Katherine Noack

University of South Australia

Hansard Data Mining Dashboard User Guide

Contents

[1 Introduction 2](#_Toc23669788)

[2 Summary Dashboard 3](#_Toc23669789)

[3 Record Search Dashboard 4](#_Toc23669790)

[3.1 Regular Expressions for Record Search 6](#_Toc23669791)

[4 Audit Team Dashboard 6](#_Toc23669792)

[5 Subject Summary Dashboard 7](#_Toc23669793)

[6 Client Summary Dashboard 8](#_Toc23669794)

# Introduction

This user guide briefly describes the features of five Tableau dashboards developed in the Hansard Data Mining Project.

The purpose of these dashboards is to allow auditors of the Auditor-General’s Department (AGD) to analyse Hansard records as part of their audit process. These dashboards were created with University of South Australia masters students (Semester 2 2019) in collaboration with the project owners James Baker and Aaron Steicke, the project sponsor Assistant Auditor-General Andrew Corrigan and a Project Reference Group consisting of members from three audit teams.

Dashboard functionality requested by the Project Reference Group included:

* Free text search of Hansard records, including being able to search for multiple terms simultaneously in a record
* Automated searching of key terms identified by three audit teams (Performance Audit Team, IT Audit Team, and Local Government Audit Team)
* Provide link to Hansard record on website
* Be able to identify relevant records without having to read the entire record text
* Search for records within specified date ranges
* Sentiment of discussion in a record

These dashboards are currently restricted to showing records of the proceeding types: Answers to Questions, Bills and Question Time.

# Summary Dashboard

The Summary Dashboard gives an overview of all Hansard records available in the project database. Metrics are displayed on the left side of the dashboard and measure:

* Number of Records
* Unique Subjects Discussed
* Most Popular Subject (and its number of mentions)
* Most Discussed Bill (and its number of mentions)

A Bill name can appear in both the most discussed subject and most discussed Bill metrics as a Bill can be recorded in metadata as the subject of a record as well as the Bill discussed. However, because these metrics search different metadata, they can display different values. For example, for Bills the “Education and Children’s Services Bill” may be mentioned in 36 records as the subject, but only 35 records as the Bill being discussed.

Several charts are displayed in the centre of the dashboard depending on the selection made using the “Select View” drop down menu. The bar charts and heat maps that can be displayed include:

* Top 50 Subjects (by number of Hansard Records and Days Discussed)
* Top 50 Bills (by number of Hansard Records)
* Top 50 Portfolios (by number of Hansard Records)
* Top Clients (by number of Hansard Records)
* Subjects Discussed over Time

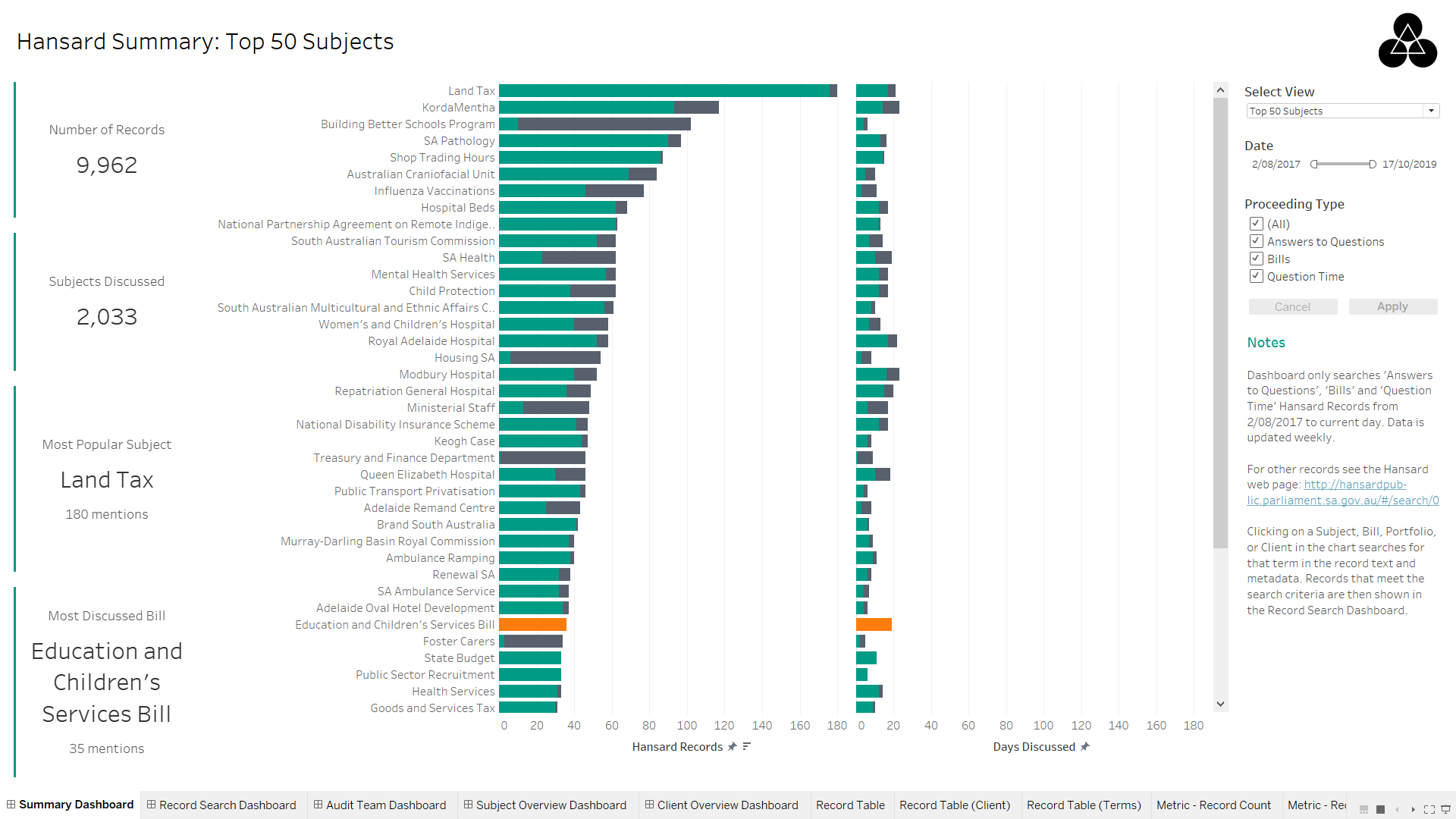


Figure 1: Hansard Summary Dashboard

Records can be flitered by:

* Date
* Proceeding Type (Answers to Questions, Bills and Question Time)

Functionality:

* Change chart in centre of dashboard by “Select View” drop down menu
* Clicking on subject, client, bill, or portfolio in charts searches for that term in the Hansard record text and metadata. Records that meet the search criteria are then shown in the Record Search Dashboard.
* Clicking Most Popular Subject or Most Discussed Bill metrics searches for that subject or bill in the Hansard record text and metadata. Records that meet the search criteria are then shown in the Record Search Dashboard.

# Record Search Dashboard

The Record Search Dashboard displays the subject, proceeding type, date and text summary of each record in a table that matches the search term specified in the top-right corner of the dashboard.

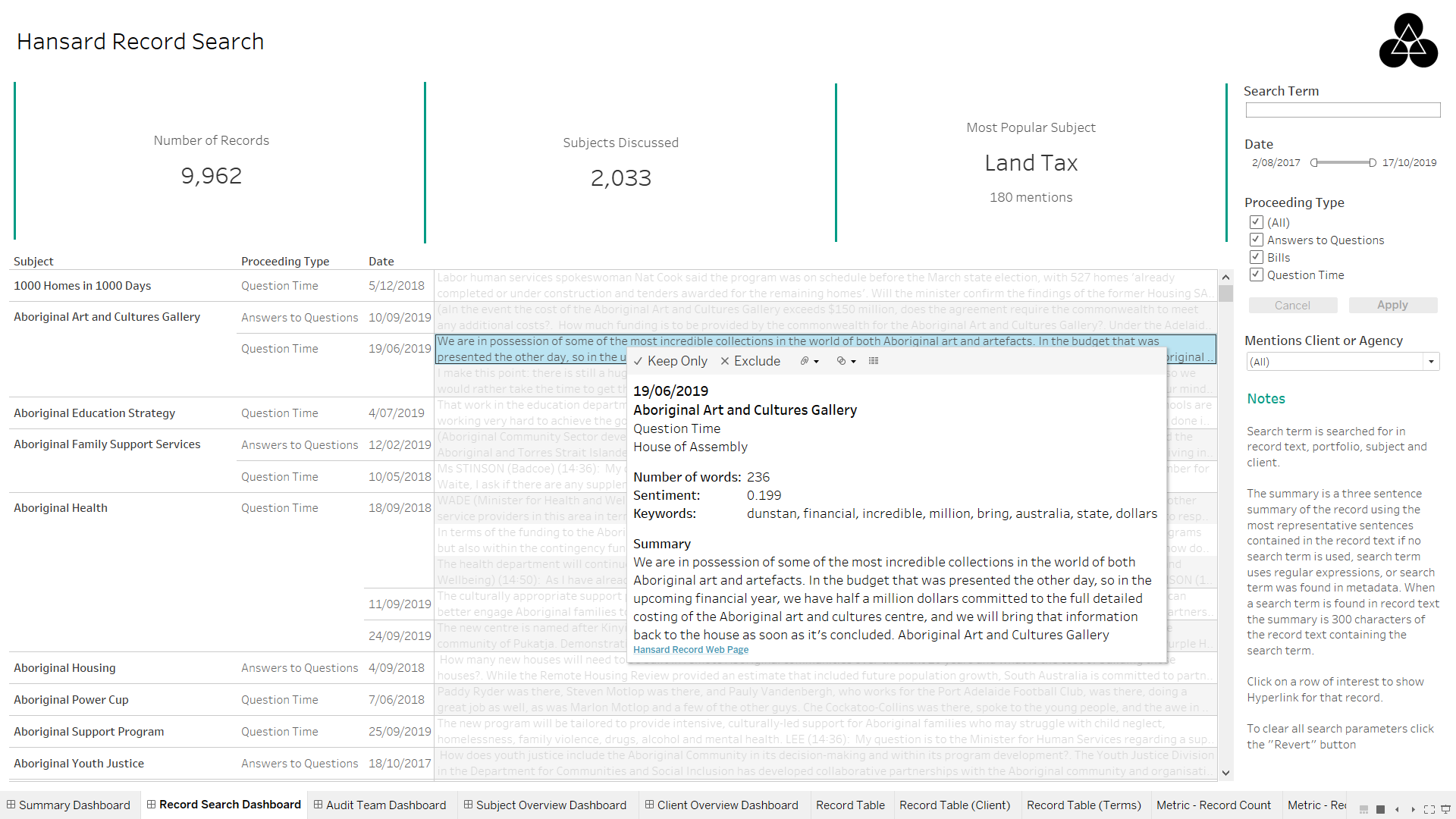


Figure 2: Record Search Dashboard

Records can be filtered by:

* Date
* Proceeding Type
* Client or agency mentioned in record text (one or more selections possible)

Functionality:

* Searches for search term in record text and metadata such as subject, Bill, Portfolio Name (e.g. Minister of Health) and Client and returns matching Hansard records. Search does not use capitalisation when finding a match.
* Can use regular expressions in search term for more precise searching such as searching for multiple terms anywhere in the document (see section “Regular Expressions for Record Search” for more information)
* Hovering over text of a record provides additional information, such as key words and sentiment
* Clicking on a Hansard record in the table displays the Hyperlink to that record

Tooltip displays:

* Number of words in record
* Sentiment of record
* Most ranked key words of the record text (maximum of 10 key words displayed)
* Record summary. Two different summary options are displayed depending on search term.
  1. 300 words contain first match of search term if search term is used without regular expressions
  2. Three-sentence summary using most representative sentences of the record. This is shown only when no search term is used, search term with regular expressions is used, or match to search term is in metadata rather than record text.

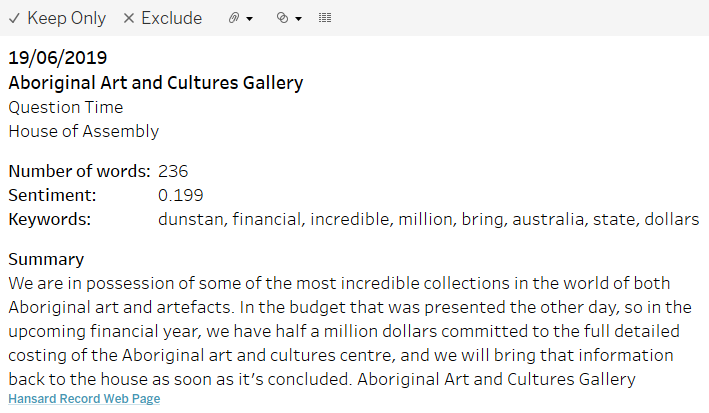


Figure 3: Hansard Record Tooltip (when clicked on text summary)

## Regular Expressions for Record Search

Regular expressions can be used in the search term for more advanced searching of Hansard records. Shown below are some examples of how regular expressions can be used in the search term.

|  |  |  |
| --- | --- | --- |
| Scenario | Term(s) | Regular Expression |
| Search for term exactly as it appears | Auditor-General | Auditor-General |
| Search for multiple terms with any text separating them | Auditor  General | Auditor.\*General |
| Search for multiple terms anywhere in the record text and in any order. This will match terms as a whole word and therefore Auditor term will not match Auditor’s | Auditor  General  Department | (?=.\*\bAuditor\b)(?=.\*\b General\b)(?=.\*\b Department \b) |

Table 1: Regular Expression Example for Search Terms

For additional information of regular expressions see: [Quick-Start: Regex Cheat Sheet](https://www.rexegg.com/regex-quickstart.html)

# Audit Team Dashboard

The Audit Team Dashboard displays the number of records key terms have been mentioned for a given audit team. The key terms displayed in the chart were provided in a spreadsheet which can be updated with new terms and additional audit teams. See Installation and Deployment Guide on how to do this.

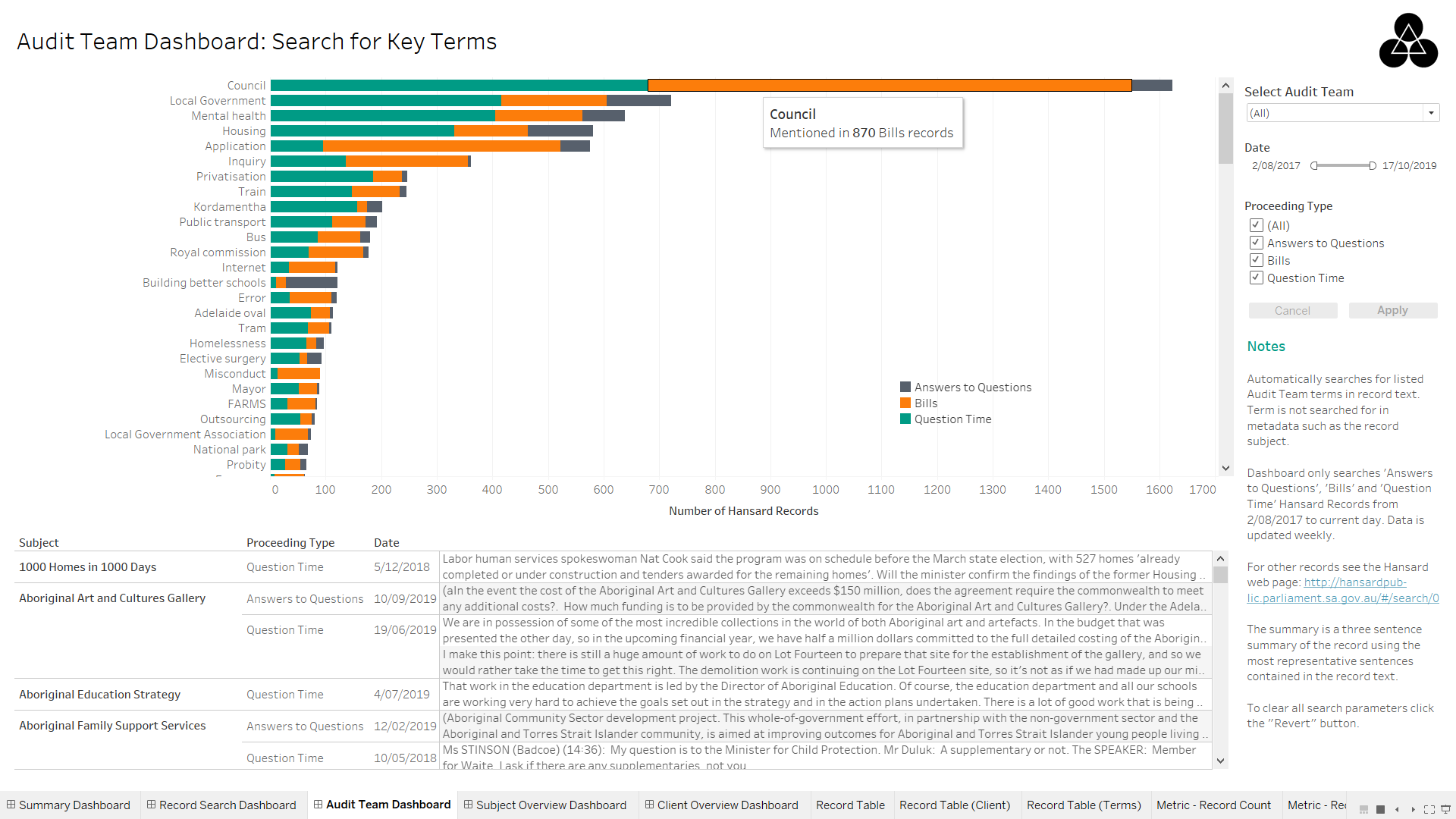


Figure 4: Audit Team Dashboard

Records can be filtered by:

* Date
* Proceeding Type

Functionality:

* Change audit team terms displayed in chart by changing “Select Audit Team” drop-down menu
* Clicking on a term in the bar chart filters the table below to show only records that contained that term (as a whole word) in the record text. These terms are not searched for in metadata such as subject

# Subject Summary Dashboard

The Subject Summary Dashboard gives an overview of the subjects discussed in Hansard records. The treemap in the centre of the dashboard displays the number of records per Proceeding Type for a selected subject or all subjects. Clicking on a Proceeding Type drills down to show the Portfolio Names that participated in that subject. Click twice to return to the original view.

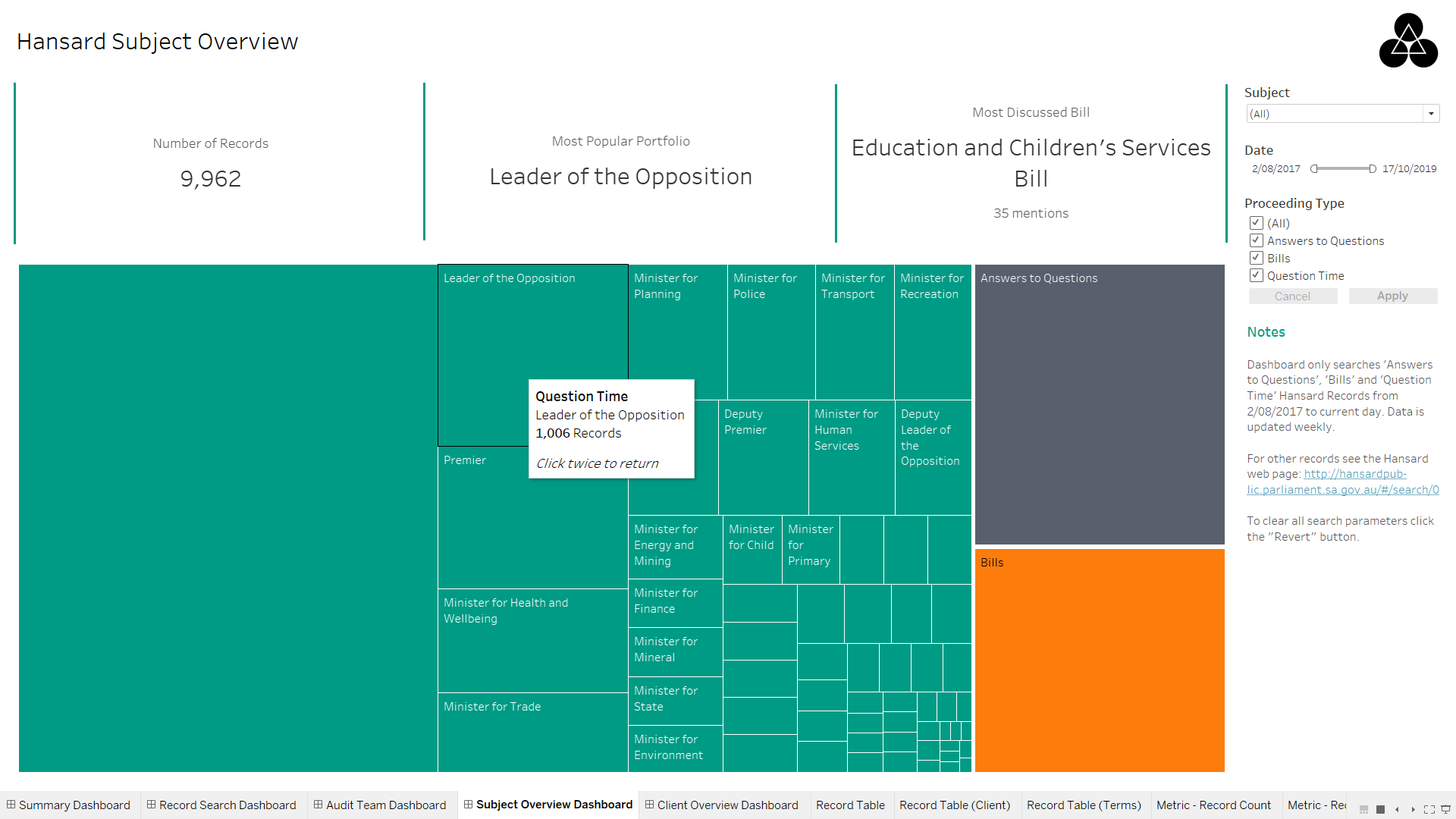


Figure 5: Subject Overview Dashboard

Records can be filtered by:

* Subject
* Date
* Proceeding Type

# Client Summary Dashboard

The Client Summary Dashboard gives an overview of AGD Clients. Initially the number of records per Client Type is shown in the treemap and clicking on a type will display the individual clients. Clicking on a specific client will filter the table shown below the treemap to display records that mention that selected client. Clicking on the select client again will return to the original treemap view.

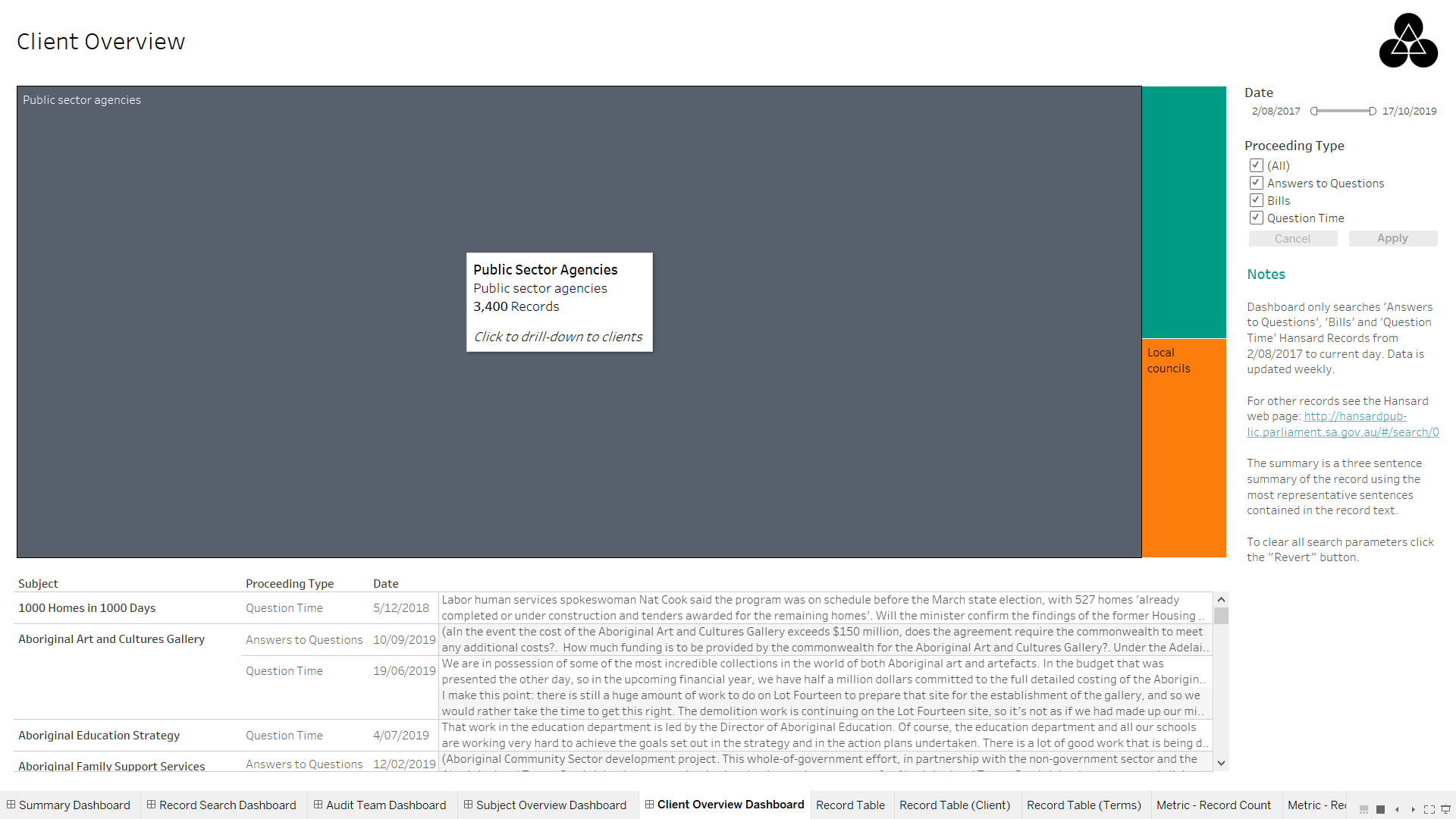


Figure : Client Summary Dashboard (original view)

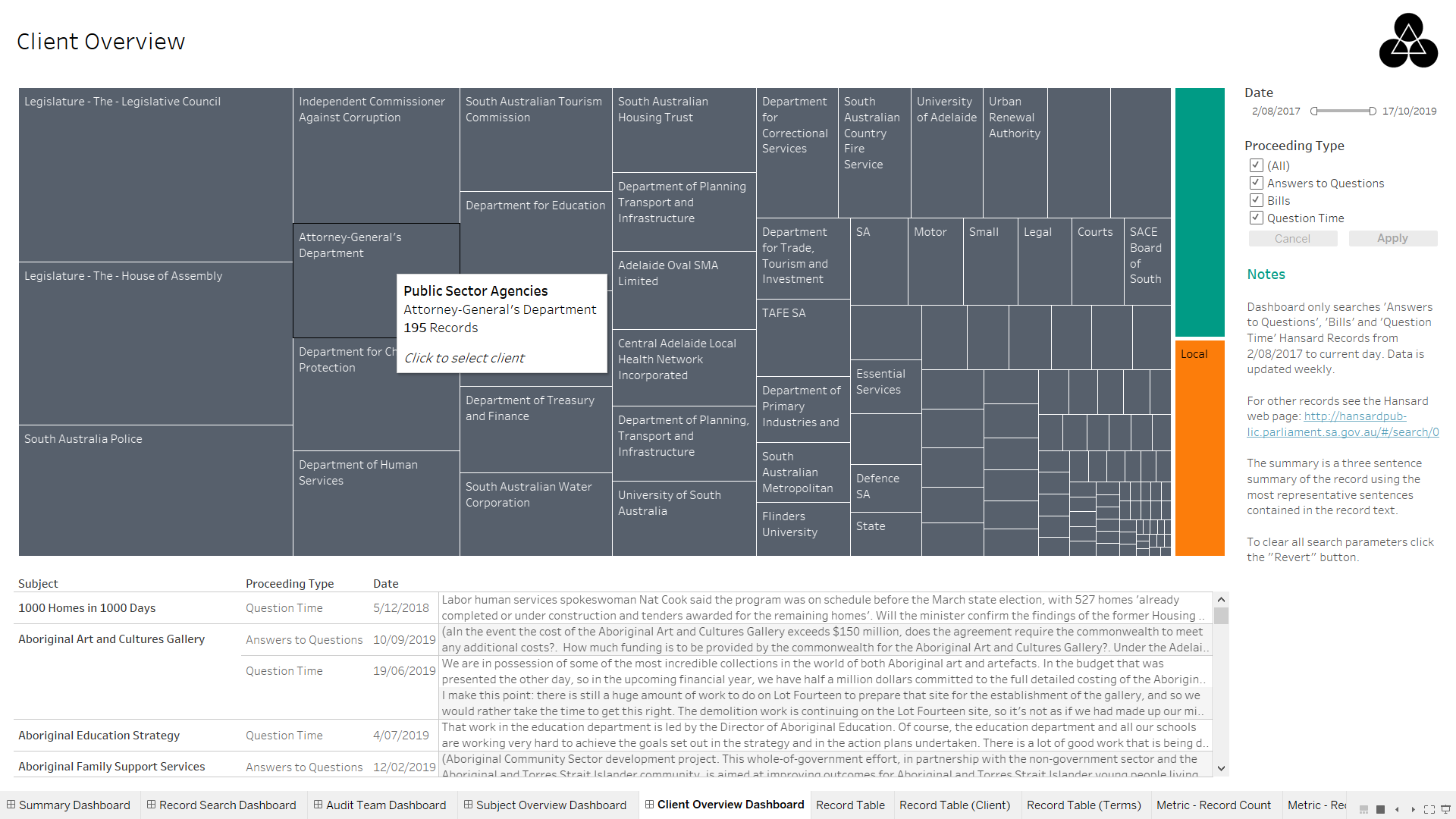


Figure 7: Client Summary Dashboard after Public Sector Agencies clicked in treemap

Records can be filtered by:

* Date
* Proceeding Type