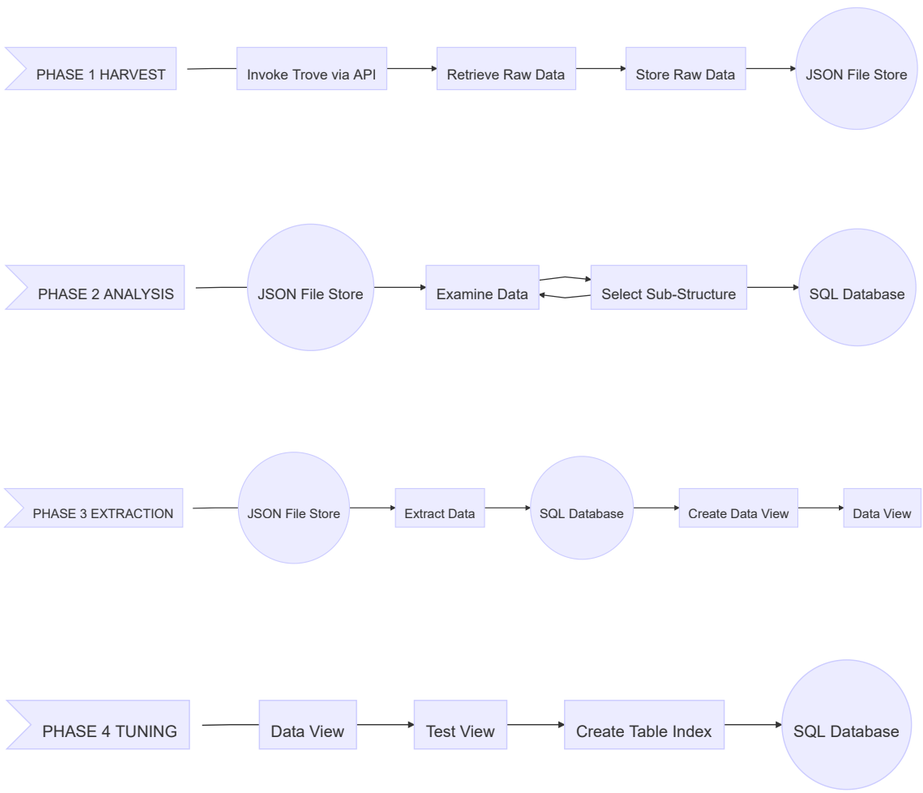

Figure 5. Complex workflow Alma Analytics and Trove API

Alma/Trove API #1

Curtin Holdings in common with other libraries

- Idea is to aid de-selection process.
- Parameters
  - *University Library/Otherwise*
  - *WA Library/Interstate Library*
  - *Not borrowed at Curtin since the year…*
  - *Maximum number of common institutions displayed*
- Of the 300,000 holdings included 90% were held by at least one other library
- Much data cleaning needed to identify common holdings due to inconsistencies across library's catalogs. Trove does well!

Alma/Trove API #1 Visualisation

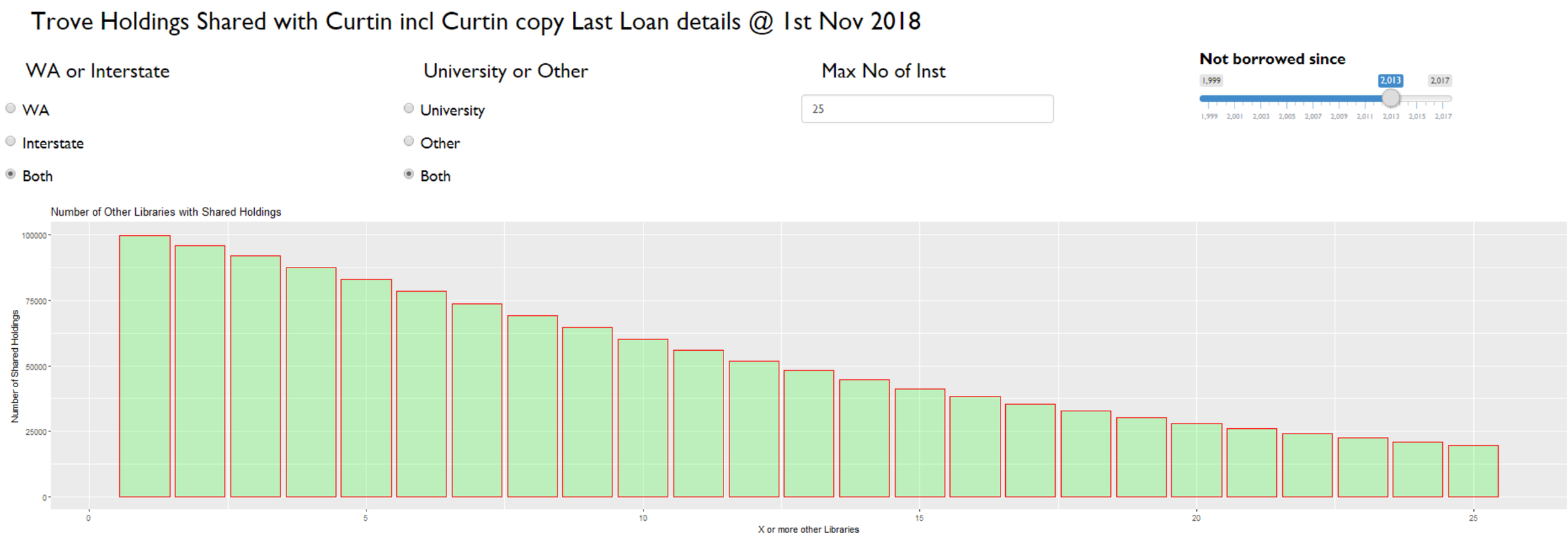

Figure 6. Robertson Library, Curtin's Bentley Campus, holdings in common with other Australian Libraries – Totals (Alma, Trove API and RShiny).

Worked Example - Treemaps

- Initial static treemap shown earlier
- Can be turned into interactive Shiny Presentation
- First create reports in Alma Analytics - Fulfillment and Physical Items

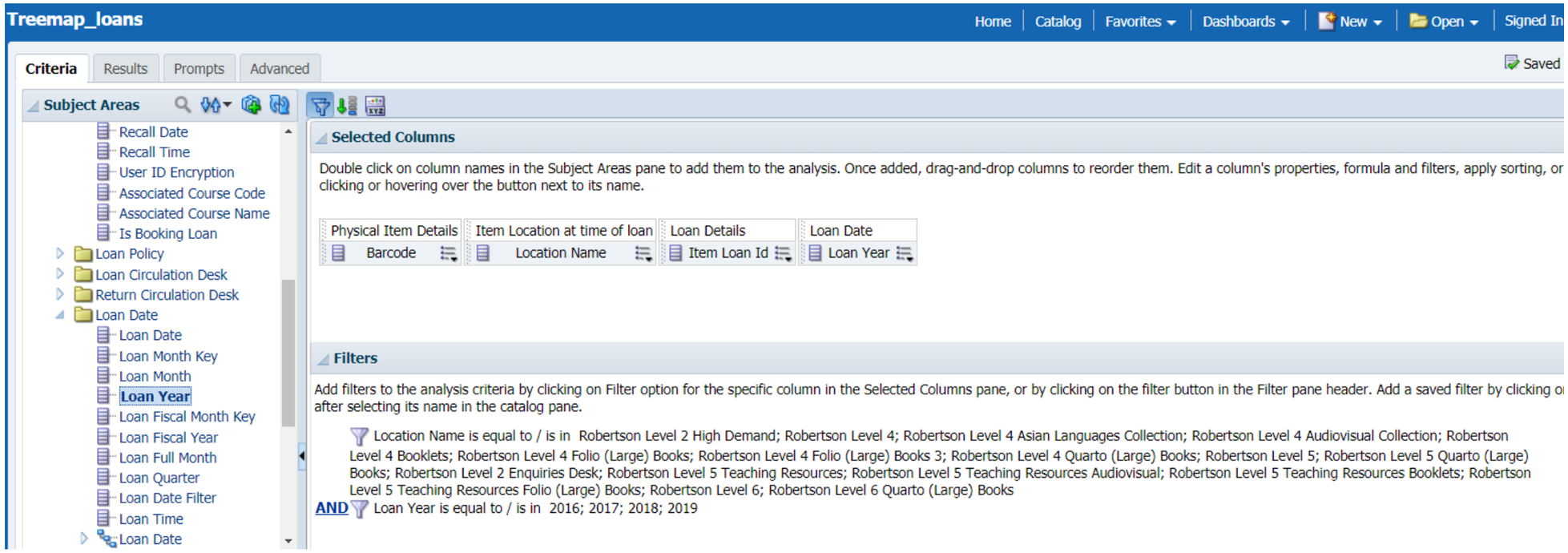

Figure 7. Alma Analytics Fulfillment report for treemap loan data.

- Export from Analytics into R

Worked Example - Treemaps2

- Clean Data

```
> Treemap_physicalitems[str_length(Treemap_physicalitems$CallNumber)==0,]
# A tibble: 88 x 3
   Barcode    `Permanent Call Number` CallNumber
   <chr>      <chr>                   <chr>
 1 DC02947555 Map                     ""
 2 DC03133825 Map                     ""
 3 DC03133874 Map                     ""
 4 DC03134480 TIM                     ""
 5 DC03142654 Unknown                 ""
 6 DC03142666 Unknown                 ""
 7 DC03142678 Unknown                 ""
 8 DC0314268X Unknown                 ""
 9 DC03142691 Unknown                 ""
10 DC03147248 Map                     ""
# ... with 78 more rows
> Treemap_physicalitems[str_length(Treemap_physicalitems$CallNumber)==1,]
# A tibble: 45 x 3
   Barcode    `Permanent Call Number` CallNumber
   <chr>      <chr>                   <chr>
 1 DC08128260 Level 2 Enquiries Desk  2
 2 DC08128300 Level 2 Enquiries Desk  2
 3 DC08128316 Level 2 Enquiries Desk  2
 4 DC08128365 Level 2 Enquiries Desk  2
 5 DC08128366 Level 2 Enquiries Desk  2
 6 DC08128383 Level 2 Enquiries Desk  2
 7 DC08128384 Level 2 Enquiries Desk  2
 8 DC08128385 Level 2 Enquiries Desk  2
 9 DC08128386 Level 2 Enquiries Desk  2
10 DC08154878 Level 2 Enquiries Desk  2
# ... with 35 more rows
> Treemap_physicalitems[str_length(Treemap_physicalitems$CallNumber)==2,]
# A tibble: 52 x 3
   Barcode    `Permanent Call Number` CallNumber
   <chr>      <chr>                   <chr>
 1 DC00537712 36 COM                  36
 2 DC00604443 57 FAL                  57
 3 DC00715912 54 CAH                  54
 4 DC0073310X 54 DET                  54
 5 DC00733135 54 DIA                  54
 6 DC00733172 54 EAS                  54
 7 DC00750922 51 SEQ                  51
 8 DC00758599 51 PET                  51
 9 DC00762046 53 GRE                  53
10 DC00762363 53 GRE                  53
#     with 42 more rows
```

Figure 8. Data cleaning in R.

- Create RShiny Visualisation, which manipulates data based on user input

Treemap Visualisation

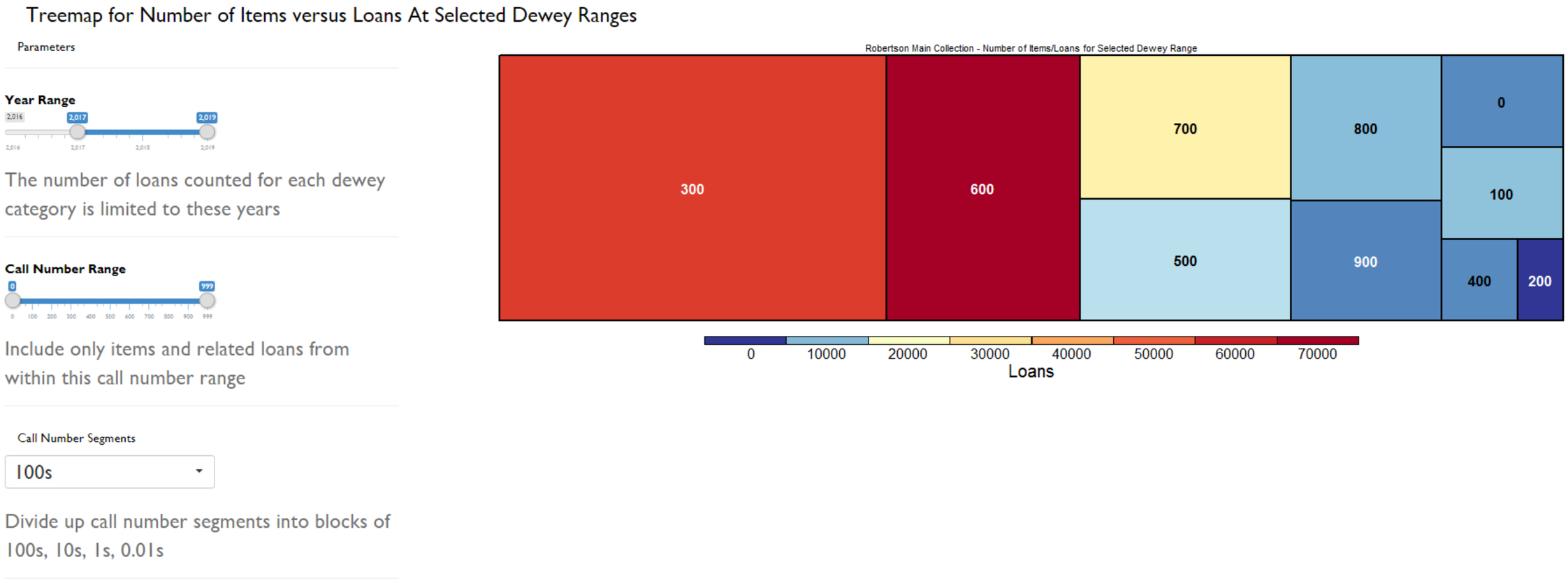

Figure 9. Treemap Visualisation of Dewey Numbers Categories (Alma and RShiny)

Conclusion

- The group did have impact through the tools used, and analysis/visualisations created
- Expanded professional networks
- With the passing of each day, the group saw more and more need for data related skills and more and more opportunities to use data

Trove API #2 – Japanese Gardens

- Idea is to show all Australian library holdings for a particular subject as a network, showing each library's holdings for a subject in context with each others.
- "Gardens, Japanese" chosen as subject
- Used the Trove API, and RShiny with igraph/visNetwork packages
- Nodes represent libraries. Hover to identify library and total holdings. Relative size of node also represents number of holdings.
- Interestingly Trove API library information doesn't always include library name (including Curtin!)
- Edges represent common holdings. Hover to give total common holdings. Relative size of edge also represents number of common holdings.
- Also select minimum number of holdings for a library to be displayed.
- Circular arrangement works best.
- Key Lesson Here - Keep libraries up to date

Trove API #2 – Visualisation

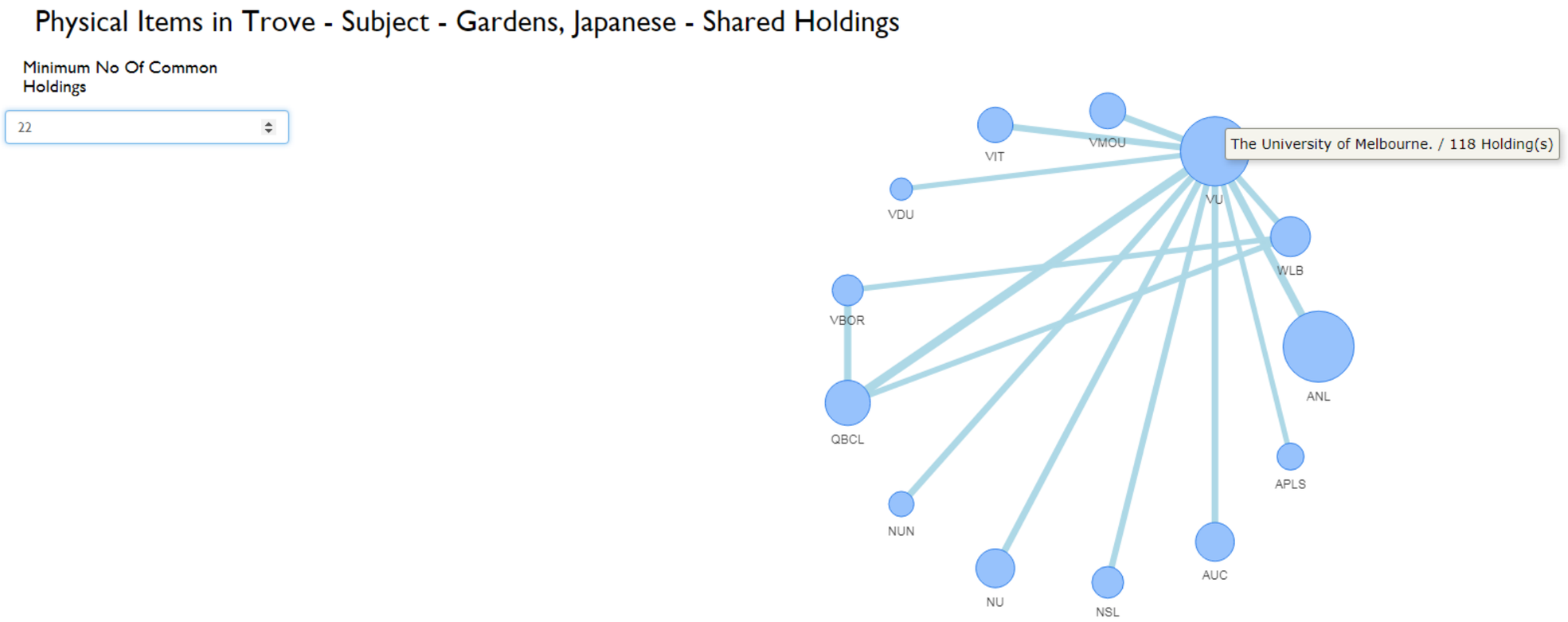

Figure 10. Shared Library Holdings in Trove – Subject : Gardens, Japanese, Minimum common holdings : 22 (Alma, Trove API and RShiny).