Generalized Web-Based Data Analysis Tool for Policy Agendas Data

Paul Wolfgang
Temple University

ABSTRACT

The Policy Agendas web site includes a data analysis tool that permits selection of the data from various datasets. Associated with each dataset there may be filters. The Pennsylvania Policy Agendas Database project has similar datasets to those in the national database, but the structure of the database tables and available filters are different. This paper describes the design of a generalized web-based data analysis tool that can be configured to work with different datasets and different filters. The tool is table driven so to add a new dataset one merely adds data to the tables that describe the dataset and the filters. The policy codes and their description are also table based to accommodate the variations of the various agenda projects. The tool is developed using the Java language and Java Server Pages, and can be easily modified to accommodate different kinds of datasets and filters.

Introduction

This paper describes the design of a web site that can be used to display and analyze the data collected by the agendas projects such as the United States Policy Agenda and the Pennsylvania Policy Database. The web site is both configurable and extensible. The datasets are described by a table in the database that defines information about each dataset and the filters associated with that dataset. Each kind of dataset and filter is defined by a Java class. If a dataset or filter that does not fit the currently defined datasets or filters, then a new Java class can be added to accommodate the dataset or filter.

This paper begins by giving the background describing the Policy Agendas data analysis tool and the initial adaptation to the Pennsylvania Policy Database. Next the two configuration tables: Tables, and Filters, are described to show how they can be used to define the different datasets. Then the detailed design of the website is given. Finally a description of how a new kind of dataset or filter can be implemented.

Background

The Policy Agendas web site includes a data analysis tool that permits selection of the data from various datasets. The Pennsylvania Policy Database Project mirrors the national project and provides corresponding data about Pennsylvania. The Pennsylvania project wants to provide the same or similar data analysis tools as the national project. The national project graciously gave us the complete source of their web site and database. The data analysis tool was written using Microsoft's Visual Basic Script

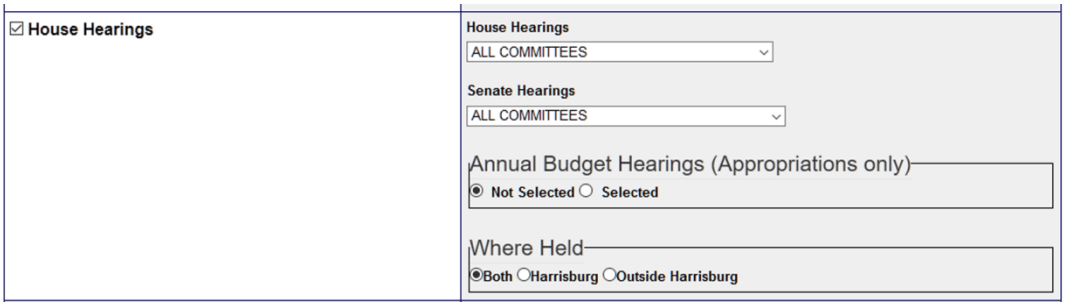
language within Active Server Pages. The original code was designed specifically for the national project, and there was no documentation. I was able to modify this code to work with the Pennsylvania data, but the results were not totally satisfactory. Whenever I made a small change or added a feature, there was a danger that I would break something. Therefore, I decided to re-write the application using Java and Java Server Pages. I choose Java because it is an object-oriented strongly typed language, and the re-design takes advantage of these features. The goal was to maintain the same look-and-feel as the original.

Analysis Page

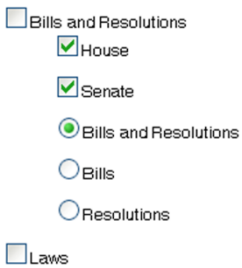
The analysis page presents a form that contains three sections: the top selection allows the selection of the dataset, the middle section the selection of the policy areas, and the bottom selects the range and output format. For each data set there are two columns: the left column gives the name of the dataset and a check-box to select it. The right column lists the filters available for that dataset. Figure 1 shows a portion of the analysis page for the Pennsylvania Policy Database Project.

Dataset Selection

The data select section consists of two columns: one to select the data set and the other to select the filters. Most datasets are selected by a single check-box. When selected the filters are displayed. For example

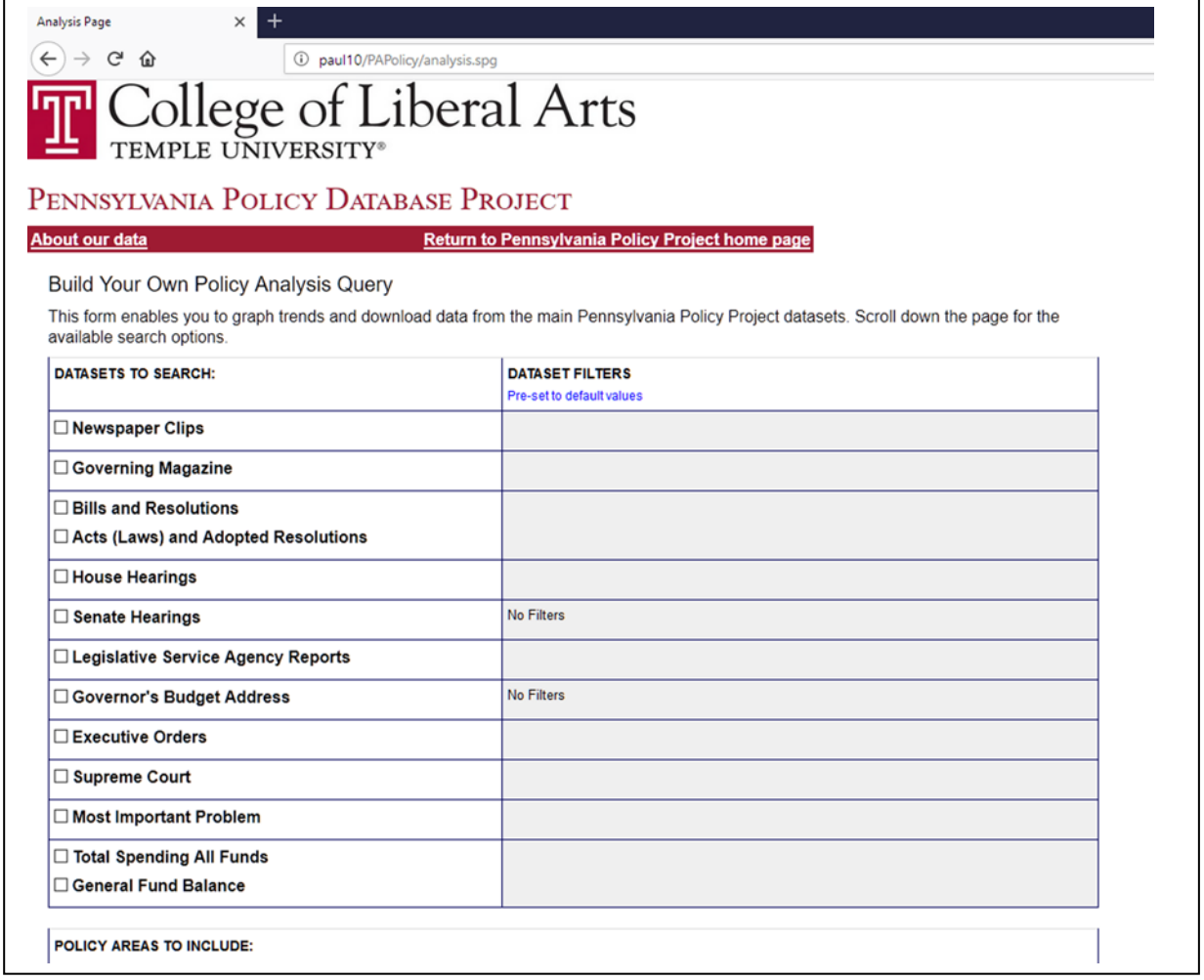


The Bills data is the most complicated. Bills originating in the House, Senate, or both can be selected. Bills, Resolutions, or both may be selected. Finally one can also (or only) select Laws. The selection of originating house and bill vs. resolution apply to the laws. Also, all filters are common between bills and laws. If selected, laws act as a separate database in the data display.

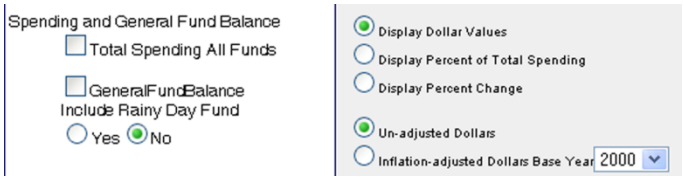


The Spending and General Fund Balance allows for selection of the Total Spending All Funds, General Fund Balance, or both. The General Fund Balance can optionally include the “rainy day fund”. While

Figure 1 A Portion of the Analysis Page



these represent separate data sets, they are displayed in a common format, and have a common inflation adjustment applied. This is specified using the filters column of the form.



The Tables Table

To specify which datasets are available a table in the database is defined. This table is given the name Tables. Table 1 shows the structure of the Tables table. Flexibility is achieved by using different Java classes to process the different kinds of datasets. The datasets which are selected by a single checkbox are processed by the Table class. Special classes are defined for the House Hearings, Bills, Public

Opinion, and spending data. Table 2 shows the table name, table title, and Java class from the Pennsylvania Policy Database's Tables table.

Table 1 Structure of the Tables table.

Column	Contents
ID	A unique identifying integer
TableName	The database table that contains the data
TableTitle	The title that is displayed in the selection column
idColumn	The column containing the ID
MajorOnly	A flag to indicate that this dataset only is classified by major category.
MinYear	The minimum year data is available
MaxYear	The maximum year data is available
Class	The Java class name that defines the kind of data set. Most datasets are described the StandardTable class. Special classes are defined for Hearings, Bills, Public Opinion, and Budget.
TextColumn	The name of the column that contains the text of interest.
YearColumn	The name of the column or SQL expression that contains the year.
LinkColumn	The name of the column that contains the hyperlink to the raw data behind the record.
DrillDownFields	The names of the columns that are to be displayed in the drilldown table.
CodeColumn	The name of the column containing the policy code
Note	Optional text to be displayed when the table is selected. Currently used for the Supreme Court data.

Table 2 Selected columns from the Pennsylvania Policy Database Tables Table

ID	TableName	TableTitle	Class
1	NewsClips	Newspaper Clips	StandardTable
2	Governing_Magazine	Governing Magazine	StandardTable
3	Bills_Data	Bills and Resolutions	BillsTable
4	Transcript	House Hearings	TranscriptTable
5	Senate_Hearings	Senate Hearings	StandardTable
6	LegServiceAgencyReports	Legislative Service Agency Reports	StandardTable
7	GovernorsBudgetAddress	Governor's Budget Address	StandardTable
8	ExecOrders	Executive Orders	StandardTable
9	SupremeCourt	Supreme Court	StandardTable
10	PublicOpinion	Most Important Problem	PublicOpinionTable
11	BudgetTable	Budget	BudgetTable

Filters

The filters are defined by the Filters table as shown in Table 3.

Table 3 Structure of the Filters table

Column	Contents
ID	A unique ID number assigned to this row
TableID	A reference to the table to which this filter applies
ColumnName	The database column containing the data to be used for selection
Description	The description of the filter
FilterClass	The Java class name for this filter. The different filter classes are described

	below.
TableReference	The table where additional filter selection criteria are stored. This applies to filters where several choices are provided.
AdditionalParam	Additional specialization data that is needed for some of the filters. Details described below.

There are several kinds of filters.

Binary Filter

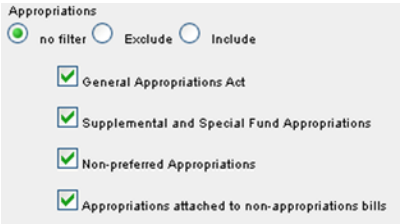
The BinaryFilter allows for three choices: no filter, Exclude, or Include

Mention Filter

A special filter is used for newspaper clips, called the MentionFilter that allows for four choices: no filter, No Mention, or Mention.

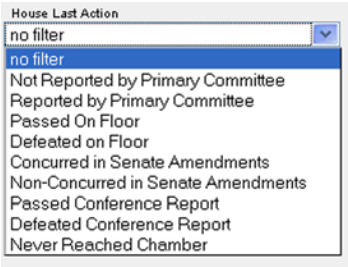
Multi-valued Filter

The multi-valued filter is an extension of the binary filter that allows the filter to either include or exclude one or more of a sub-category selection. It is used with the newspaper clips to specify the document type and with the Bills and Laws to select the final state of Constitutional Amendments and the kind of Appropriation Bill. The list of choices is specified by a separate table that is referenced in the Filters table.



Drop-down Filters

The drop-down filters display their selection choices using a drop-down box. They are used to display the committees and the last house/senate action on a bill.



Since special processing is required to select the committees, special Java classes are defined for hearings committees and the bills committees.

Check-Box Filters

The check-box filters display a selection option with a check-box. By checking the box the selected item is included, leaving it unchecked the item is excluded. This is similar to the binary-filter except that the "no filter" option is not available.

Where Held Filter

A special filter is defined for the committee hearings to select hearings that are held in the state capitol, outside the state capitol, or both. The name of the state capitol is defined by the AdditionalParam column of the Filters table.

Most Important Problem Filter

This is not actually a filter, but is used to specify the display format. The Most Important Problem can be displayed either as a percentage or by rank.

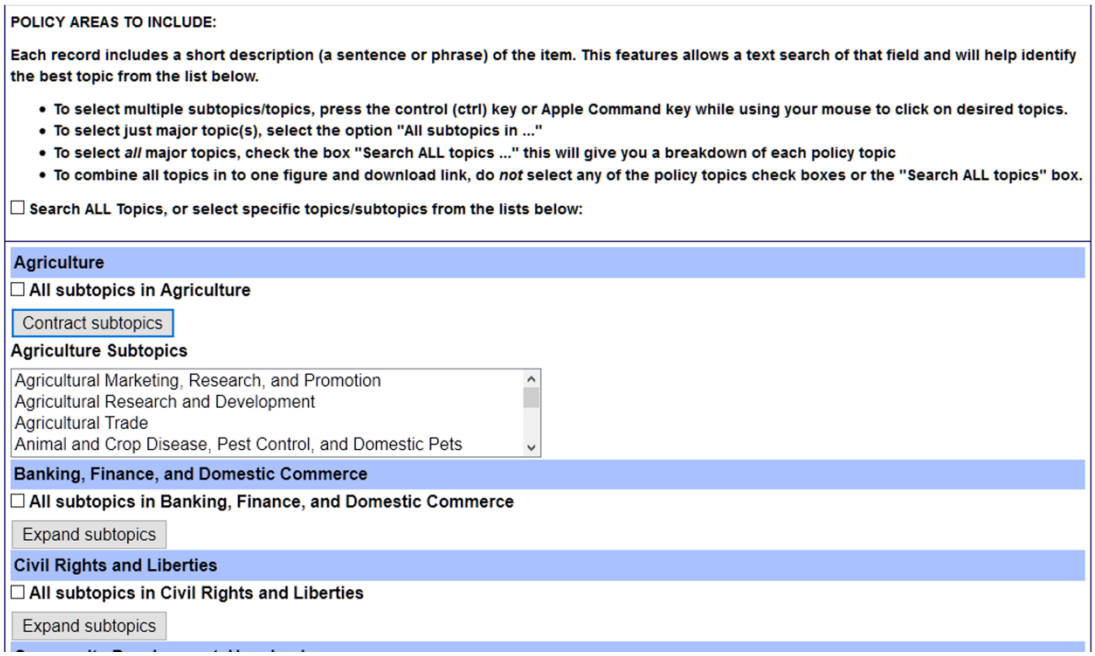
Budget Filters

This is also not a filter, but is used to select the display format. Spending and general fund balance data can be displayed in dollars, percent of total spending, or percent change. The raw data values can be displayed or the values can be adjusted for inflation with the base year specified.

Selection of Policy Area

The selection of policy area is table driven, just as it was in the Policy Agendas project. The names of the tables are "hard-coded" into the program currently. The table MajorCode contains the major codes and their description and the table Code contains the minor codes and their description. The selection form was copied from the Policy Agendas project as shown in Figure 2.

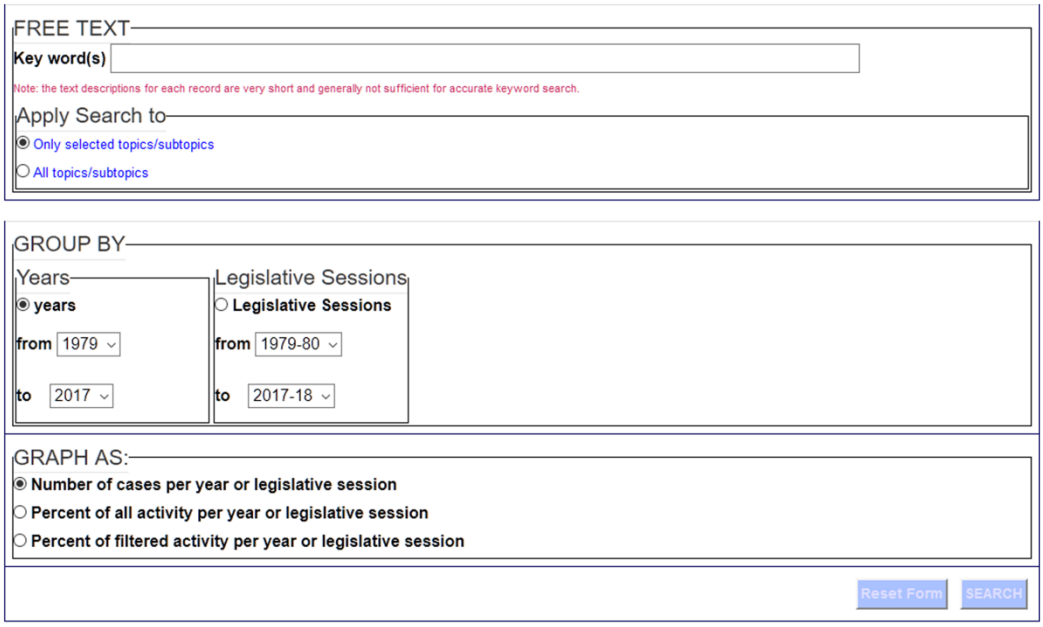
Figure 2 Policy Areas Selection



Remainder of the form

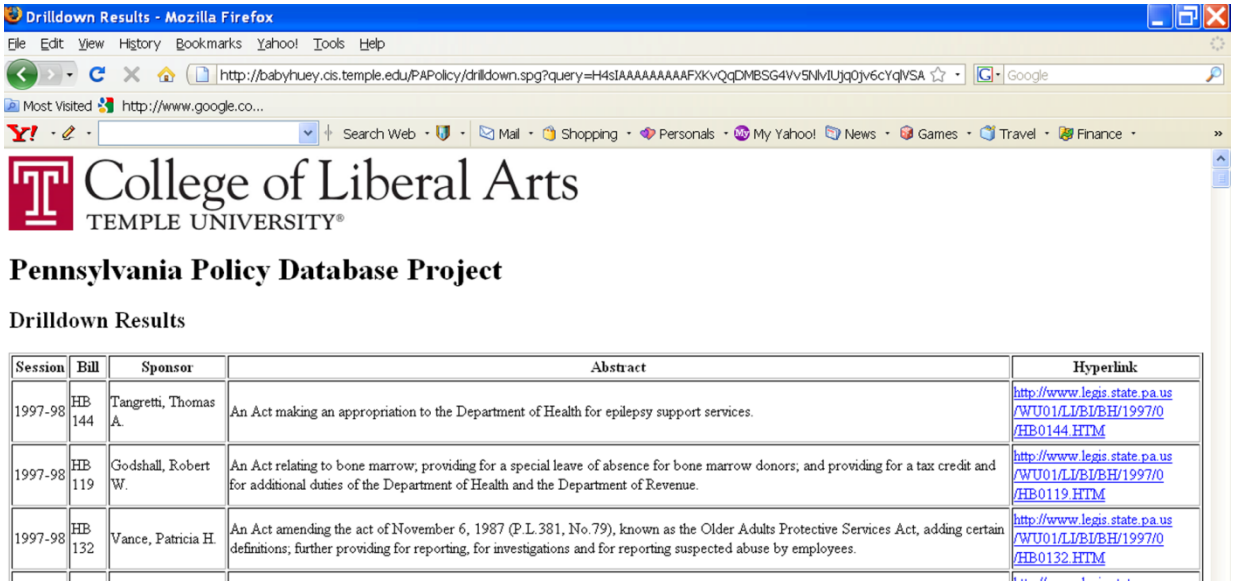
The remainder of the form allows for selection of free text, grouping of the data either by year or legislative session, and graph format as shown in Figure 3.

Figure 3 Remainder of the Analysis Form



Results Page

An example of the results is shown in Figure 4. At the top are links to download the selected data as an Excel spreadsheet. This is followed by a line-graph of the data. Finally there is the drill-down table that lists the values for each row and column. The entries in the drill-down table are links that will display the detail data for that row and column. The data columns displayed are specified by the DrillDownFields in the Tables table. An example of the drill-down results is shown below.



Session	Bill	Sponsor	Abstract	Hyperlink
1997-98	HB 144	Tangretti, Thomas A.	An Act making an appropriation to the Department of Health for epilepsy support services.	http://www.legis.state.pa.us/WU01/LI/BI/BH/1997/0/HB0144.HTM
1997-98	HB 119	Godshall, Robert W.	An Act relating to bone marrow; providing for a special leave of absence for bone marrow donors; and providing for a tax credit and for additional duties of the Department of Health and the Department of Revenue.	http://www.legis.state.pa.us/WU01/LI/BI/BH/1997/0/HB0119.HTM
1997-98	HB 132	Vance, Patricia H.	An Act amending the act of November 6, 1987 (P.L.381, No.79), known as the Older Adults Protective Services Act, adding certain definitions; further providing for reporting, for investigations and for reporting suspected abuse by employees.	http://www.legis.state.pa.us/WU01/LI/BI/BH/1997/0/HB0132.HTM

Figure 4 Sample Analysis Results

Analysis Page

paul10/PAPolicy/display.spg

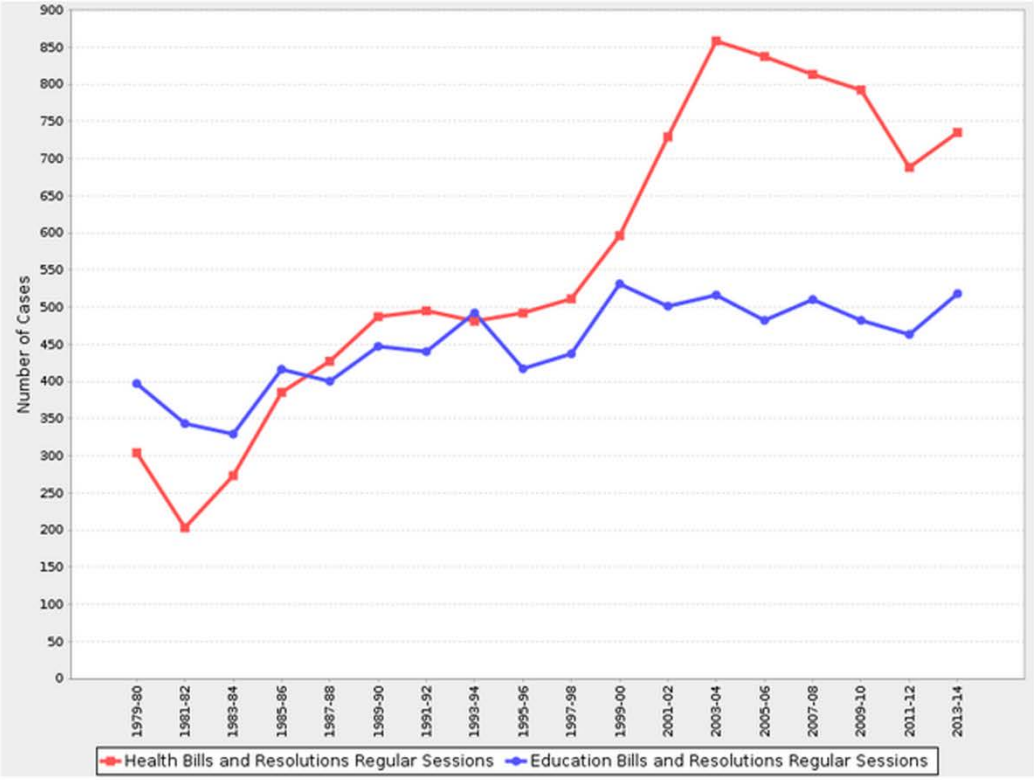
College of Liberal Arts
TEMPLE UNIVERSITY®

PENNSYLVANIA POLICY DATABASE PROJECT

About our data | Return to Pennsylvania Policy Project home page

To download the raw data as an Excel spreadsheet, click on the link

Health Bills and Resolutions Regular Sessions
Education Bills and Resolutions Regular Sessions



To download the image, right-click on the image

Session	Health Bills and Resolutions Regular Sessions	Education Bills and Resolutions Regular Sessions
1979-80	304	397
1981-82	203	343
1983-84	273	329
1985-86	385	416

How it works and how to customize

The Table classes

Table classes are responsible for:

- Generating the HTML that is used to create the dataset selection cell in the analysis form.
- Generate the SQL query to select the data by topic for a date range.
- Generate the SQL query to select the total over all topics for a date range.

The AbstractTable class provides default implementation for all methods required by the Table interface except for getTitleBox. The StandardTable implements this method. For most datasets the StandardTable is configured using the parameters defined in the Tables table. For the bills and resolutions, budget data, hearings data, and most import problem data, special Java classes are defined that "override" some of the methods of the AbstractTable class. This allows for more complicated processing for these datasets. While the StandardTable class is fairly general these other classes are coded based on the specific structure of the PA policy data. To adapt this system to the another project data the StandardTable class could be used for most of the datasets, but special classes will be needed for the budget, hearings, and most import problem data.

The filters

For each of the filters described above there is a Java class that is responsible for:

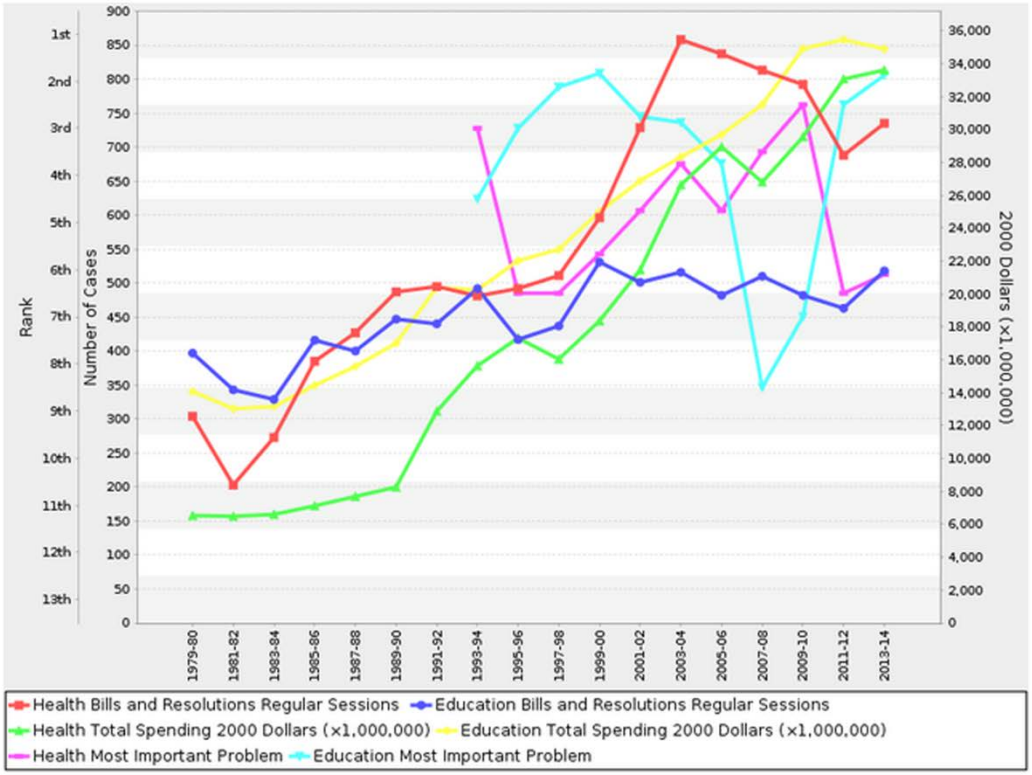
- Generating the HTML that is used to create the filter selection cell in the analysis form.
- Generate part of the SQL query to restrict the selection to the filter criteria.

The most projects would only need the BinaryFilter and the MultiValuedFilter.

Graph Display

To generate the graph an open-source Java package known as JFreeChart is used. This package can generate a variety of charts. If the selected data is for one dataset over several years (sessions), a bar chart is generated. For multiple datasets over several years (sessions) line chart is generated, and if multiple datasets are selected for a single year (session) a pie-chart is generated. JFreeChart allows for the display of multiple axes the budget data (units of dollars), most import problem (units rank), and count of bills and resolutions can all be displayed as shown in Figure 5.

Figure 5 Chart with multiple axes.



To download the image, right-click on the image

References and Sources

The current PA Policy Database is at http://policydb.cla.temple.edu/PAPolicy

Documentation on Java and Java Server Pages is at http://java.sun.com

The Spring framework is used, documentation at http://www.springsource.org/about

JFreeChart is at http://www.jfree.org/jfreechart/