Discoverer Web for AQS

Goals for this class

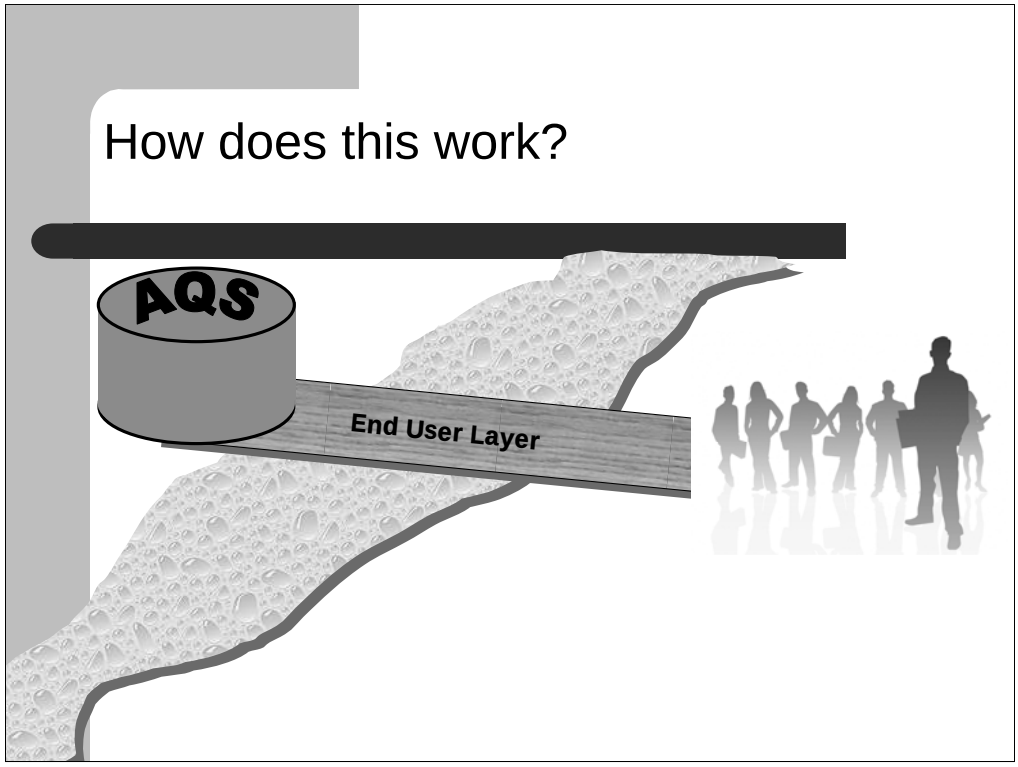
- Introduce Discoverer Plus
- Learn About the Data in AQS
- Practice Finding Data Using Discoverer
- Have Fun!

2

What is Discoverer?

- A business intelligence tool for customizable queries, reporting, analysis, and Web publishing from Oracle Corporation
- A retrieval only tool that works on a relational database

3



The End User Layer is the interface to AQS data. It

•contains metadata (i.e., the data about the data in the database).

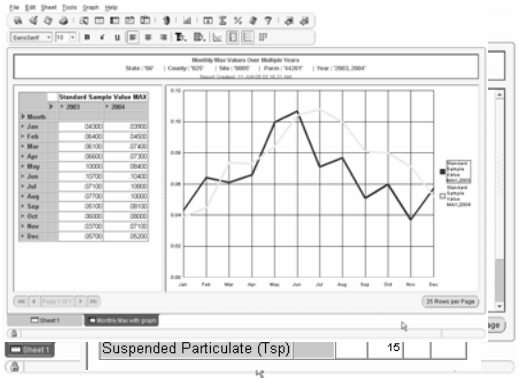
•consists of a collection of about 50 tables in the database.

•provides a more intuitive view of the data so end users don't have to be so familiar with the physical structure of the database

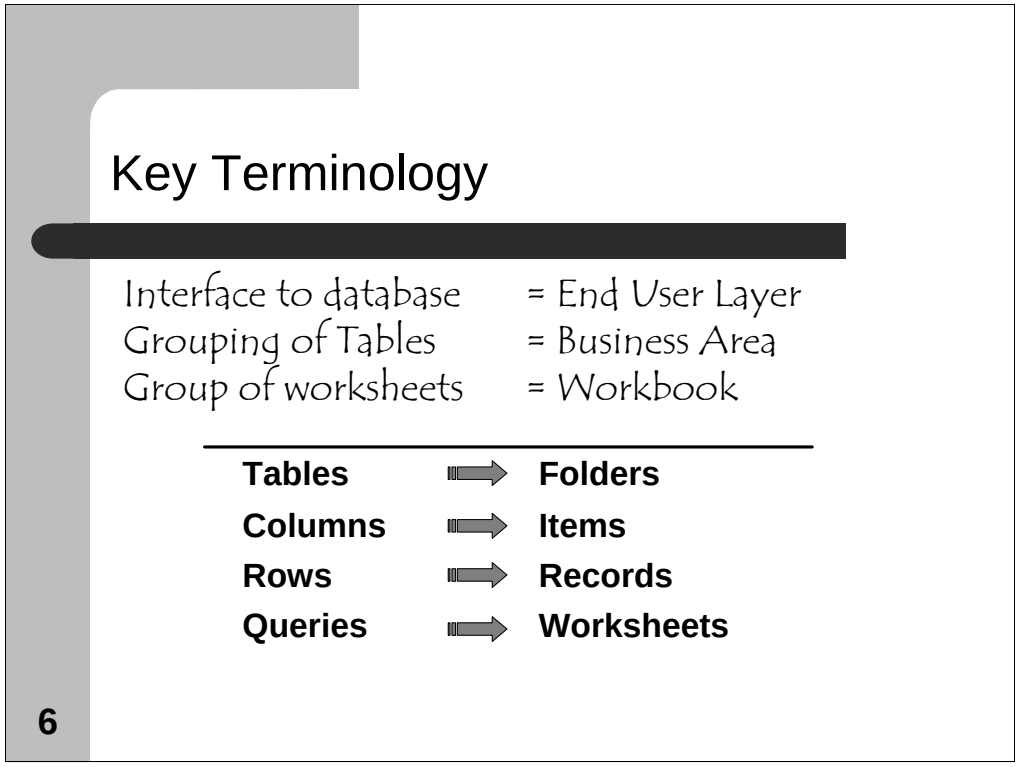
•protects the data in the database – neither you nor the Discoverer manager can affect the integrity of the database with Discoverer.

What Can Discoverer Produce?

- Data Listing Reports
- Cross-Tabulation Reports (Pivot Tables)
- Graphs & Charts



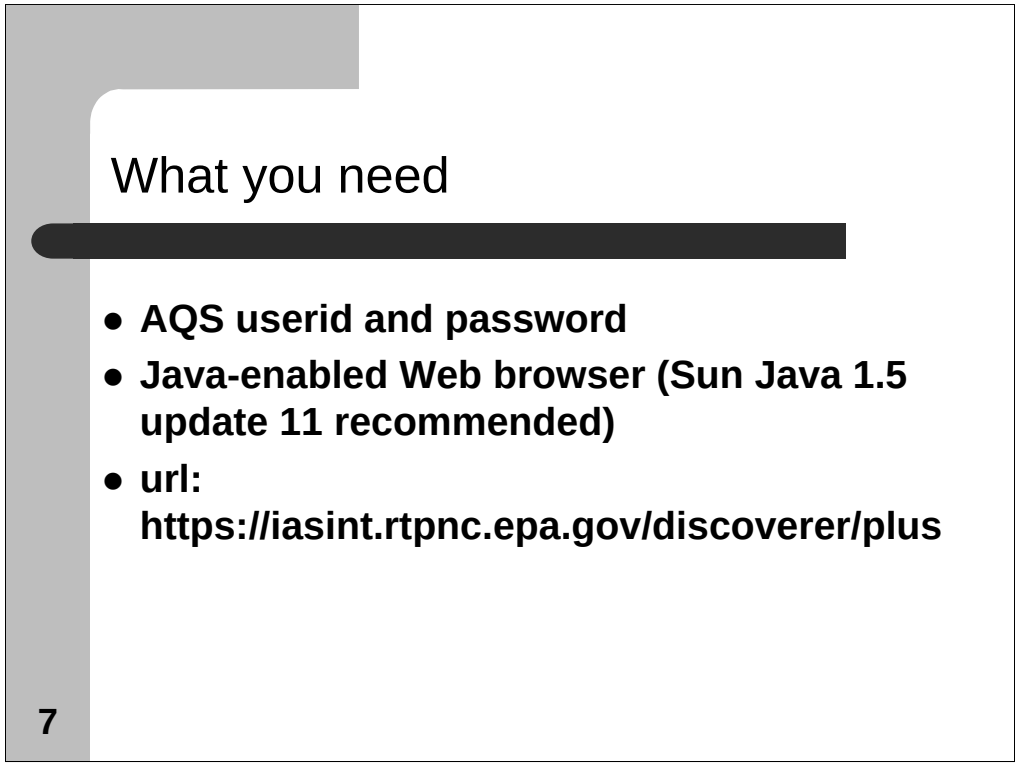
5



Business Areas are subsets of tables that contain related information. Business Areas are organized by the Discoverer Manager.

If you're familiar with database terminology, you'll have no trouble with this. Many times the terms "tables and folders", "columns and items", "rows and records", and "queries and worksheets" are used interchangeably.

Workbooks are Discoverer files that contain worksheets displaying related data retrieved from the database. Worksheets contain the data that you want to analyze, together with a number of Discoverer components to help you analyze the data. If you are familiar with spreadsheet applications (e.g. Microsoft Excel), think of a workbooks as spreadsheet files and worksheets as different sheets in a spreadsheet file.



All AQS users have access to Discoverer for AQS.

How To Build a Query in Discoverer

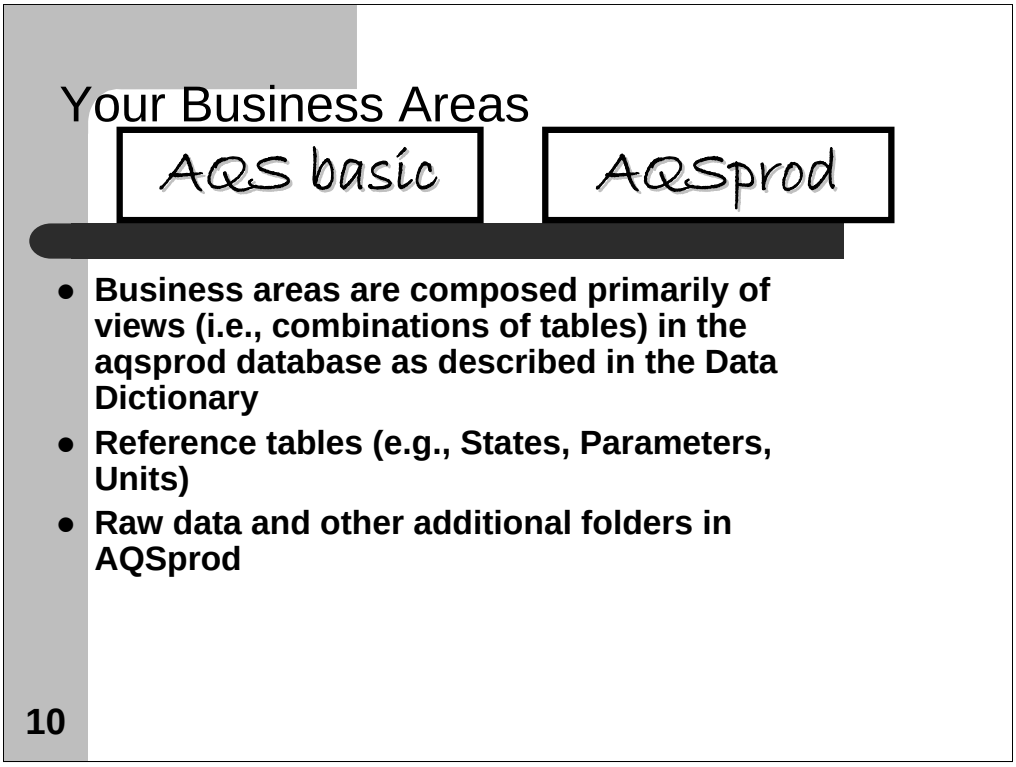
- Connect
- Choose a Layout
- Select Items
  - To Display
  - To Filter By
- Specify Totals and other Calculations (Optional)
- Specify Sorting (Optional)
- Specify Run-Time Parameters (Optional)
- Run Report

8

The REAL Way to Build a Query…

- FORM A GOOD QUESTION!!
- Understand the Data You Need to Answer the Question

9

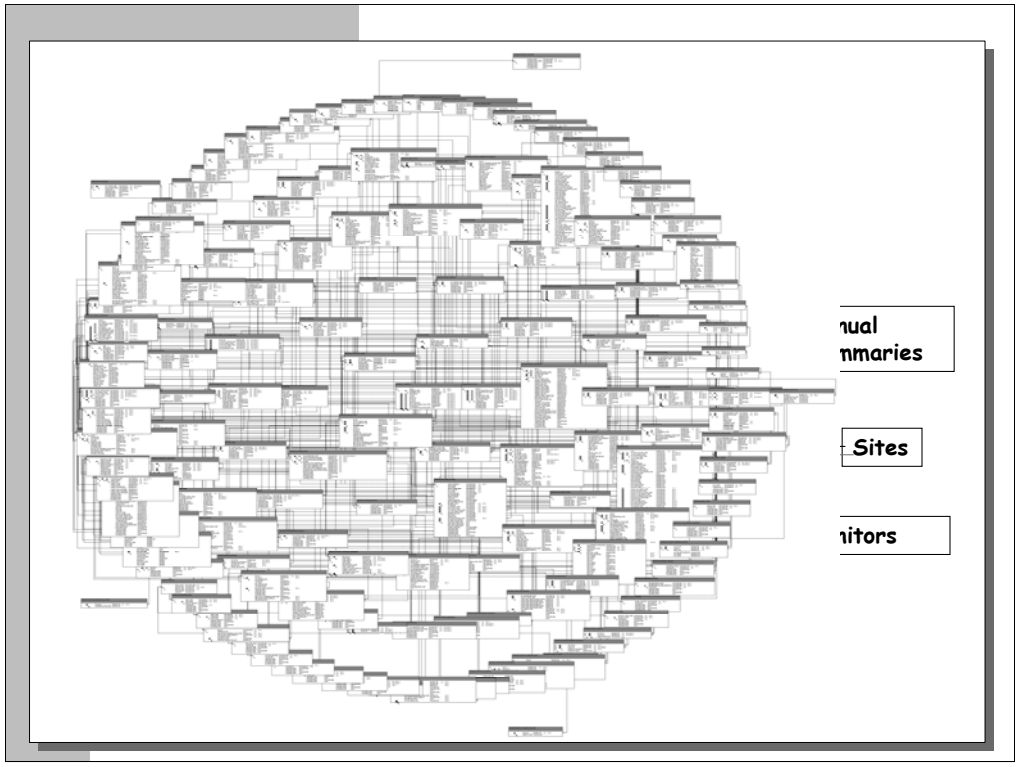


There are two business areas available – AQS basic and AQSprod.

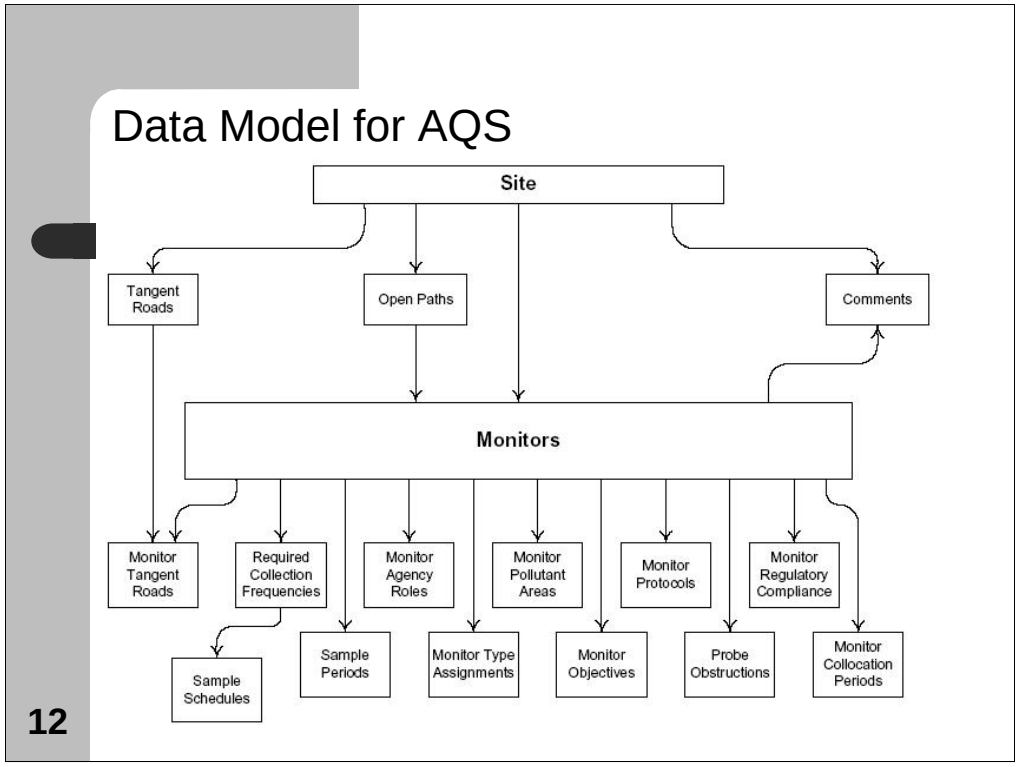
AQS basic contains reference tables, site and monitor data, and summary data, including raw data daily, quarterly, and annual summaries. AQSprod includes the raw data.

There are links to tables of items in each of the business areas on the AQS Discoverer web page

http://www.epa.gov/ttn/airs/airsaqs/aqsdiscover

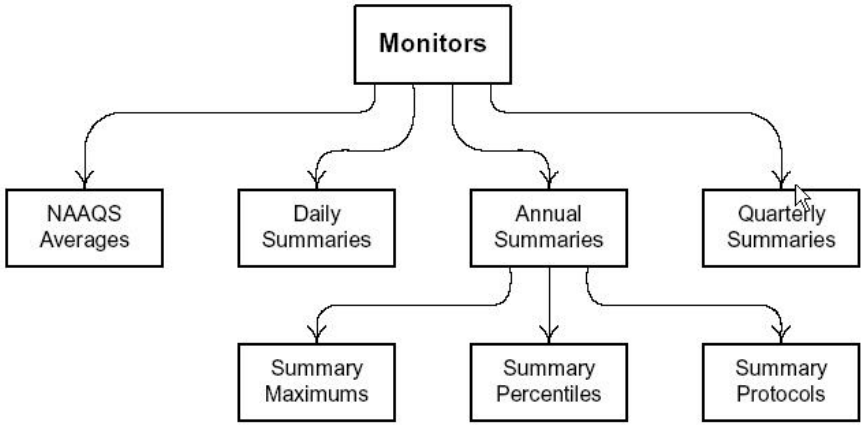


This is the Entity Reference Diagram (ERD) for AQSPROD, shown only to illustrate the complexity of the database.



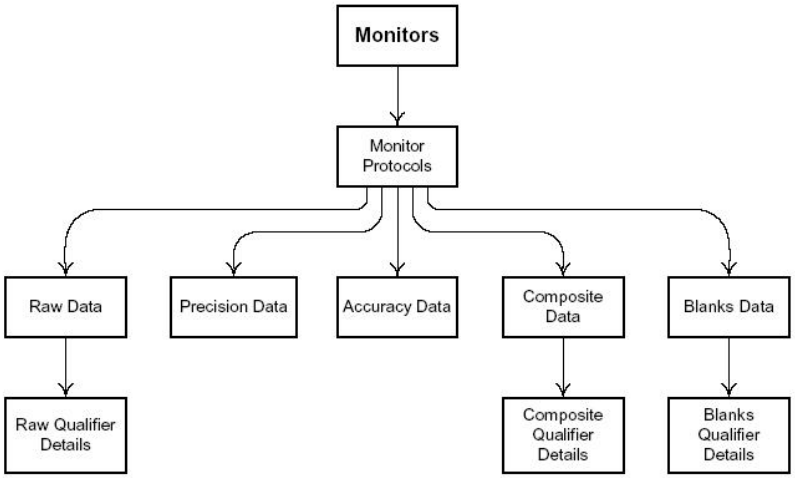
Briefly stated, the structure of the data in AQS is: **Sites contain Monitors that contain Raw Data, Summary Data, Precision Data, and Accuracy Data**. This and the following diagrams illustrate that structure. Many of the folders in the two available Business Areas in Discoverer will correspond to a box in these diagrams.

Data Model for AQS, cont'd



13

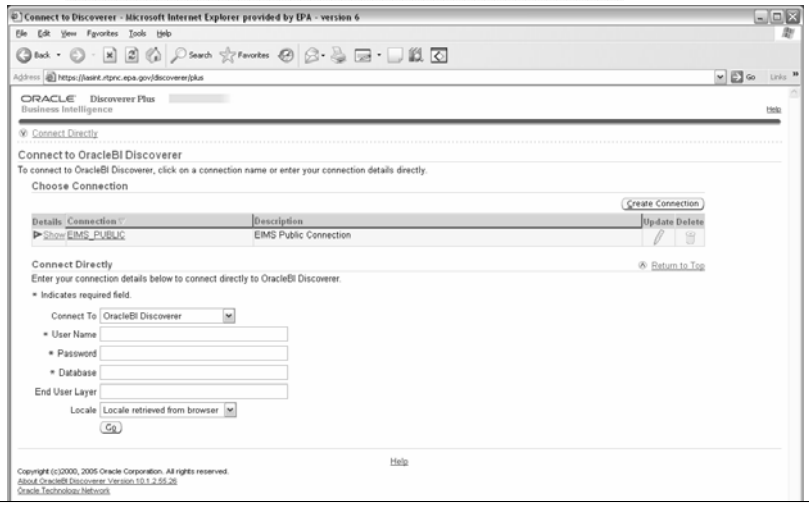
Data Model for AOS. cont'd



14

Starting Discoverer Plus

https://iasint.rtpnc.epa.gov/discoverer/plus



15

You can get to Discoverer via the direct URL:
https://iasint.rtpnc.epa.gov/discoverer/plus

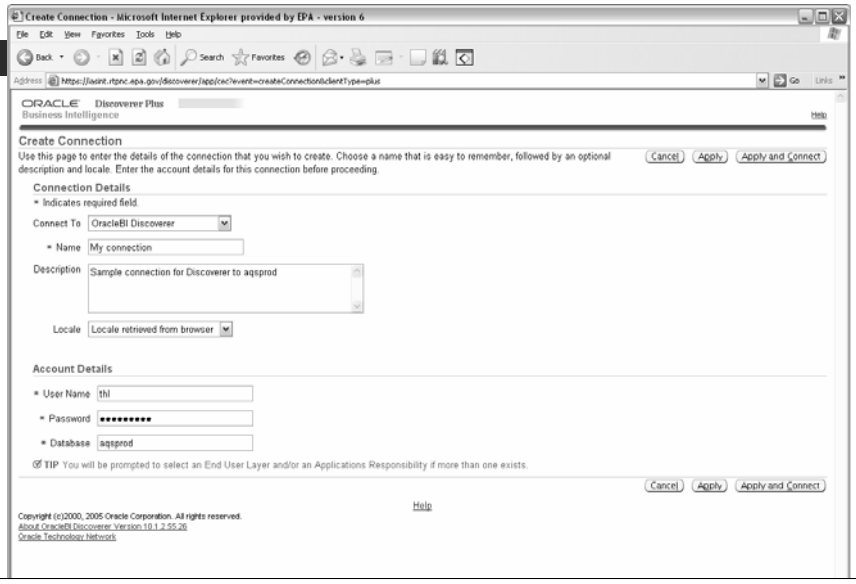
or you can follow the links on the AQS Web site. Start with the AQS Discoverer link on the left side bar.

If you can logon to AQS, then you most likely have a sufficient version of Java to run Discoverer. Java Version 5 update 11 is the recommended version. The most recent update levels of Java 6 from java.com do not work well with Discoverer.

To connect just enter your User Name (three character userid, same as for AQS), your AQS password, and aqsprod for the Database. Click on Go.

A Discoverer connection stores Discoverer login details that can be used in subsequent logins. If you want to create a connection, click on Create Connection.

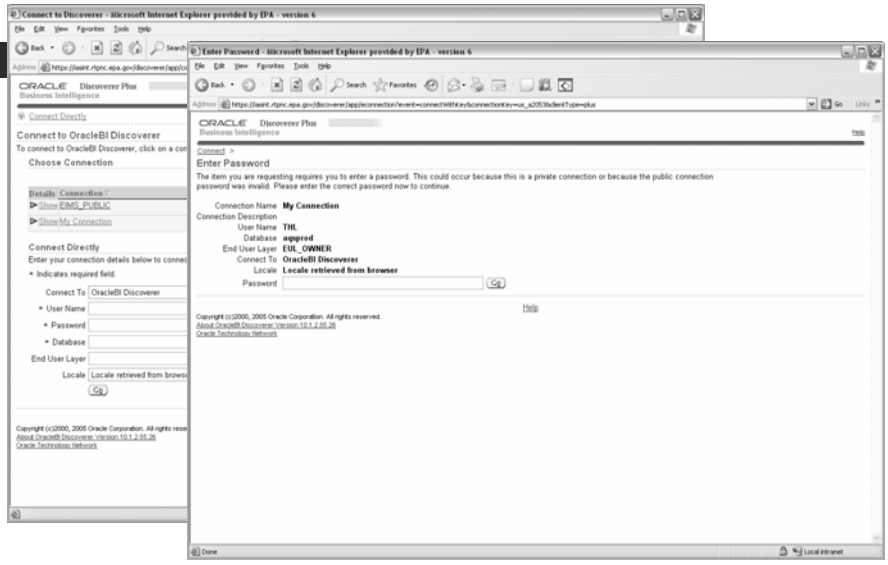
Create a Connection



16

To create a connection, enter the fields and click either “Apply” or the “Apply and Connect” buttons.

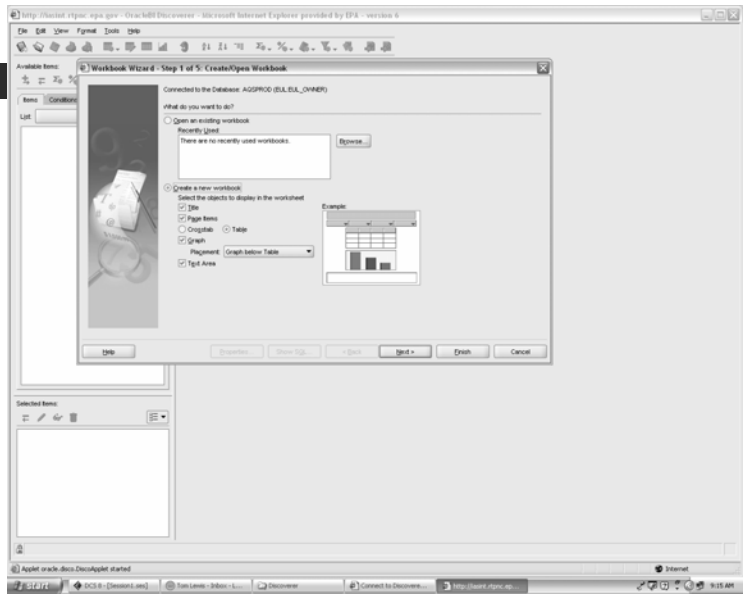
Create a Connection, cont'd



17

If you use a Connection, you'll still need to enter your Password. When you define a connection, it will show up in your Connection list (shown in background screen). If you have to delete your browser cookies or cache, you will lose your saved Connection.

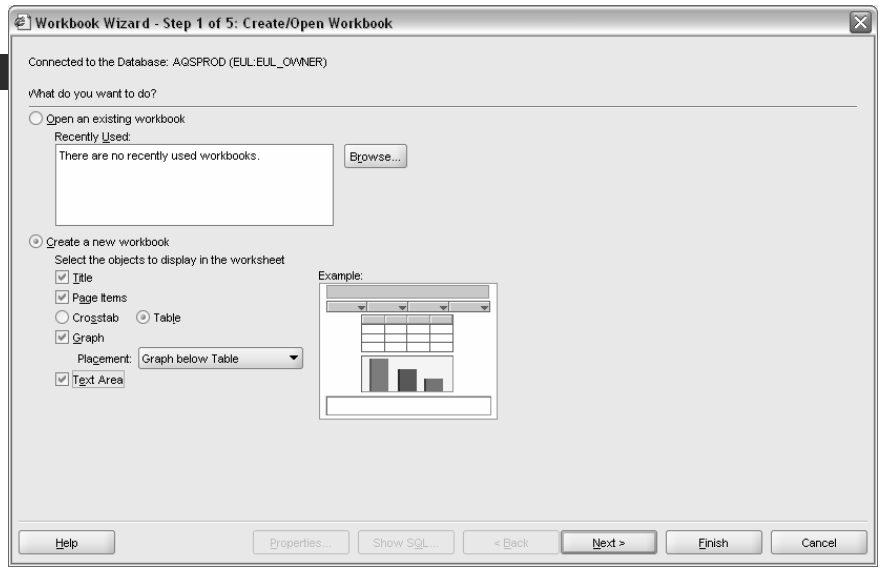
Successful connection



18

Here's the result of a successful connection. Step 1 of the Workbook Wizard is presented. Select either "Create a new workbook" (set up as the default) or "Open an existing workbook".

Create/Open a Workbook

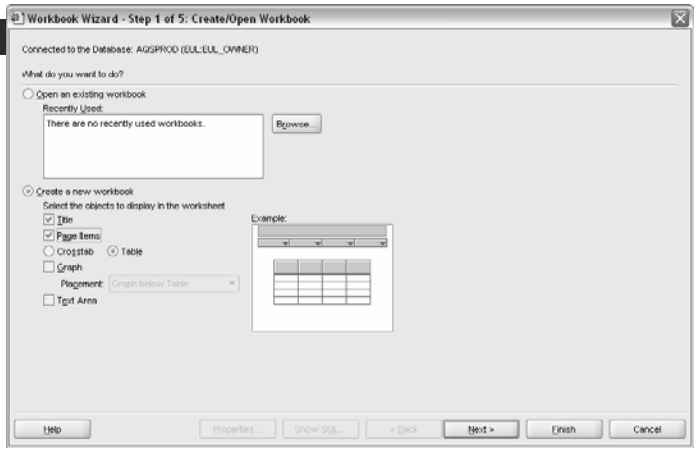


19

To open an existing workbook, select that option and click on Browse. You'll be presented a list of your saved workbooks and workbooks that others have shared with you.

Creating a New Workbook

- Use the Workbook Wizard
- Start with a simple goal in mind – e.g., **Find the states in your region**



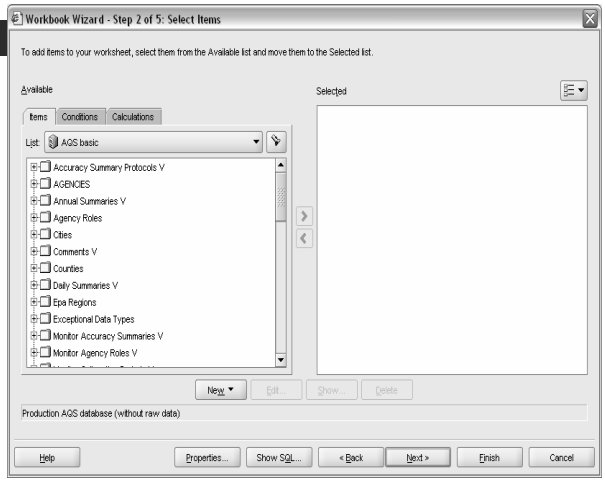
20

To create a new workbook, select that option and choose the page layout by selecting/deselecting objects. As you deselect/select objects, they will disappear/appear in the Example. Initially, keep it simple, e.g., deselect the graph, text area, and page items. (If an item is made a page item, there will be a separate worksheet page for each value of the item.)

Once you've selected your page layout, click on Next to proceed to Step 2 of the Workbook Wizard.

Selecting Items

- AQS basic is the simplest Business Area
- Scroll through list of folders
- Expand a folder to see its items

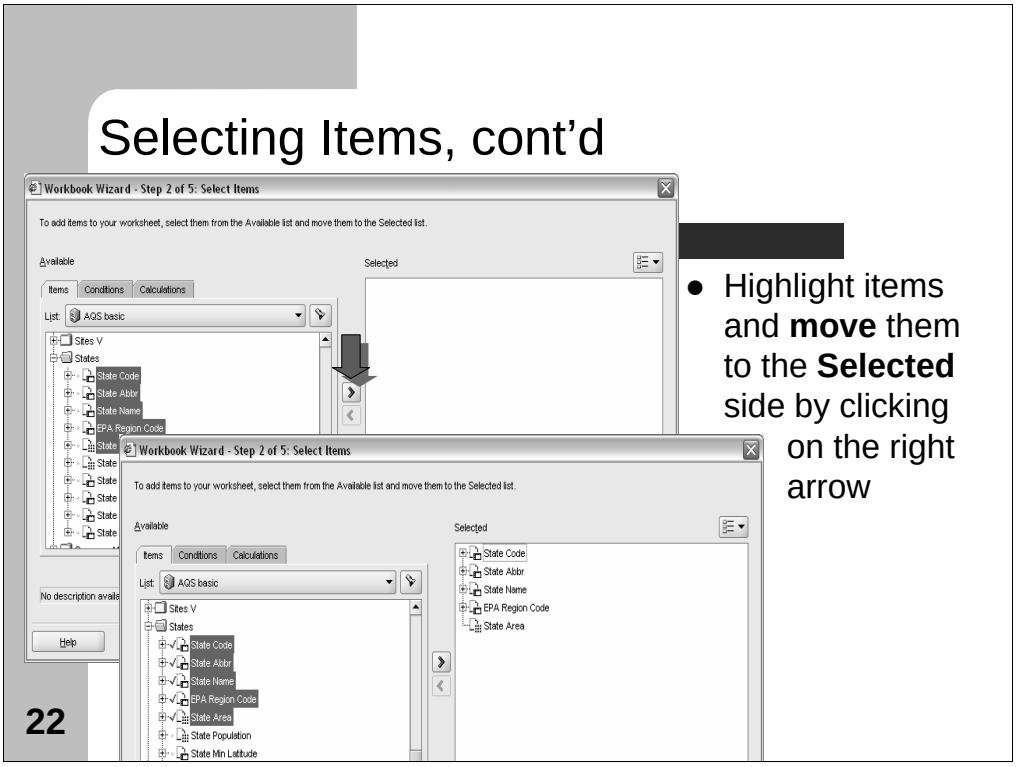


21

There's a lot to do in Step 2. From the Items tab, select what you want to appear in the table; from the Conditions tab you specify conditions so you only get data that meets certain criteria.

Select the business area from List, in this case, AQS basic. To find the items you want, scan through the list of folders. The type of data is usually fairly self-evident from the folder name. The V appended to the folder name indicates it's a view, i.e., a combination of tables. Expand the folders to see the items they contain.

To search for folders and items that contain a string, use the Find icon (flashlight).

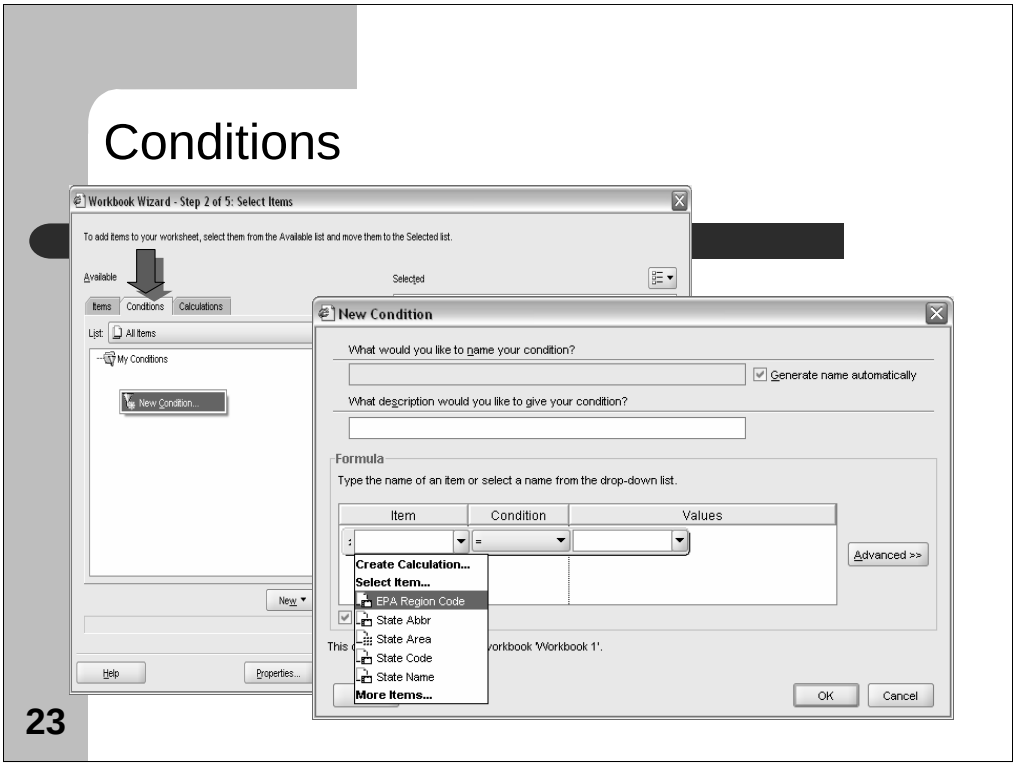


To select items, highlight them and move to the Selected area with the right arrow. You can also remove an item from Selected with the left arrow. You can drag and drop items back and forth as well.

To select all items in a Folder, select the Folder itself and move it over.

Here, items from the States table have been highlighted and selected. Selected items will be checked in the Items list.

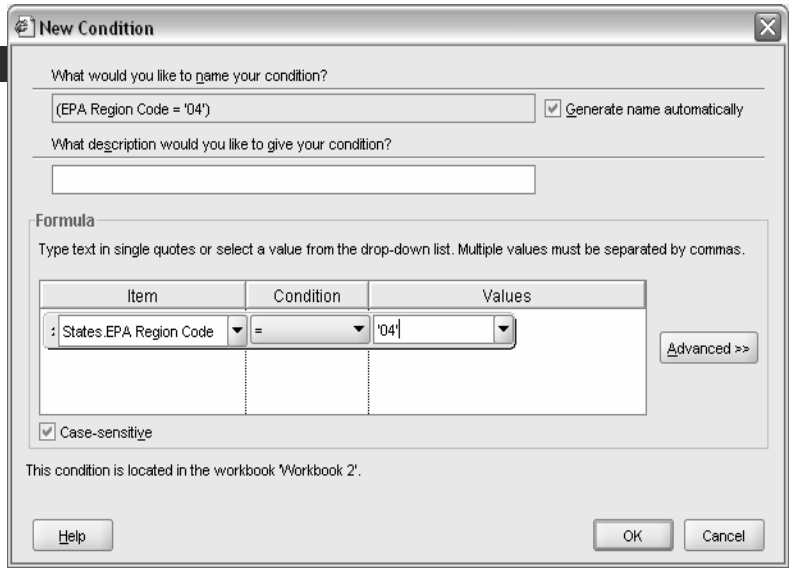
Once items are selected, some folders will become grayed out, indicating that their items cannot be joined with what has been selected. For example, with these items from the States folder selected, the Sampling Methodologies table is no longer available.



Conditions are probably the most important part of creating a worksheet. When creating a new worksheet it's a good idea to use conditions to limit the amount of data that will be selected until you have a good feel for how much data will be returned.

To create a condition, go to the Conditions tab. Right click in the Conditions pane and select New Condition, or click on the New button and select New Condition. This will bring up the New Condition dialog. Select the item to test against, here EPA Region Code.

Conditions, cont'd



24

Specify the conditional operator and value(s). Some items with a discrete number of possible values will give you that list of values (LOV) to pick from.

The Advanced button will allow you to specify compound conditions separated by logical operators AND or OR. There is an implied AND between individual conditions.

Three conditional operators of note:

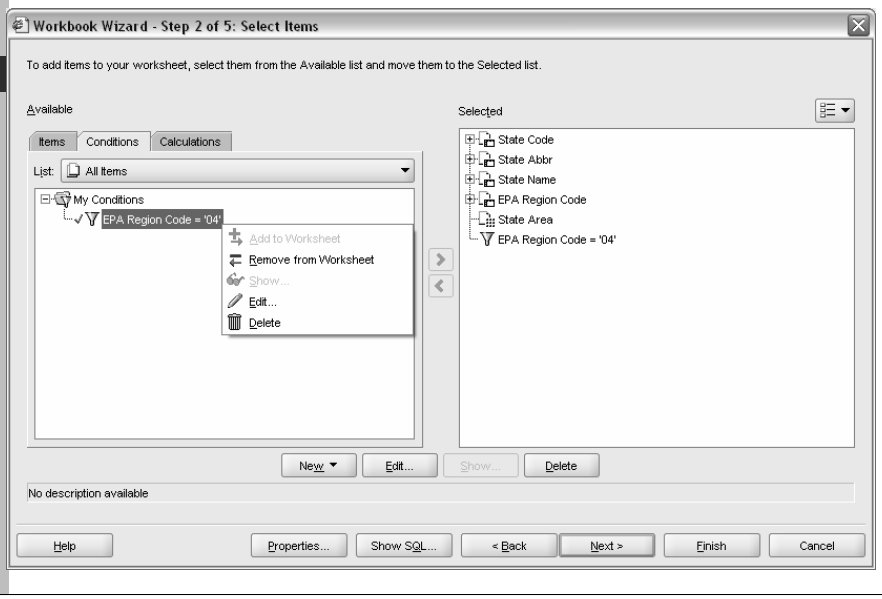
LIKE – allows wild card specifications in matching the value of the item. The _ (underscore) matches a single character, and the % matches any number of characters. For example, if testing against a monitor ID, both '37%-88101-%' and '37-___-____-88101-_' would select all PM 2.5 monitors in North Carolina.

BETWEEN – two values are specified that the selected item's value must fall between, particularly useful for date ranges.

IN – a list of possible values of the selected item is specified, separated by commas.

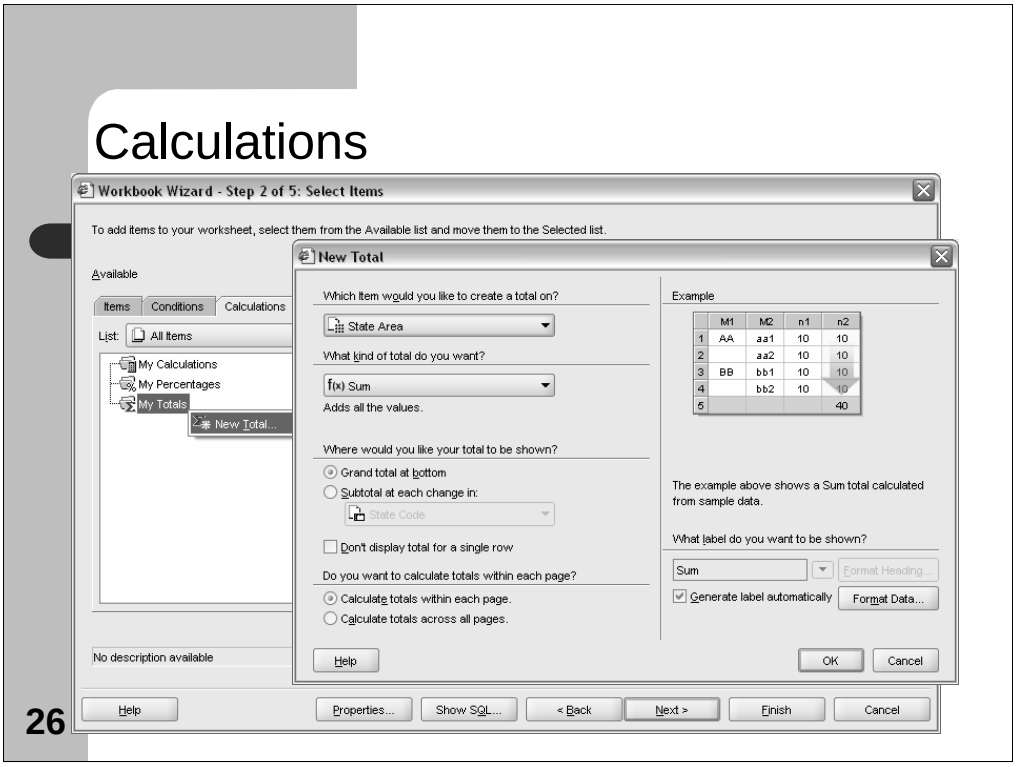
If quotation marks are required in the value specification, Discoverer will supply them. In other words, "Type text in single quotes …" in the instructions for specifying values can be ignored.

Conditions, cont'd



25

Once a condition has been created, you can remove it by right clicking on it and selecting "Remove from Worksheet". It will still be available, just not active. It can be added back with "Add to Worksheet".

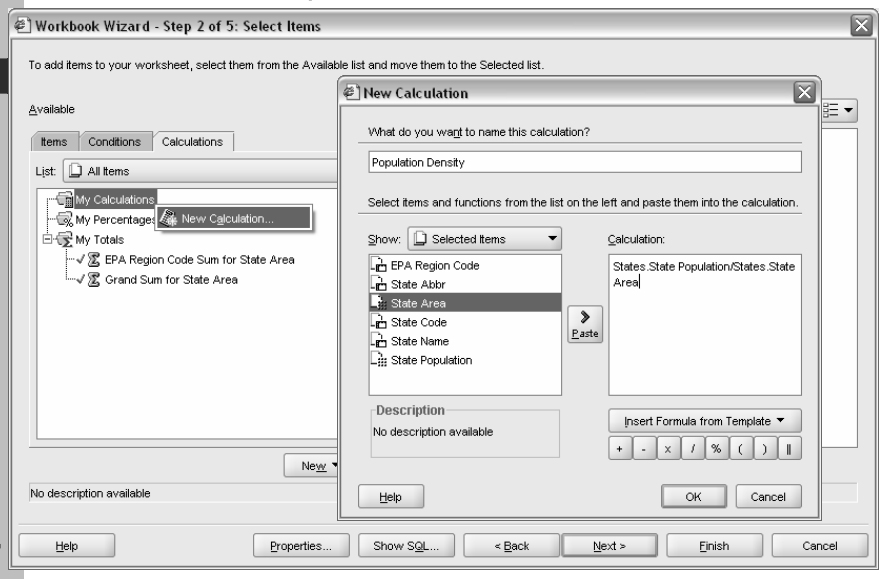


From the Calculations tab, you can add totals or percentages, and you can calculate a new column.

To create a total, right click on My Totals and select New Total, right click anywhere in the Calculations pane and select New Total (New Percentage and New Calculation will also be options), or click on the New button and select New Total.

Here a grand total of State Area is created.

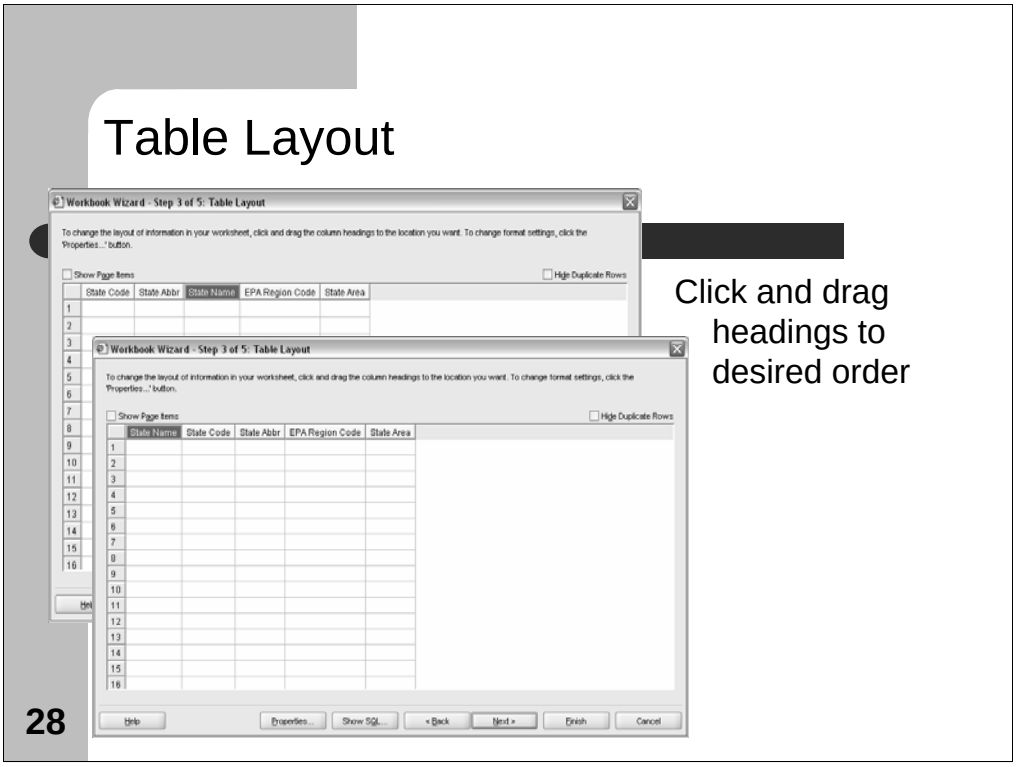
Calculations, cont'd



27

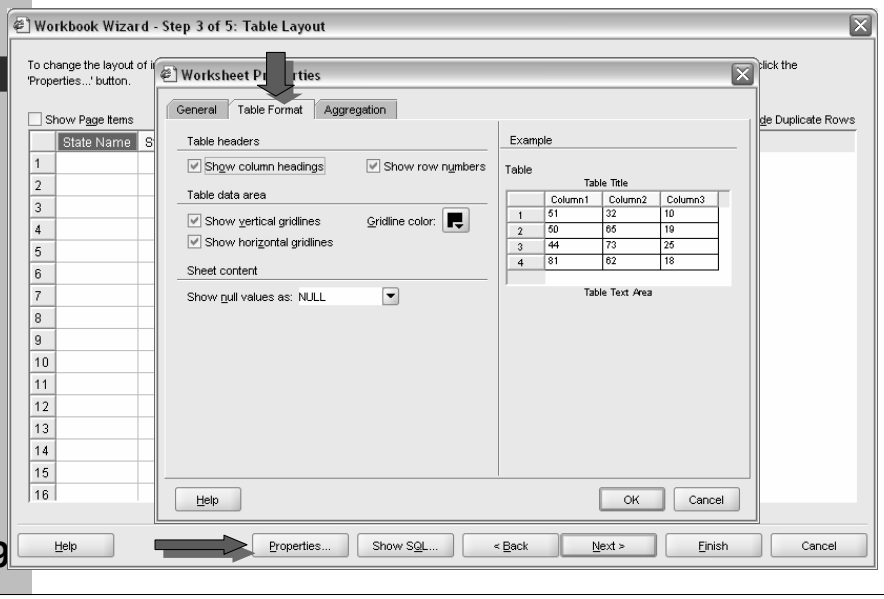
Start the process of creating a new calculation in similar fashion as for a new total, i.e., right click on My Calculations and select New Calculation, etc.

Construct the calculation in the Calculation pane of the New Calculation window. Move items and operators into the pane. Here, a new column, Population Density, is created as the ratio of State Population to State Area. The calculation was built by pasting State Population, followed by clicking on /, and pasting State Area. Under Show, you can select a wide range of functions to paste into the Calculation (Show/f(x) Functions). The "Insert Formula from Template" button will help you construct a formula for a limited set of specific purposes, e.g., calculating a running total.



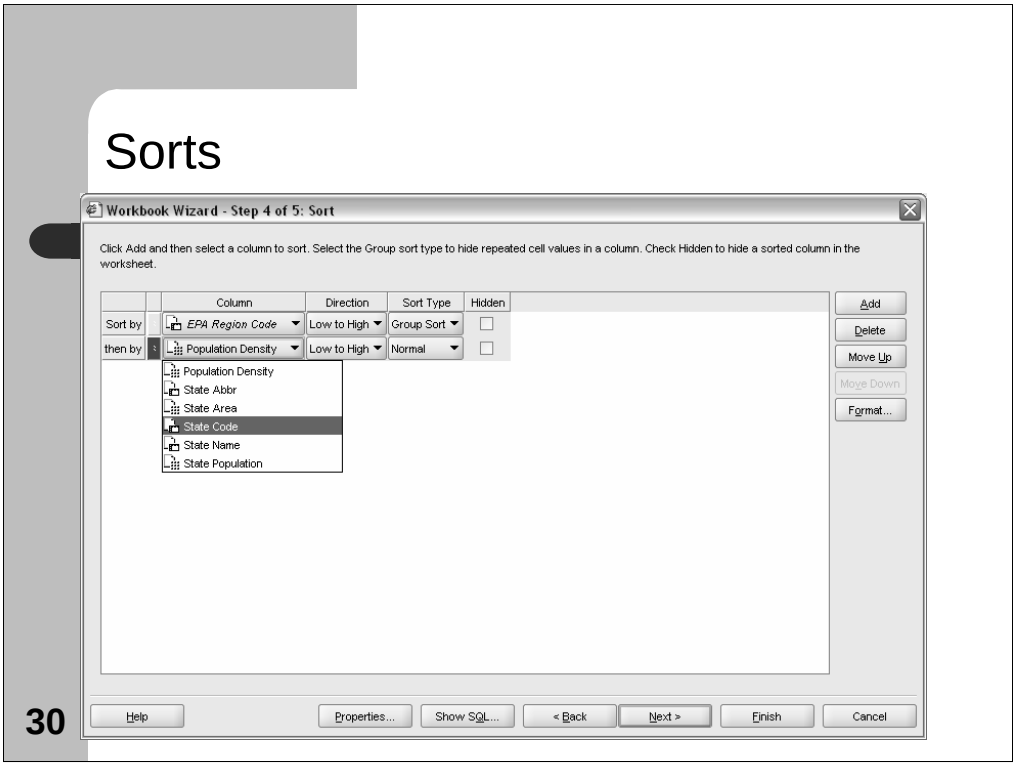
Step 3 of the Workbook Wizard is Table Layout. To change the order of columns click and drag the headings to the desired location. As you drag, lines between the columns will highlight as you pass over them. Drop the heading where you want the column.

Table Layout, cont'd



29

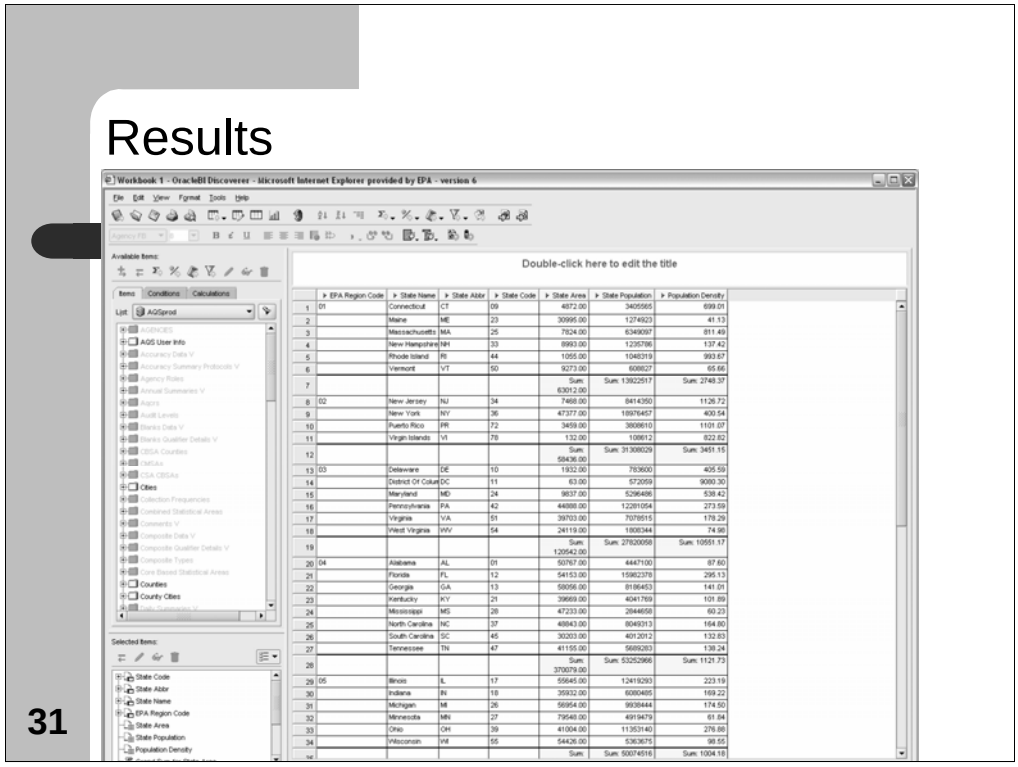
Some other general characteristics of the layout can be set from the Table Format tab of the Worksheet Properties. Click on the Properties button to bring it up. The settings shown are the defaults except for "Show row numbers".



Sorting is specified in Step 4 of the Workbook Wizard. To Add a Sort, click on Add and you will be given a new sort line to specify the sort column.

For a group sort, a given value of the sort column only appears in its first row.

If you have previously specified a group total by a column, Discoverer will automatically include a group sort for that column, e.g., the group sort for EPA Region Code.

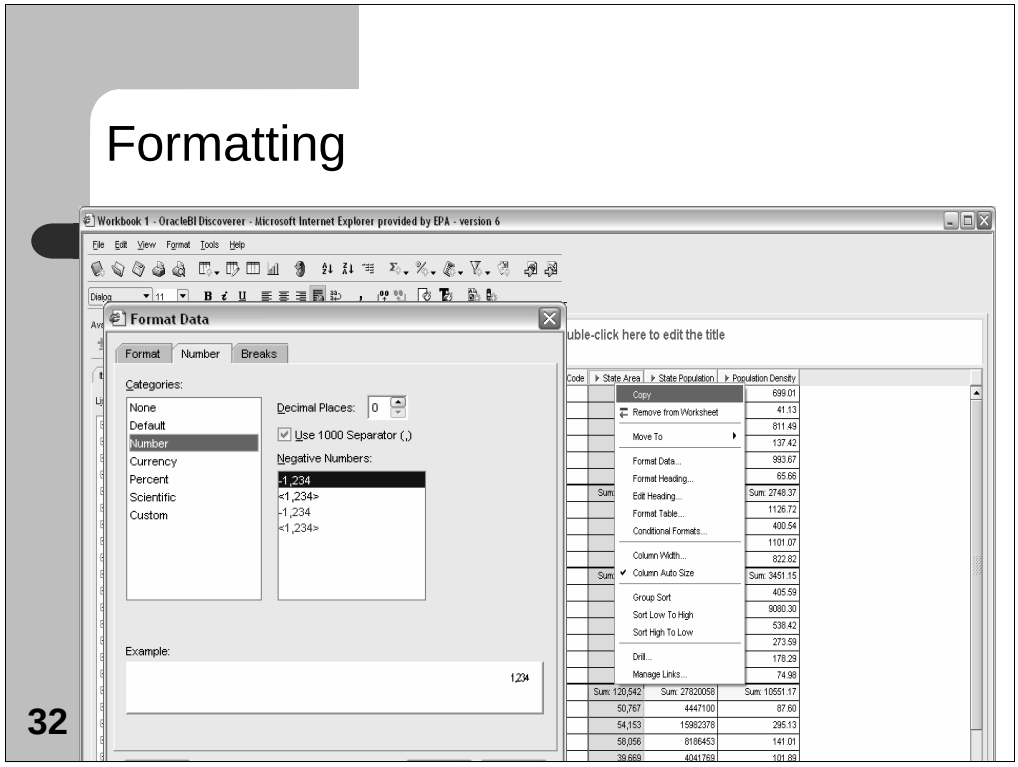


From the Workbook Wizard, click on Finish to run the query you've set up.

Here are the results of the query, with the EPA Region Code condition disabled, i.e, all rows in the States table were selected.

The Available Items pane and the Selected Items pane appear on the left. They can be turned off from the View menu, if desired. From the Available Items pane, you can duplicate the function of Step 2 of the workbook wizard. You can select and remove items, and add/remove/edit conditions and calculations from the Items, Conditions, and Calculations tabs.

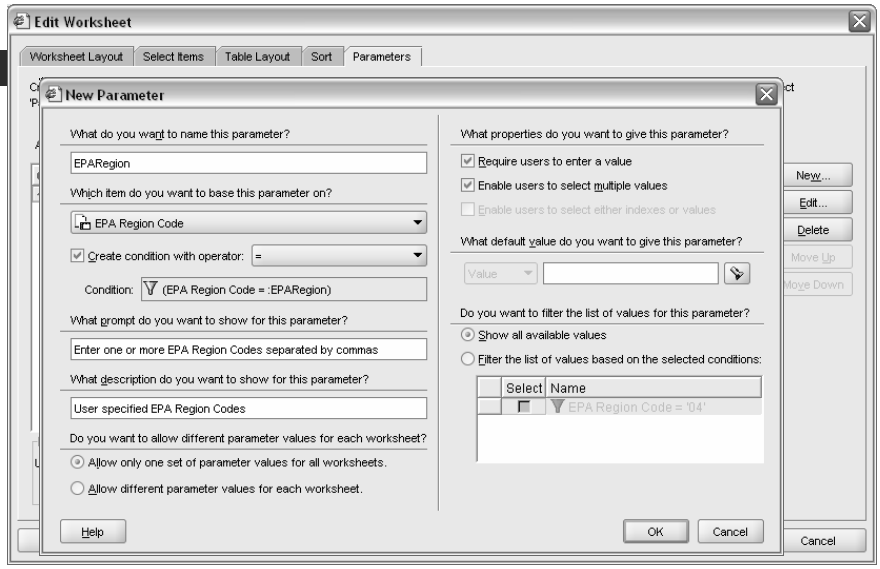
From the results pane, you can change the order of columns by dragging and dropping. And, you can drag and drop items to and from the Available Items pane.



From the results pane, you can right click on a column or cell and modify various aspects. For example, you can change the formatting of the data. Depending on the type of data (character, numeric, date, etc.) you will have different options for the format. Here, the format for State Area data is changed to include commas.

Note the other things you can do from this drop down menu. If you click in a data cell, only the options that apply to the data will be available.

Parameters

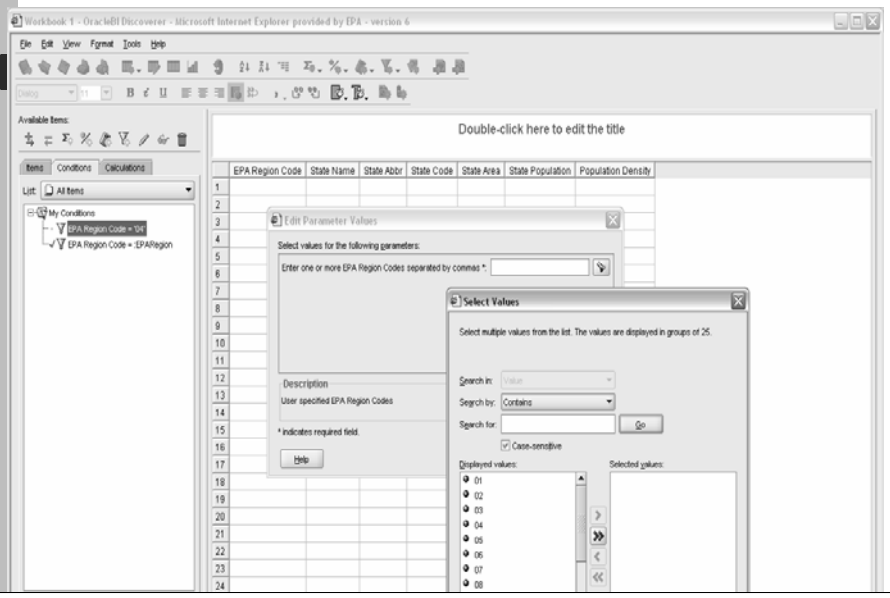


33

Parameters can be used to generalize a query by defining conditions that are based on user input. Parameters are Step 5 of the workbook wizard (skipped earlier). To get to them after the fact, select Edit/Worksheet and click on the Parameters tab.

Here, parameter EPARegion is defined. When the query is run, you will be prompted for one or more EPA Region codes (note the “Enable users to select multiple values” option.) Discoverer did most of the work. The condition is created along with the parameter and the default options were used.

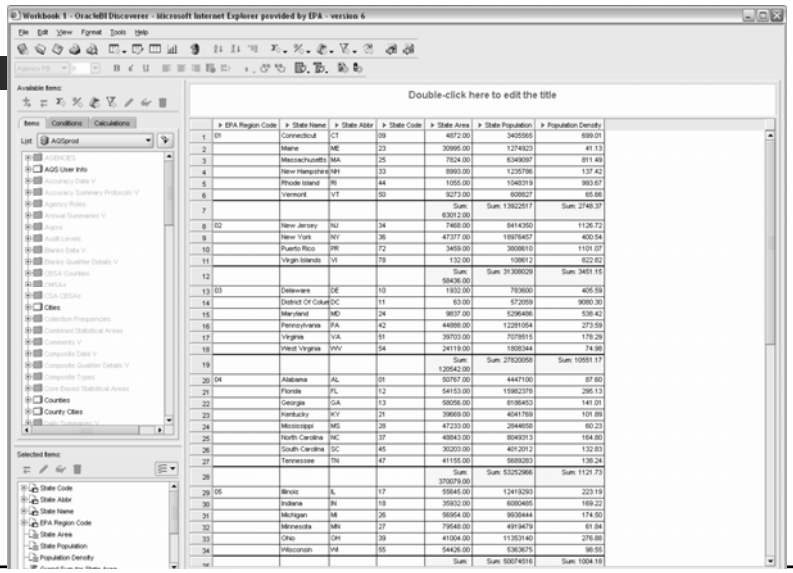
Parameters, cont'd



34

After defining the parameter, running the query prompts for values of Region Code. You can enter the values, or click on the Find icon (flashlight) to select from a list of values, as has been done here. The right double arrows will move all items to Selected Values; the double left arrows will remove all selected values.

The Same Results



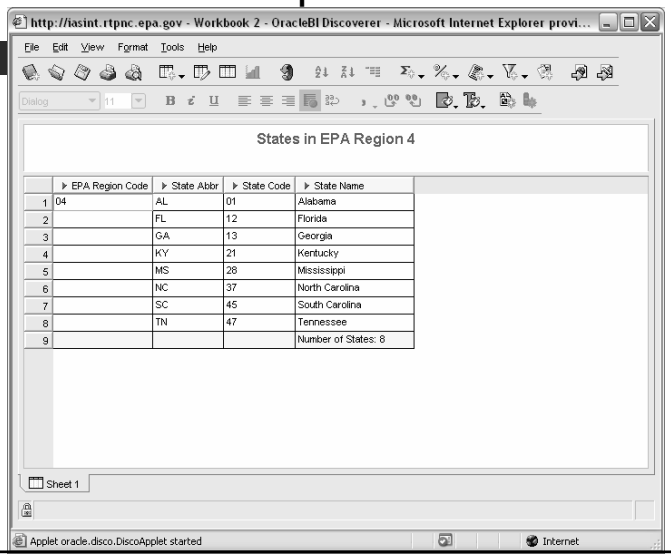
35

Exercise 1

1. Connect to Discoverer.
2. Create a new workbook using the AQS basic business area to show all states within your EPA region.
3. Rearrange the order of columns. Make EPA Region column 1.
4. Sort by EPA Region, then State Abbreviation. Change to a group sort on EPA Region.
5. Center headings for all columns. (Extra credit if you can find where to do this for all headings at once.)
6. Add a title.
7. Add a count of states by Region.

36

Exercise 1: Sample Results



37

Exercise 2

1. List all the parameters (pollutants) measured at site 06-001-0007
2. Modify #1 to include only parameters being sampled since 2003
3. Modify #1 to show sampling end dates

38

Exercise 2 Results

All parameters at site

	State Code	County Code	Site ID	Parameter Code
1	06	001	0007	42101
2	06	001	0007	42601
3	06	001	0007	42602
4	06	001	0007	43101
5	06	001	0007	43201
6	06	001	0007	44201
7	06	001	0007	68101
8	06	001	0007	68102
9	06	001	0007	68103
10	06	001	0007	68105
11	06	001	0007	68106
12	06	001	0007	68107
13	06	001	0007	68108
14	06	001	0007	68109
15	06	001	0007	81102
16	06	001	0007	81102
17	06	001	0007	82301
18	06	001	0007	82306
19	06	001	0007	82403
20	06	001	0007	85101
21	06	001	0007	68104
22	06	001	0007	82203
23	06	001	0007	88101

All parms since 2003

	State Code	County Code	Site ID	Parameter Code
1	06	001	0007	42101
2	06	001	0007	42601
3	06	001	0007	42602
4	06	001	0007	43101
5	06	001	0007	43201
6	06	001	0007	44201
7	06	001	0007	68101
8	06	001	0007	68102
9	06	001	0007	68103
10	06	001	0007	68105
11	06	001	0007	68106
12	06	001	0007	68107
13	06	001	0007	68108
14	06	001	0007	68109
15	06	001	0007	81102
16	06	001	0007	82301
17	06	001	0007	82306
18	06	001	0007	82403
19	06	001	0007	85101
20	06	001	0007	68104
21	06	001	0007	82203
22	06	001	0007	88101

All parms w/end date

	State Code	County Code	Site ID	Parm Code	Date Sampling Ended
1	06	001	0007	42101	
2	06	001	0007	42601	
3	06	001	0007	42602	
4	06	001	0007	43101	
5	06	001	0007	43201	
6	06	001	0007	44201	
7	06	001	0007	68101	
8	06	001	0007	68102	
9	06	001	0007	68103	
10	06	001	0007	68105	
11	06	001	0007	68106	
12	06	001	0007	68107	
13	06	001	0007	68108	
14	06	001	0007	68109	
15	06	001	0007	81102	31-MAY-2000
16	06	001	0007	81102	
17	06	001	0007	82301	
18	06	001	0007	82306	
19	06	001	0007	82403	
20	06	001	0007	85101	
21	06	001	0007	68104	
22	06	001	0007	82203	
23	06	001	0007	88101	

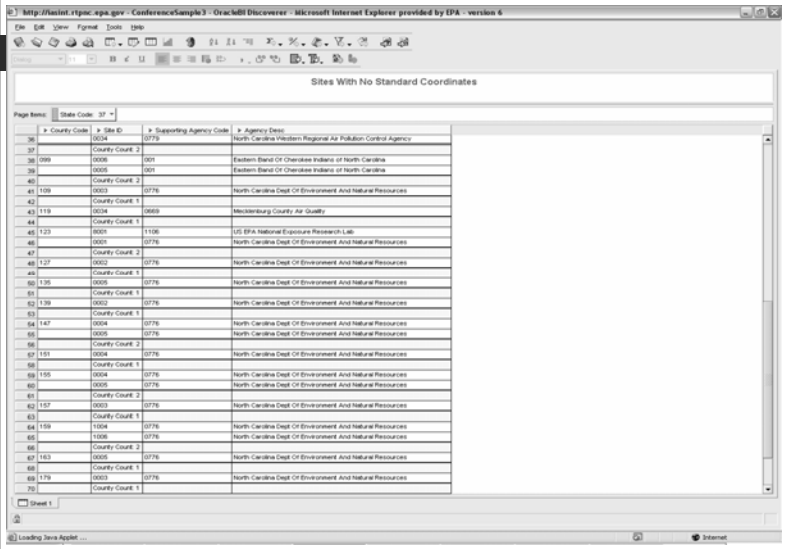
39

Exercise 3

1. Identify all active sites in your state that do not have the standard coordinate fields populated.
2. Add a title.
3. Include a count of the number of sites by county and a grand total for the state.

40

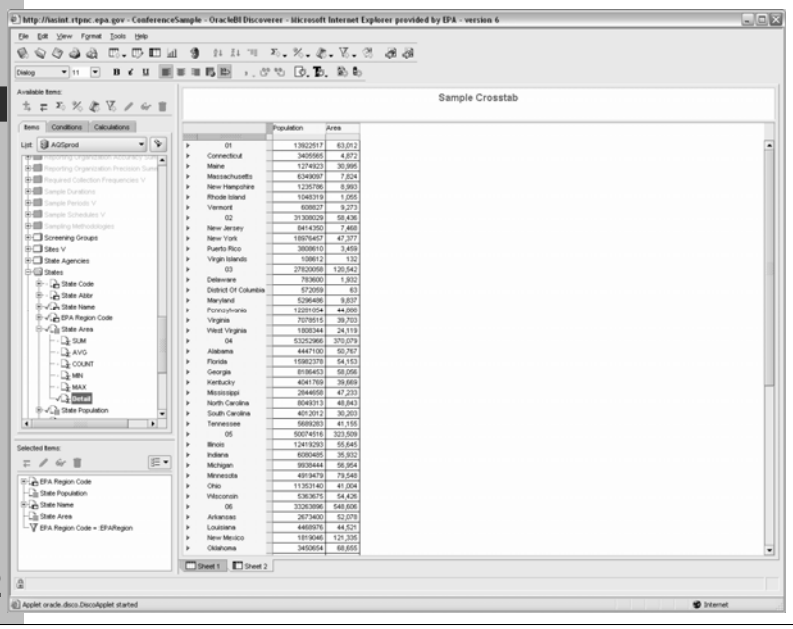
Exercise 3 sample results



41

Other Views -

Crosstab



42

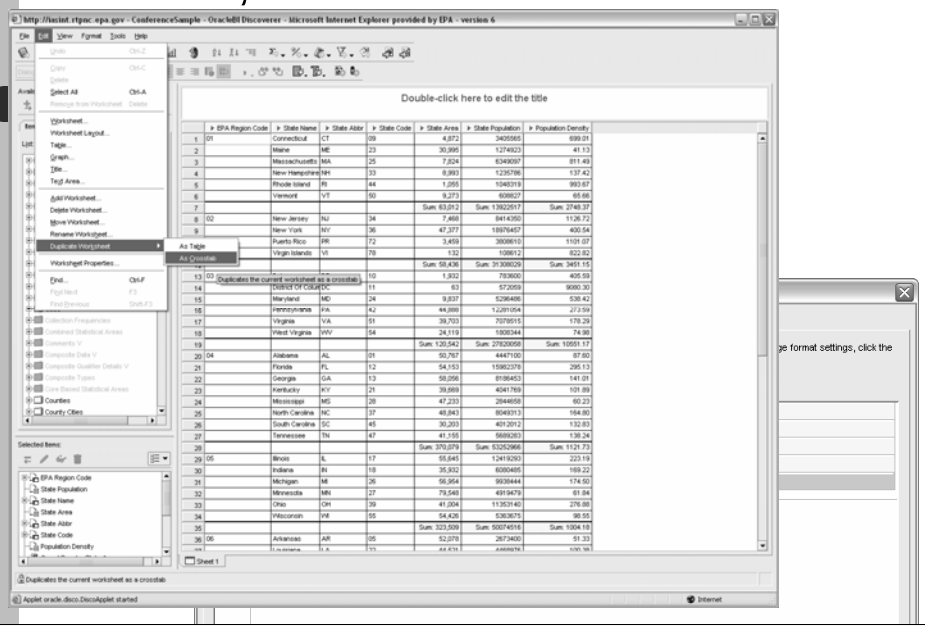
Crosstabs typically summarize the relationship of two sets of data in terms of a third set of data. For example, suppose you have several sites, each with an ozone, PM 2.5 FRM, and an FRM like continuous PM 2.5 monitor. The sites are one set of data, the parameters the second, and the third could be some annual statistics for the monitors.

Pictured here is a trivial example of a crosstab – trivial in the sense that's it's missing one of the sets of data and is essentially just a table.

The next couple of slides illustrate how it was created from the States/EPA Regions worksheet.

Other Views -

Crosstab, cont'd

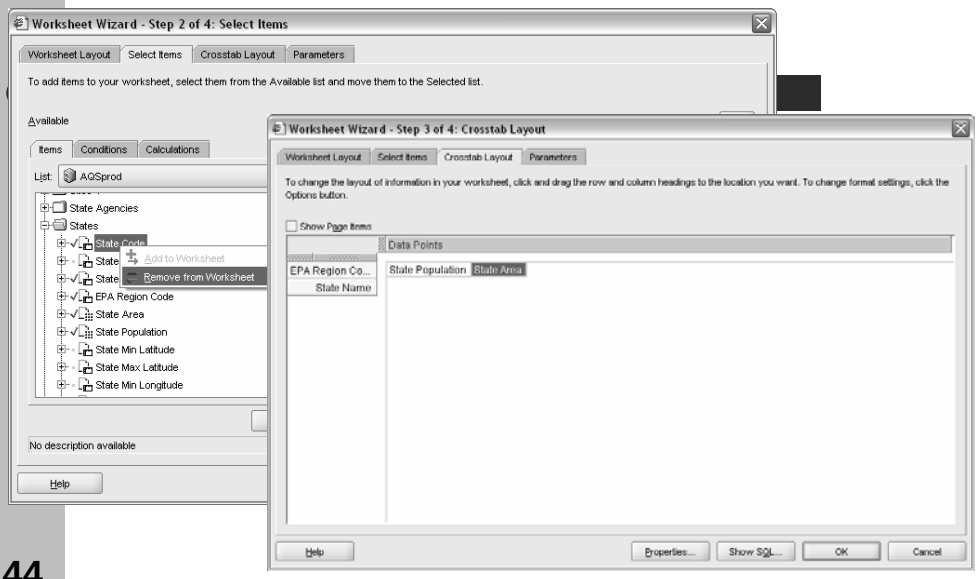


43

From the original worksheet of states by EPA Region, copy it as a crosstab by selecting Edit/Duplicate Worksheet/As Crosstab. You'll get the warning pictured here, because a table converted to a crosstab will initially have no rows.

Other Views -

Crosstab, cont'd

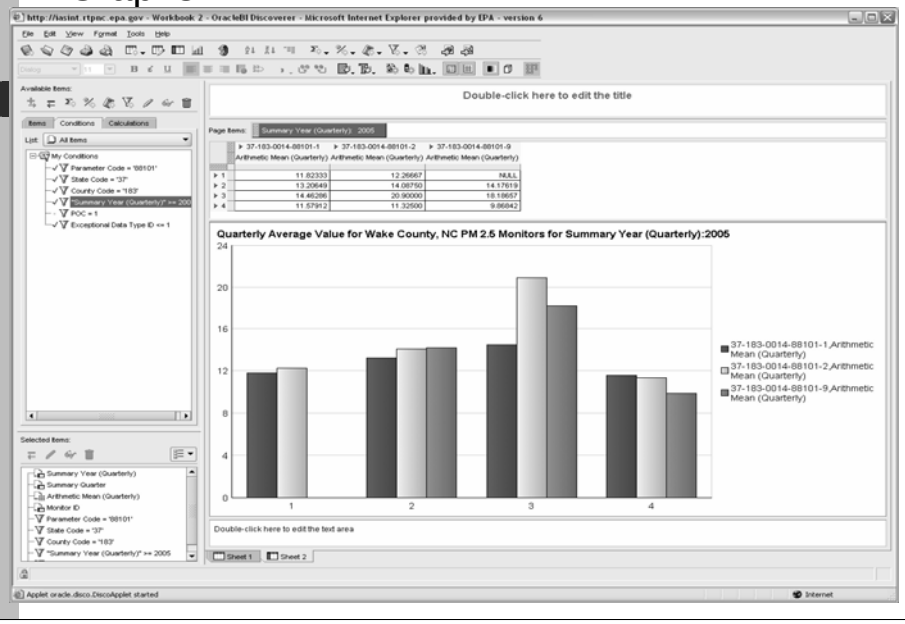


44

All items except EPA Region Code, State Name, State Population, and State Area have been removed. EPA Region Code and State Name have been made row items.

Running the query yields the results previously shown. Population and Area are summed for each EPA Region Code.

Graphs

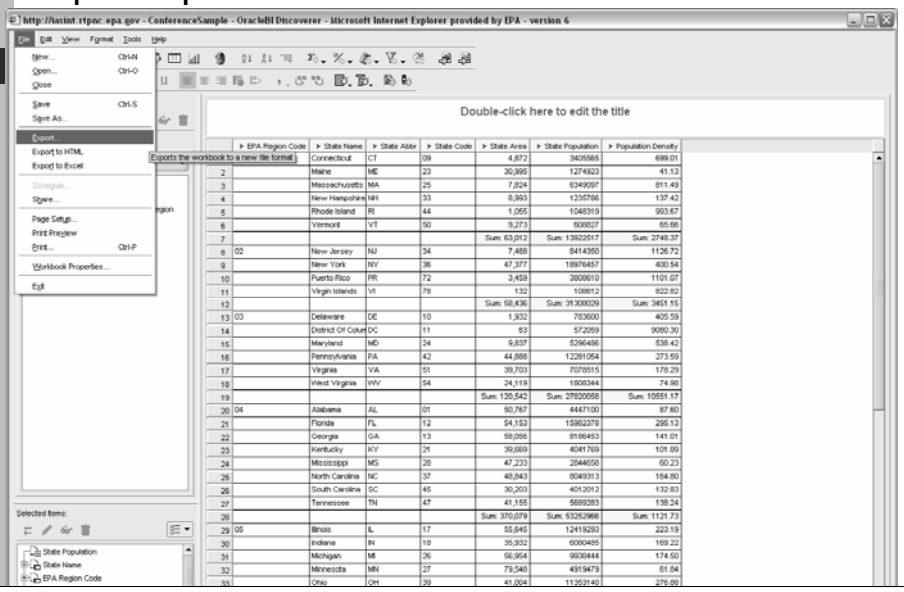


45

Here is a sample query with a graph. It also has a more representative example of a crosstab.

Check out all the options for customizing the graph. You can get to them from Edit/Graph or by right clicking on the graph and selecting Edit Graph.

Export Options

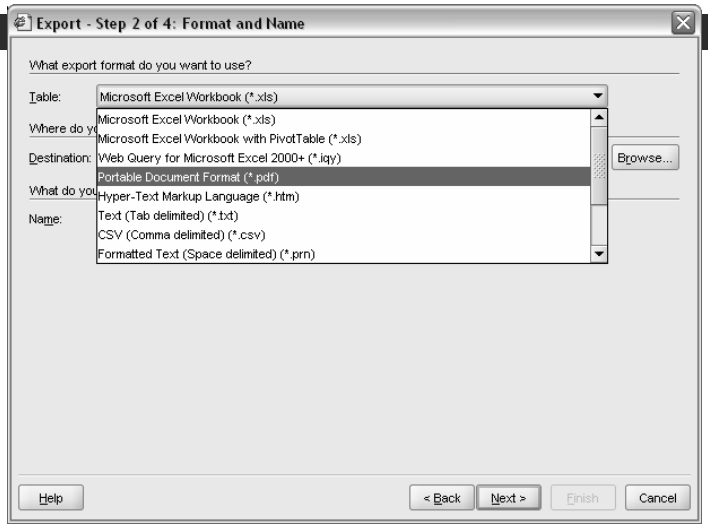


46

To start the Export process, select File/Export. You can use the "Export to HTML" and "Export to Excel" selections for those particular formats. They will write a file to your default user directory, Documents and Settings, and you will not have a chance to specify the name. Using Export starts up the Export Wizard, which gives you more output types and allows explicit control of the destination directory and file name.

The last two steps of the export wizard complete the process and are not pictured here.

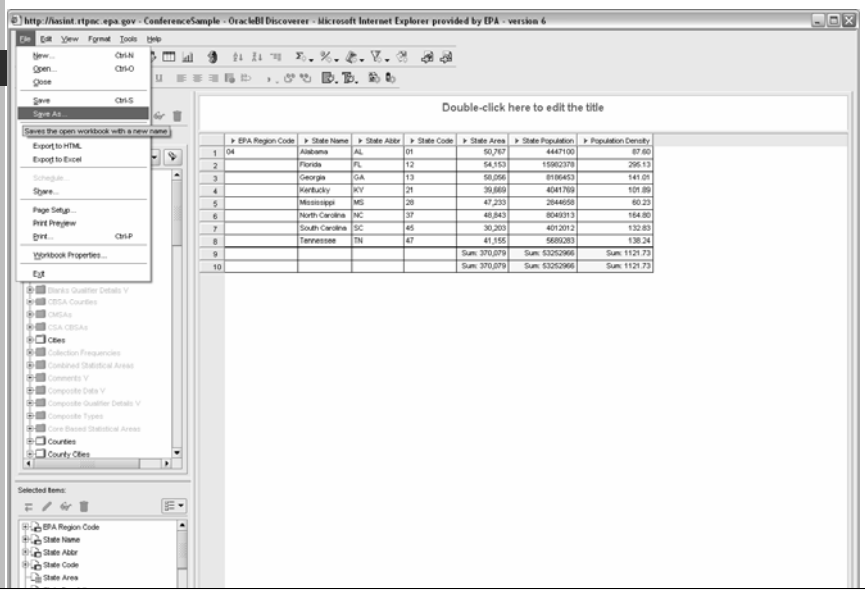
Other Output Options



47

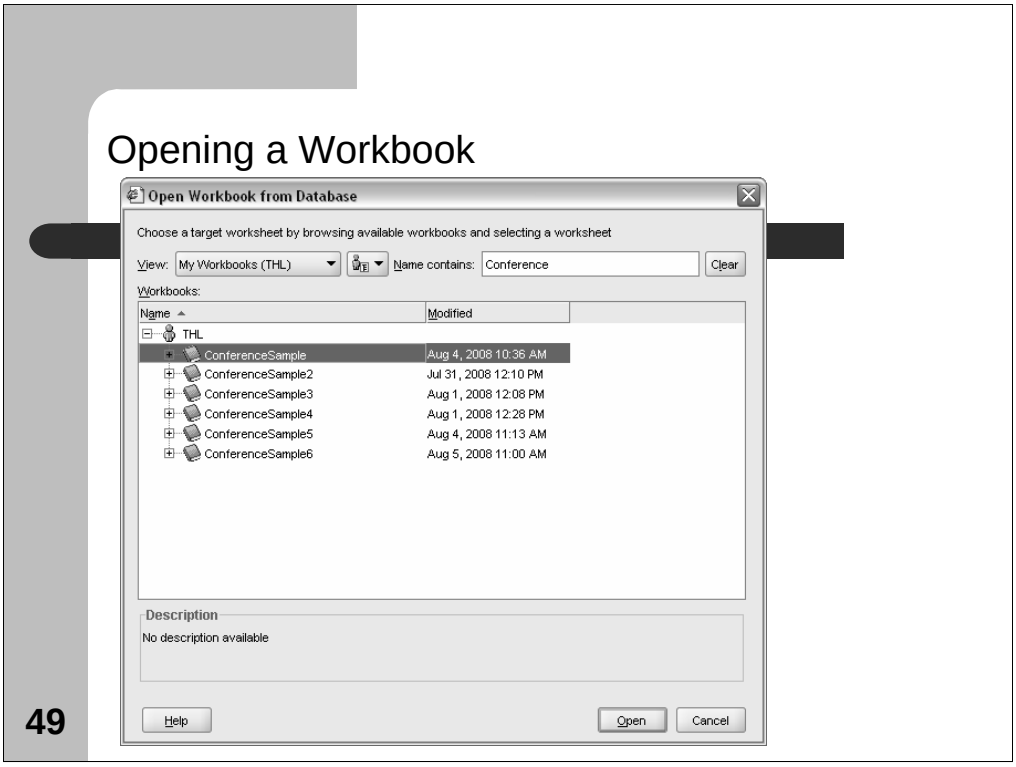
File/Export gives you more options for the output format.

Saving a Workbook



48

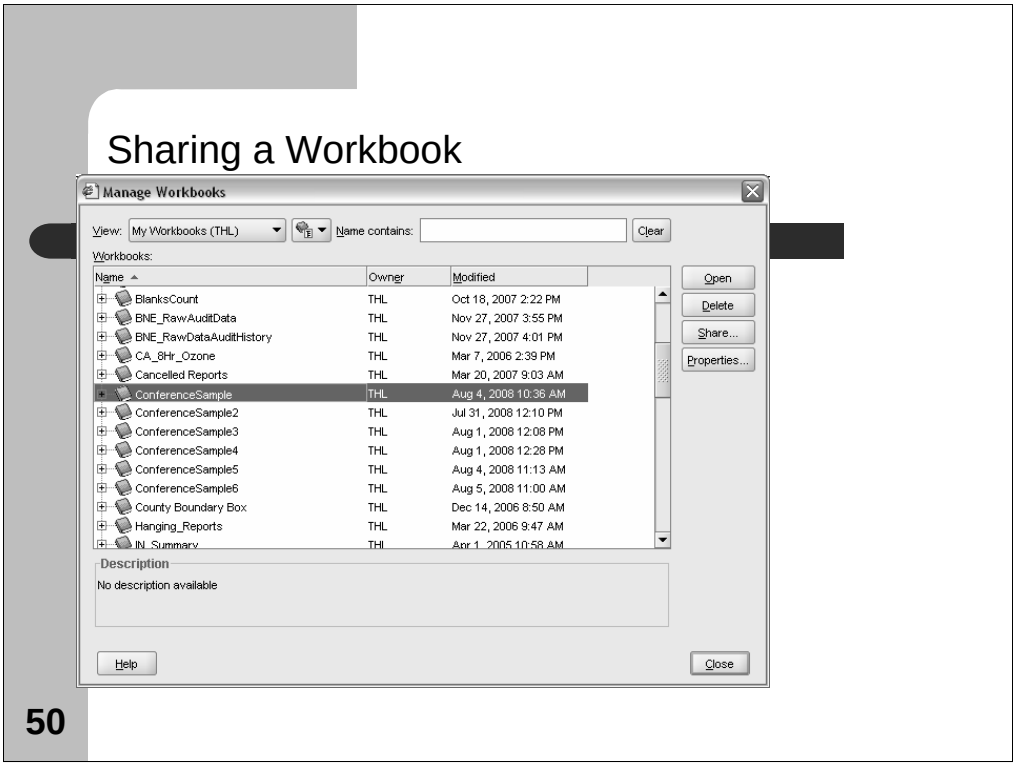
To save your workbook, select File/Save As, specify a name, and Save.



You can open a Workbook by selecting “Open an existing workbook” from the Workbook Wizard or from File/Open.

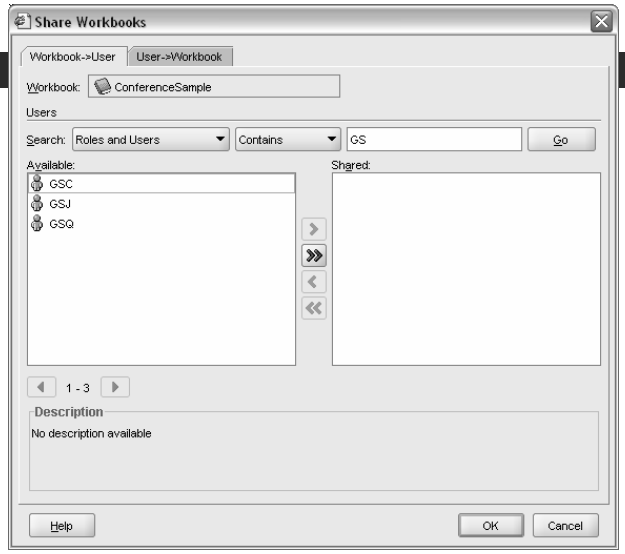
View allows you to select whether or not to see only your own workbooks “My Workbooks” or all workbooks that you have access to. The button beside View allows you to view workbooks by User (“User tree”) or alphabetical by workbook name (“Workbook Tree” or “Table”). You can also filter by entering a character string that must be contained in the name of the workbook. If you’re in User Tree mode, Discoverer does have the annoying habit of collapsing at the user/role level when you start entering a new “Name contains” string, so you’ll have to expand to see the selected workbooks.

The slide depicts all workbooks for user THL that contain “Conference”. You would just click on Open to open the selected workbook.



To share a workbook, select Tools/Manage Workbooks to display the dialog shown here. Select the workbook and click on Share.

Sharing a Workbook, cont'd



51

After you've clicked on Share on Manage Workbooks, you need to select who you want to share your workbook with, by userid. Unless you're looking for userid AAA, you'll likely want to use the Search feature. There are a lot of AQS users, and you only get 100 at a time. To filter the userids you see, type in the search string and click on Go. There are several, more specific, options other than Contains. Check them out from the dropdown if you like.

The above shows all users with GS in their userid. To share your workbook with one or more of them, select them and move them to the Shared pane. The double arrow will move them all. You can perform other searches to add additional users. Once you've populated the Shared pane, click on OK and you've shared your workbook.

If you want to share the active workbook, provided it has been saved, you can select File/Share to get to the Share Workbooks dialog.

If you want to share multiple workbooks with a single user, check out the "User=>Workbook" tab.

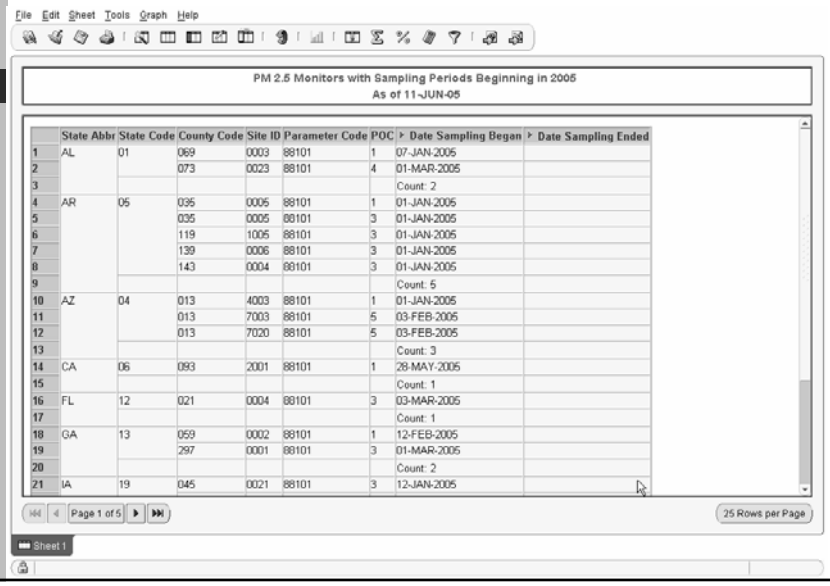
Using More Features

Goal:

1. Find all PM 2.5 monitors with sampling periods beginning in 2005
2. Sort the results by state abbreviation, county, site, and POC
3. Display start date and end date (if any) for sampling periods
4. Display count of monitors in each state.
5. Add an appropriate title
6. Save the workbook on the database
7. Share your workbook with your neighbor

52

Desired Results



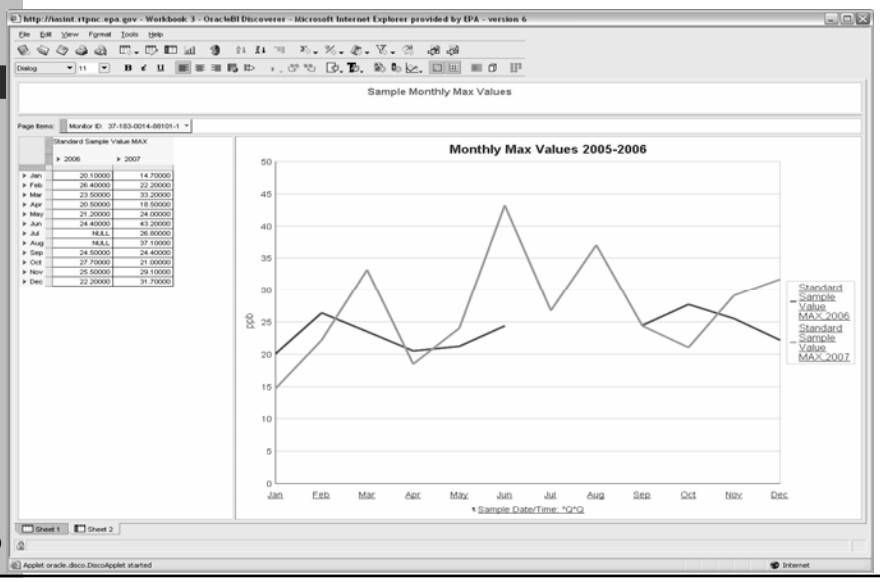
	State Abbr	State Code	County Code	Site ID	Parameter Code	POC	Date Sampling Began	Date Sampling Ended
1	AL	01	069	0003	88101	1	07-JAN-2005	
2			073	0023	88101	4	01-MAR-2005	
3							Count: 2	
4	AR	05	035	0005	88101	1	01-JAN-2005	
5			035	0005	88101	3	01-JAN-2005	
6			119	1005	88101	3	01-JAN-2005	
7			139	0006	88101	3	01-JAN-2005	
8			143	0004	88101	3	01-JAN-2005	
9							Count: 5	
10	AZ	04	013	4003	88101	1	01-JAN-2005	
11			013	7003	88101	5	03-FEB-2005	
12			013	7020	88101	5	03-FEB-2005	
13							Count: 3	
14	CA	06	093	2001	88101	1	28-MAY-2005	
15							Count: 1	
16	FL	12	021	0004	88101	3	03-MAR-2005	
17							Count: 1	
18	GA	13	059	0002	88101	1	12-FEB-2005	
19			297	0001	88101	3	01-MAR-2005	
20							Count: 2	
21	IA	19	045	0021	88101	3	12-JAN-2005	

53

More Sample Queries . . .

54

Monthly Max Values with Graph



55

Sample Monthly Max, Min, Mean by Monitor

File Edit Sheet Tools Graph Help

Monthly Max, Min and Arithmetic Mean Values
State : '06' | County : '025' | Site : '0005' | Parm : '44201' | Year : '2003, 2004'
Report Created: 11-JUN-05 02:56:09 AM

		Standard Sample Value MAX		Standard Sample Value MIN		Standard Sample Value Arith Mean		Number of Obs	
		2003	2004	2003	2004	2003	2004	2003	2004
Monitor ID	Month								
06-025-0005-44201-1		.10700	.10800	.00200	.00200	.01954	.02519	8309	8326
	Jan	.04300	.03900	.00200	.00200	.01067	.01174	682	707
	Feb	.06400	.04500	.00200	.00200	.01769	.01846	659	659
	Mar	.06100	.07400	.00200	.00200	.02194	.02528	712	700
	Apr	.06600	.07300	.00200	.00200	.03203	.03770	687	680
	May	.10000	.08400	.00200	.00200	.03198	.03358	712	696
	Jun	.10700	.10400	.00200	.00200	.02840	.02790	682	682
	Jul	.07100	.10800	.00200	.00200	.02381	.02680	709	727
	Aug	.07700	.10000	.00200	.00200	.01918	.02471	695	710
	Sep	.05100	.08100	.00200	.00200	.01468	.02766	680	682
	Oct	.06000	.08000	.00200	.00200	.01227	.02683	703	703
	Nov	.03700	.07100	.00200	.00200	.01022	.02358	683	677
	Dec	.05700	.05200	.00200	.00200	.01116	.01816	705	703

Page 1 of 1 | 25 Rows per Page

Sheet 1 | Monthly Max, Min, Avg

56

Sample Quarterly Max, Min, Mean

File Edit Sheet Tools Graph Help

Quarterly Max, Min and Arithmetic Mean Values
State : '06' | County : '025' | Parm : '44201' | Year : '2004'
Report Created: 11-JUN-05 02:22:16 AM

		Yr	2004																
		Q	Q1				Q2				Q3				Q4				
			Std Sample Value MAX	Std Sample Value MIN	Std Sample Value AVG	Number of Obs	Std Sample Value MAX	Std Sample Value MIN	Std Sample Value AVG	Number of Obs	Std Sample Value MAX	Std Sample Value MIN	Std Sample Value AVG	Number of Obs	Std Sample Value MAX	Std Sample Value MIN	Std Sample Value AVG	Number of Obs
Scmg Grp	Monitor ID																		
CALIFORNIA 2			.07400	.00200	.02050	4126	.12400	.00200	.03766	4122	.10800	.00200	.02892	4212	.08000	.00200	.02262	4143
	06-025-0005-44201-1		.07400	.00200	.01847	2066	.10400	.00200	.03306	2058	.10800	.00200	.02636	2119	.08000	.00200	.02265	2083
	06-025-0006-44201-1		.06500	.00200	.02253	2060	.12400	.00600	.04225	2064	.09300	.00600	.03148	2093	.06500	.00200	.02238	2060
CALIFORNIA 3			.06700	.00200	.02274	8250	.10900	.00200	.04090	8322	.10600	.00200	.02997	8364	.08700	.00200	.02027	7829
	06-025-0004-44201-1		.06900	.00200	.01803	2041	.10500	.00200	.03020	2089	.09400	.00200	.01934	2116	.06000	.00200	.01078	2108
	06-025-1003-44201-1		.06700	.00200	.02203	2033	.09000	.00200	.04657	2056	.09600	.00200	.03697	2050	.08700	.00200	.02657	2074
	06-025-4003-44201-1		.07700	.00200	.01646	2089	.10900	.00200	.04076	2091	.10600	.00200	.03140	2096	.08300	.00200	.02273	1581
	06-026-4004-44201-1		.07900	.00700	.03433	2087	.09200	.00200	.04616	2086	.08300	.00200	.03241	2102	.08100	.00200	.02176	2066

Page 1 of 1 | 25 Rows per Page

Sheet 1 | Qtrly Max, Min, Avg by Monitor ID

57

Hints and Tips and Other Good Stuff

- Discoverer is accessing the live production database (AQSprod – same as the AQS application)
- Your Discoverer userid and password are the same as your AQS application userid and password
- Only registered AQS users can access AQSprod via Discoverer
- There is one EUL for AQS; multiple Business Areas
- AQS basic has the most commonly used tables
- AQSprod includes all the tables in AQS basic plus raw data – current and a few other related tables
- Online Help is pretty good!
- Any workbook you share with others will have your userid appended to the beginning of the workbook name
- You can get to Discoverer from the AQS Web Application link on the EPA TTN AQS page - http://www.epa.gov/ttn/airs/airsaqs/
- Many of the reference tables are viewable from http://www.epa.gov/ttn/airs/airsaqs/manuals/codedescs.htm

58

Hints and Tips and Other Good Stuff, cont'd

- Max time for Discoverer queries = 60 mins (Query Governor tab from Tools/Options)
- Max number of rows=100,000 (Query Governor tab from Tools/Options)
- Many of the date fields are defined as YYYYMonDD HH:MM – e.g., 2005Jun11 01:22. Many date fields are also defined for just the YYYY, or Q, Mon, or DD, or HH, or MM
- If you can't get a total to work within Discoverer, you may want to export to a spreadsheet to create your totals (and other calculations)
- Oracle Technology Network is a good source for information – free registration – look for Discoverer with Data Warehousing and Business Intelligence (http://www.oracle.com/technology/products/discoverer/index.html)

59

There’s lots you can do!
Discover the data
with ***DISCOVERER***!

60