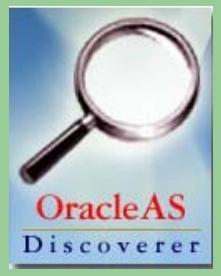
Discoverer with AQS for New Users



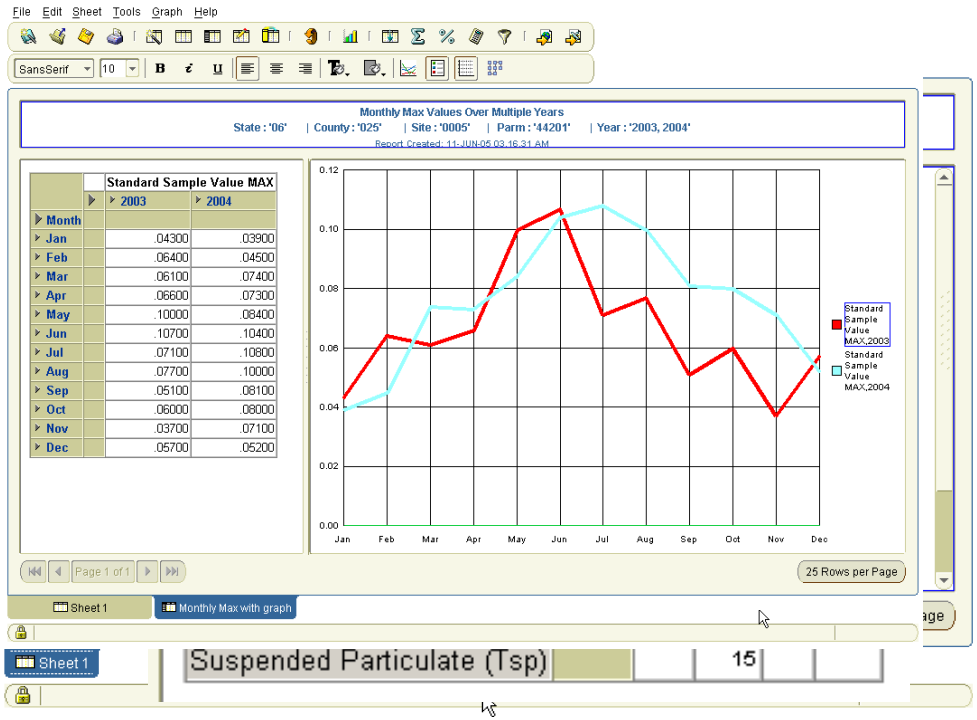
Nick Mangus

Goals for this class

- Become familiar with the mechanics of using Discoverer
- Practice using Discoverer
- Learn about the data in AQS
- Have Fun!

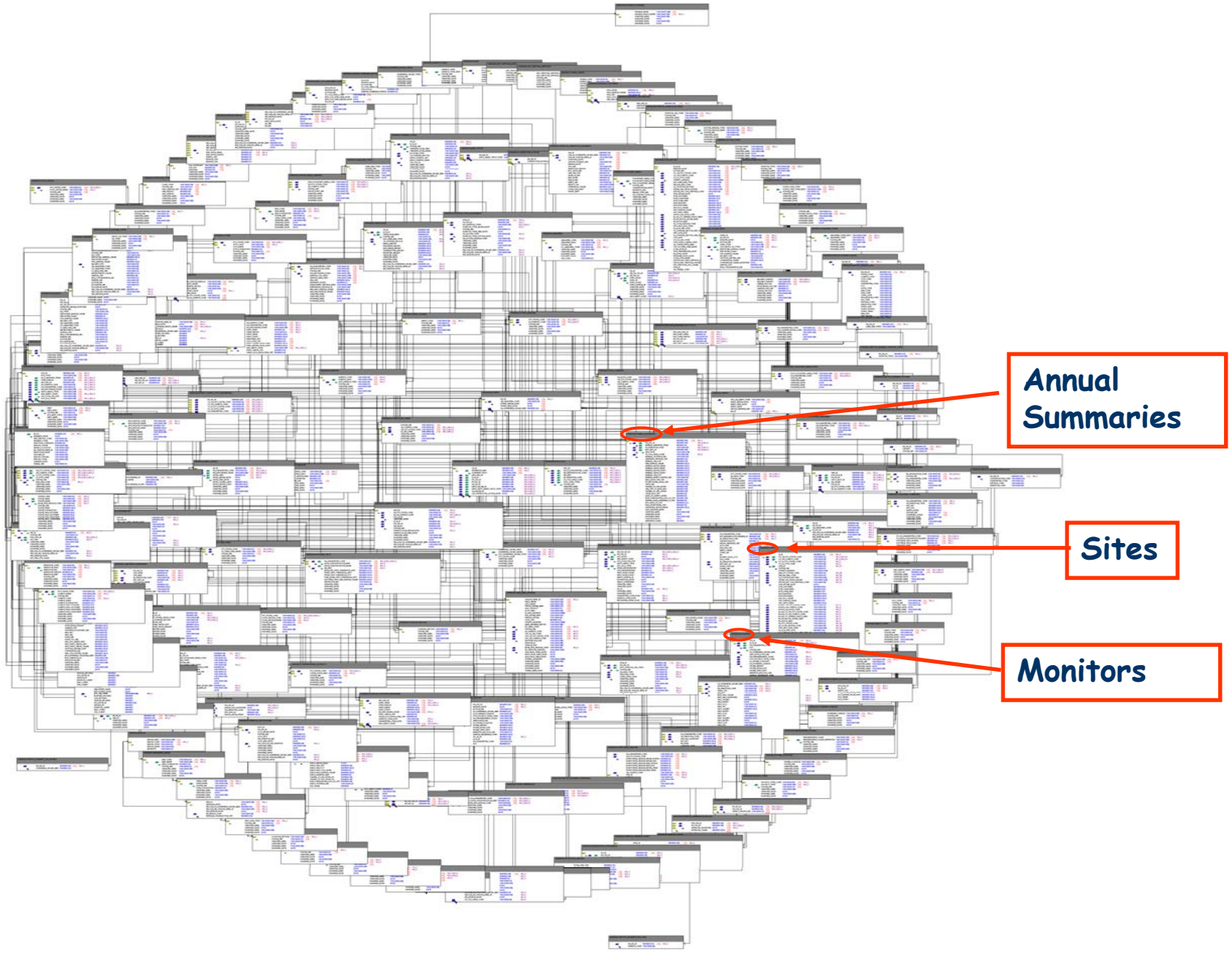
What Can Discoverer Produce?

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

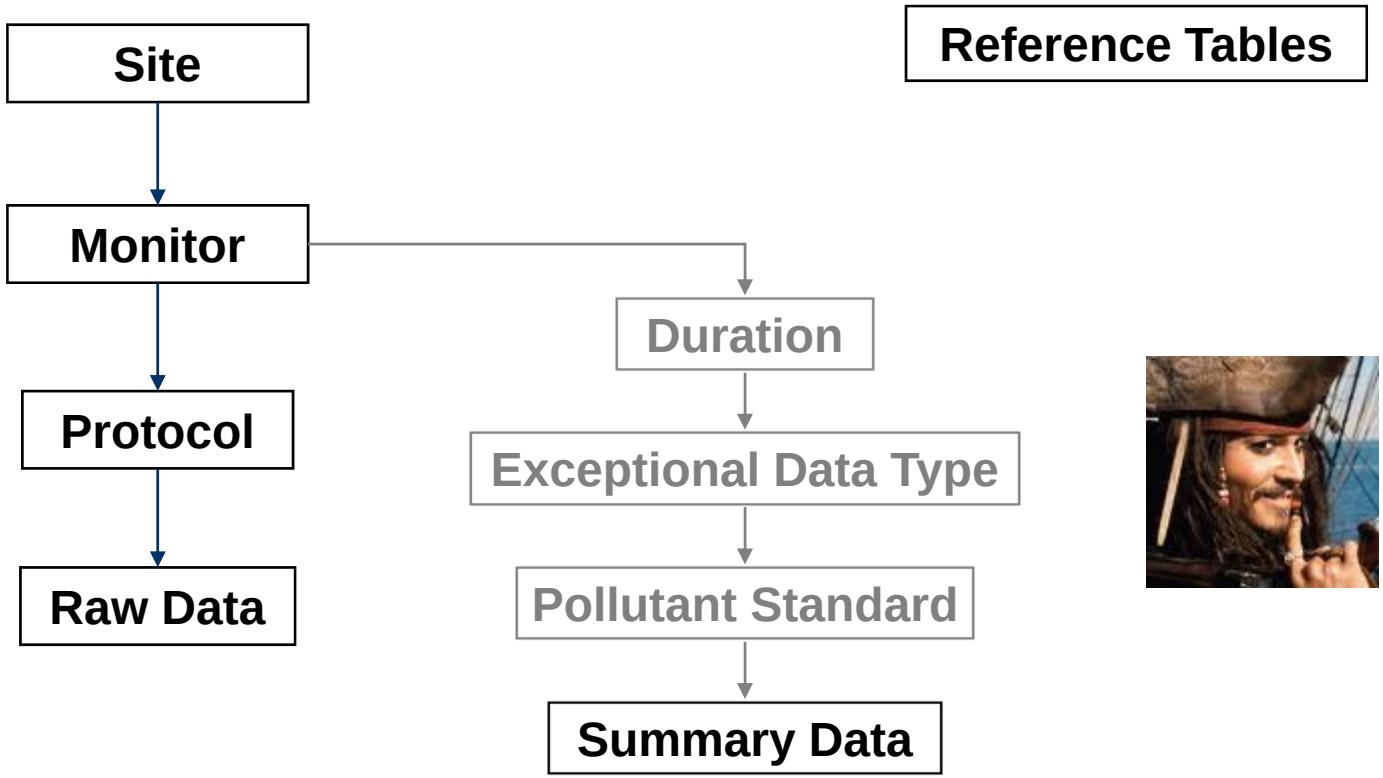


What is Discoverer?

- A business intelligence tool for customizable queries from Oracle Corporation (“OBI”)
- A retrieval only tool that works on a relational database
- Key terms
  - The End User Layer is a filter that makes the database more understandable
  - A Business Area is a grouping of similar data related to a common task
    - You should use only “AQS prod”


Annual Summaries
Sites
Monitors

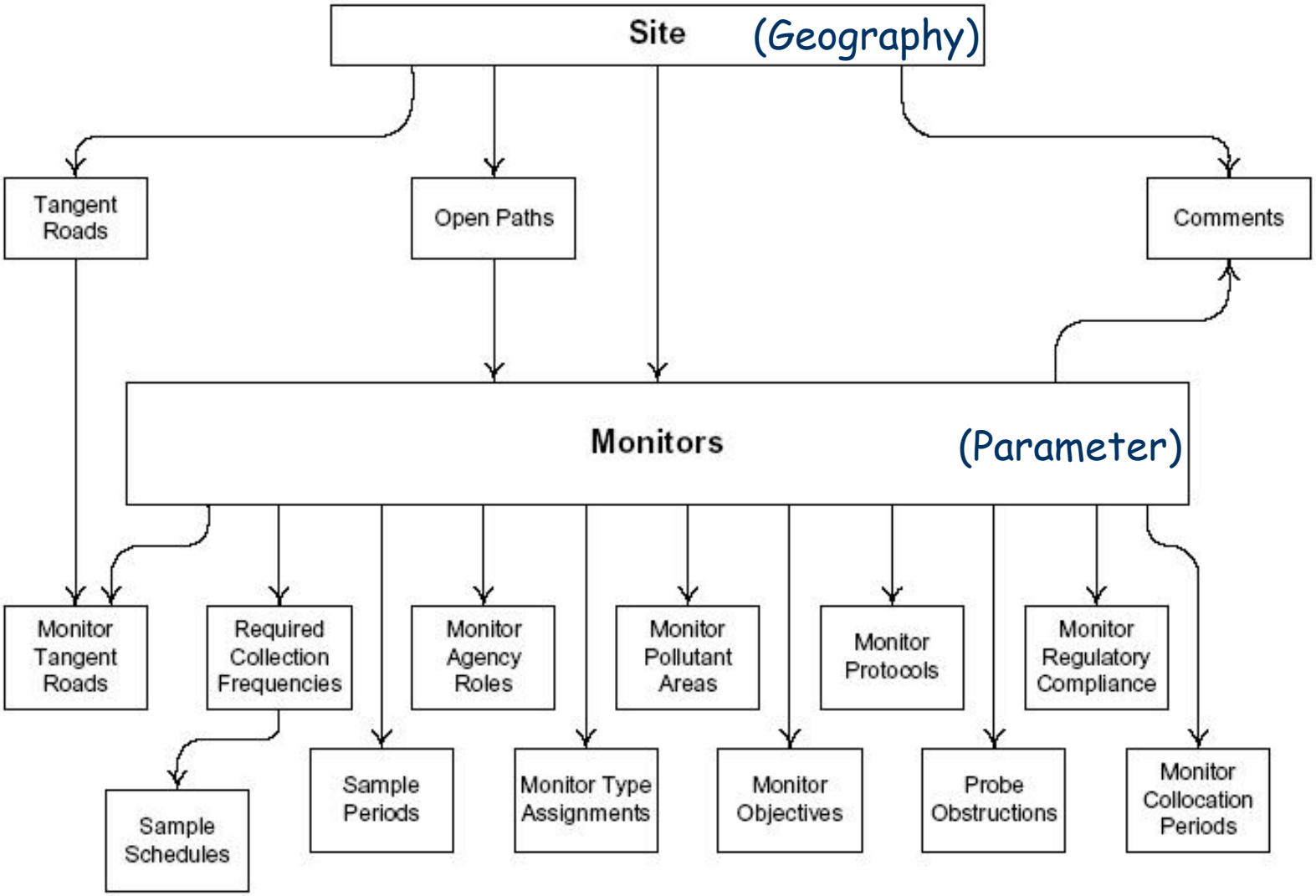
Data Model for AQS (Overview)



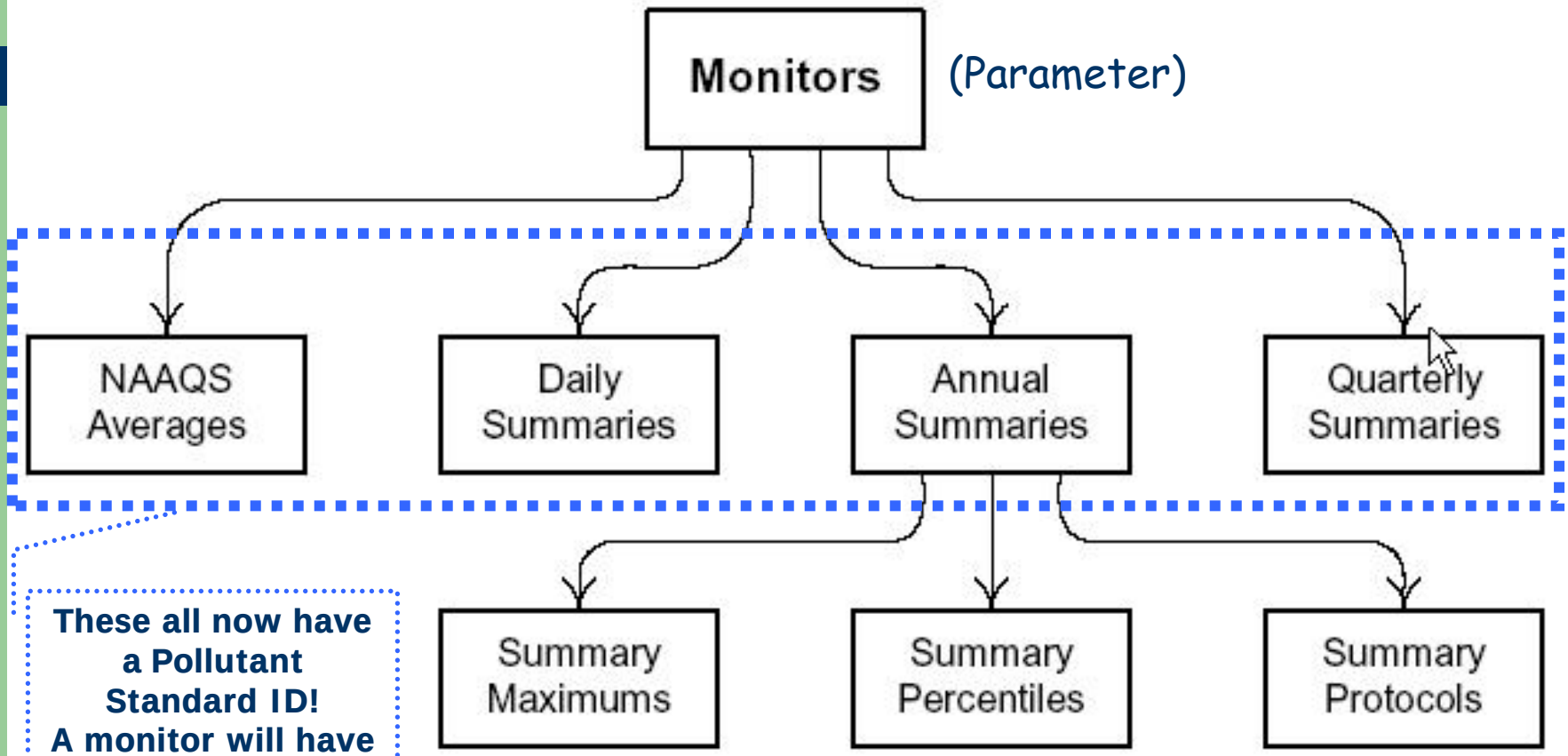
Pollutant Standard IDs

PS_ID	PARAM. CODE	PROMUL. DATE	PRIMARY STD LEVEL	SEC'RY STD LEVEL	ROUND TRUNC IND	SHORT_DESCRIPTION	DESCRIPTION
1	12128	05-Oct-1978	1.500	1.500	2	Lead Quarterly Historical	Original lead standard based on calendar quarterly average.
3	42101	13-Sep-1985	35	35	2	CO 1-hour 1971	Carbon Monoxide 1-hour standard from 1971.
4	42101	13-Sep-1985	9.000	9.000	2	CO 8-hour 1971	Carbon Monoxide 8-hour running average from 1971.
5	42401	22-May-1996		0.500	2	SO2 3-hour 1971	Sulfur Dioxide 3-hour Block Average from 1971.
6	42401	22-May-1996	0.140		2	SO2 24-hour 1971	Sulfur Dioxide 24-hour standard from 1971.
7	42401	22-May-1996	0.030		2	SO2 Annual 1971	Sulfur Dioxide Annual Mean Standard from 1971.
8	42602	19-Jun-1985	0.053	0.053	2	NO2 Annual 1971	Nitrogen Dioxide Annual Mean from 1971.
9	44201	18-Jul-1997	0.120	0.120	2	Ozone 1-hour Daily 2005	Ozone 1-hour Daily Max value based on data complt'ness from 9am to 9pm.
10	44201	18-Jul-1997	0.080	0.080	2	Ozone 8-Hour 1997	Ozone 8-hour running average stored in begin hour from 1997.
11	44201	27-Mar-2008	0.075	0.075	1	Ozone 8-hour 2008	Ozone 8-hour running average stored in begin hour from 2008.
12	81102	01-Jul-1987	150	150	2	PM10 24-hour 2006	PM10 24-hour standard. Violations include actual exceedences & expected excd's where
16	88101	17-Oct-2006	35	35	2	PM25 24-hour 2006	PM25 24-hour standard made more restrictive in 2006.
18	88101	17-Oct-2006	15	15	2	PM25 Annual 2006	PM25 annual wtd mean was reaffirmed in 2006 with same level as originally promulgated.

Data Model for AQS (Metadata)

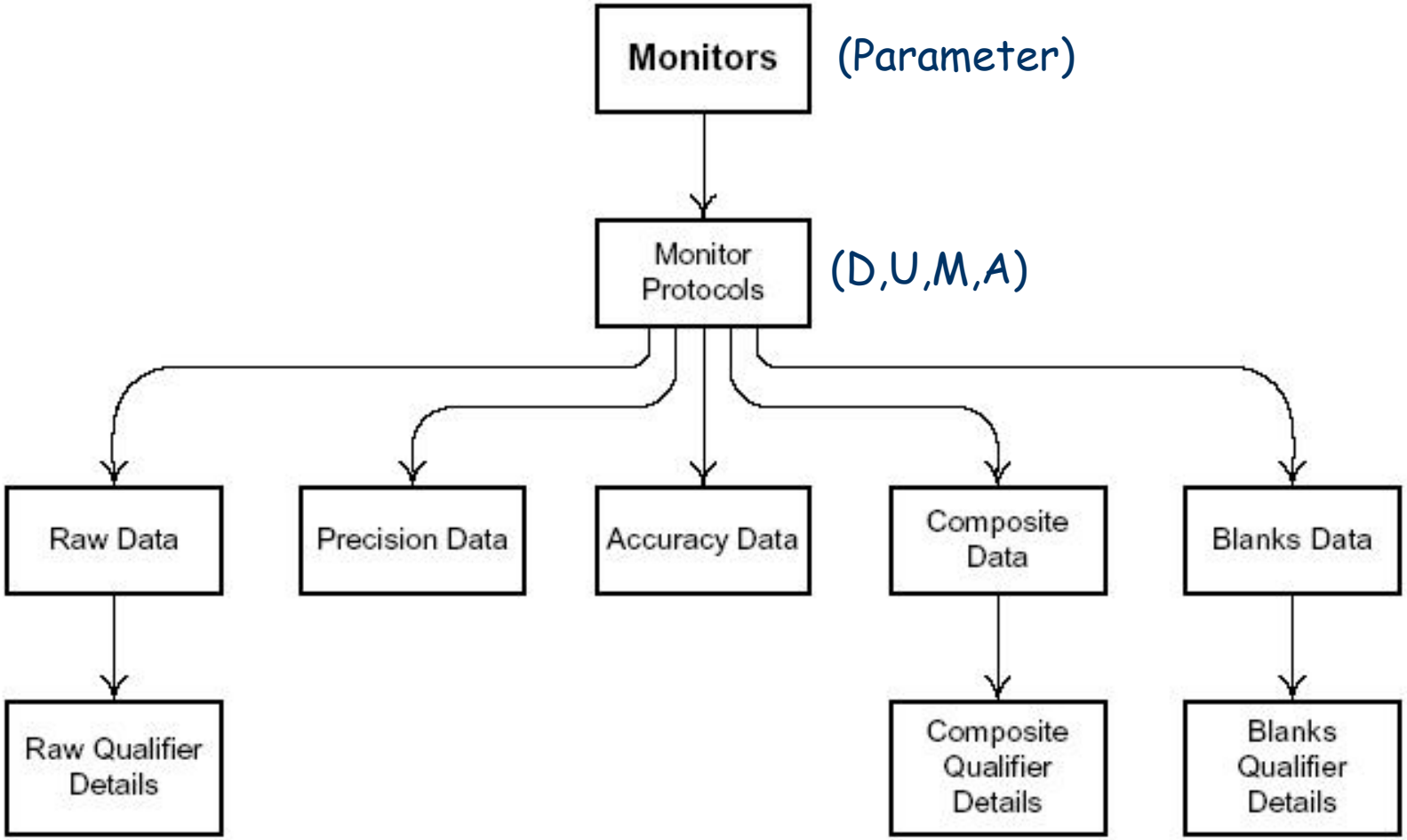


Data Model for AQS (Summary Data)



These all now have a Pollutant Standard ID!
A monitor will have more than one annual summary for the same duration + events!

Data Model for AOS (Measured Data)



How To Build a Query in Discoverer

- Connect to Discoverer (log on)
- Choose an Output Format
- Select Data
  - To Display
  - To Filter By (Conditions) – NOT Optional
- Specify Calculations and Totals (Optional)
- Adjust the Layout (Optional)
- Specify Sorting (Optional)
- Specify Run-Time Parameters (Optional)
- Run Report

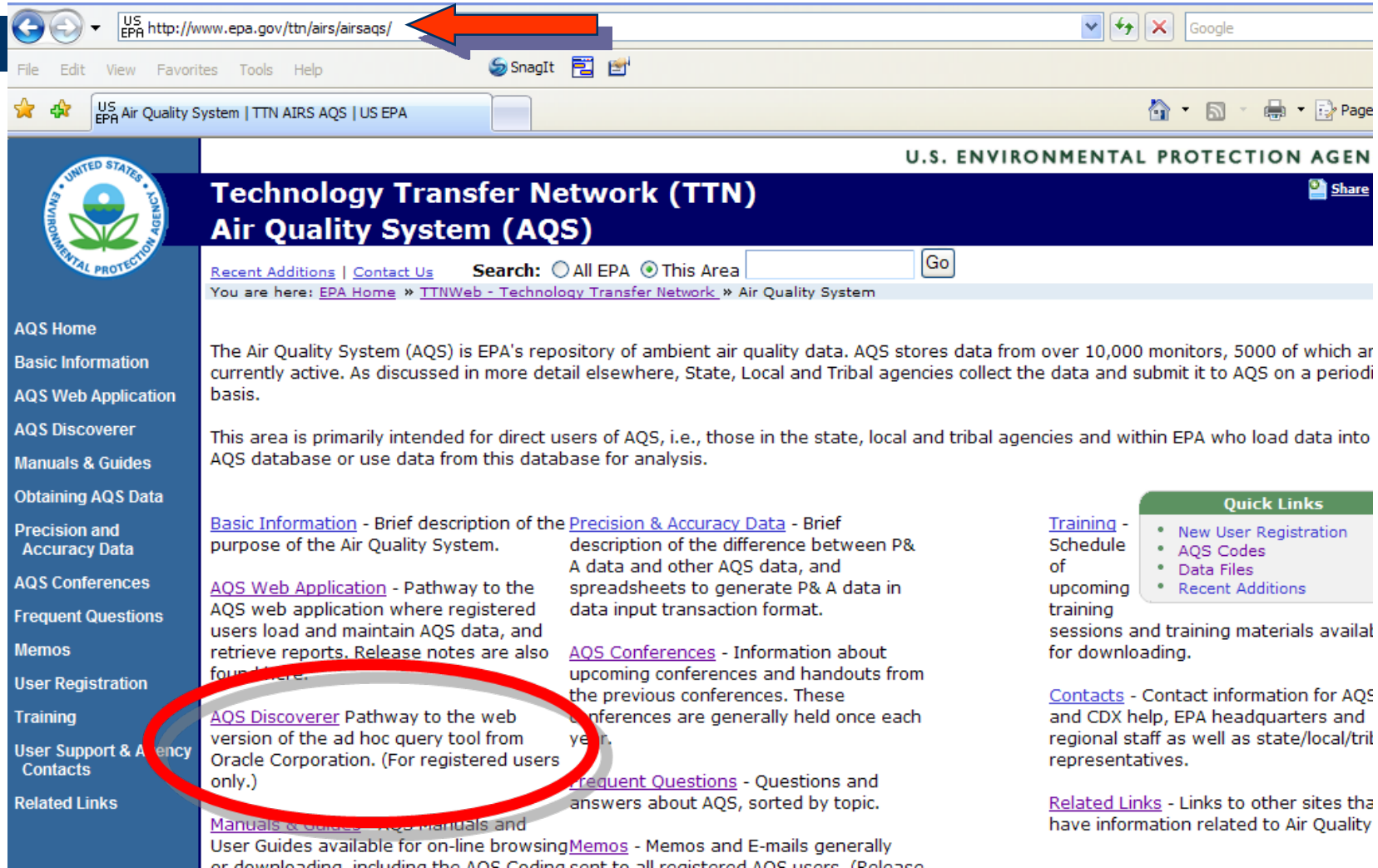
The REAL Way to Build a Query…

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

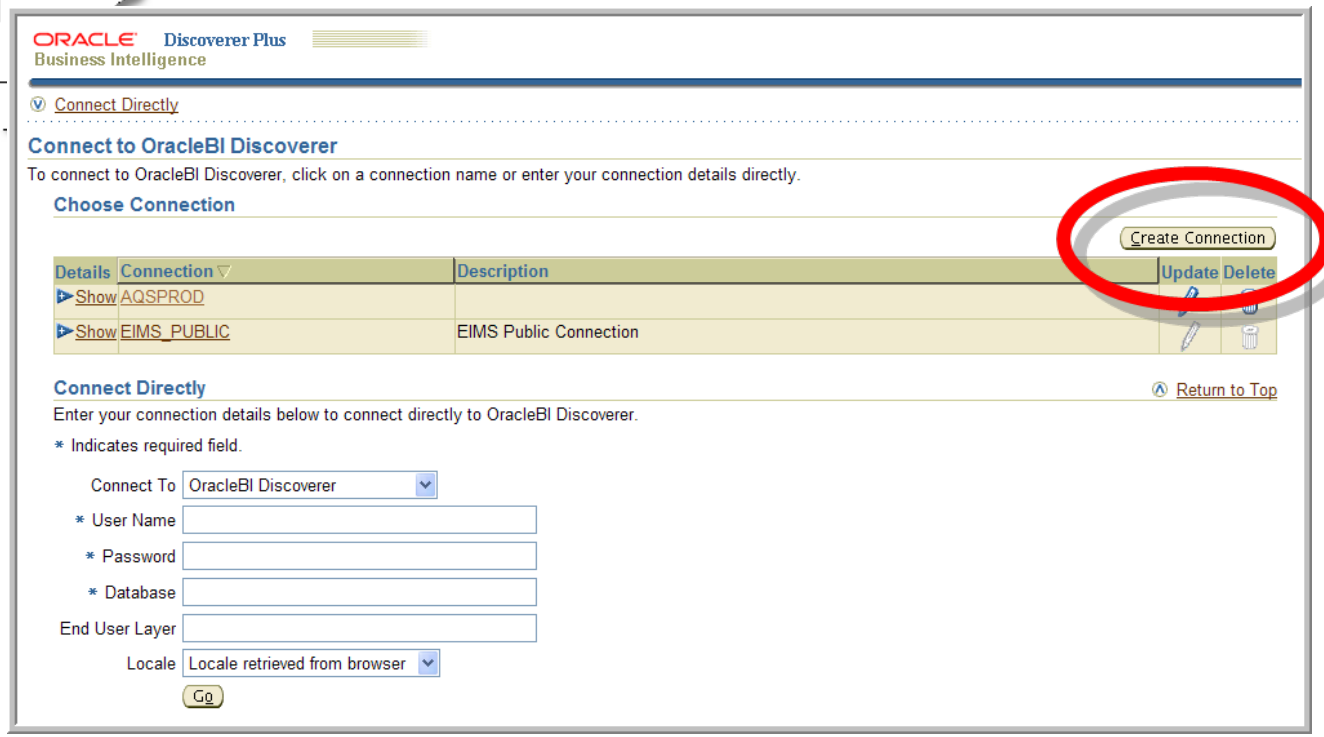
What you need to Run Discoverer

- AQS user ID and password
- Web browser
- Java (Virtual Machine)
- URL: http://www.epa.gov/ttn/airs/airsaqs/aqsdiscover/

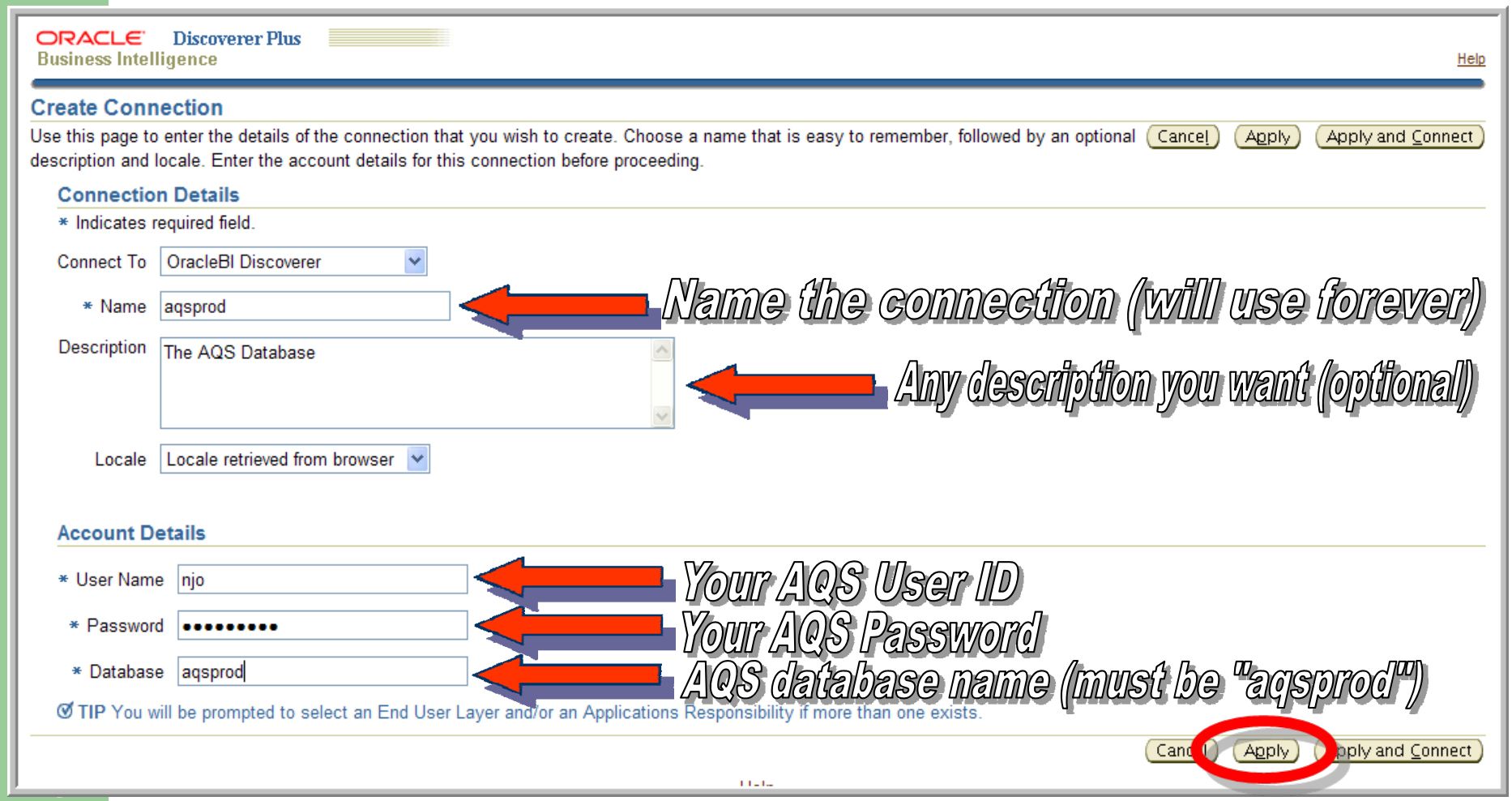
Let's Light This Candle



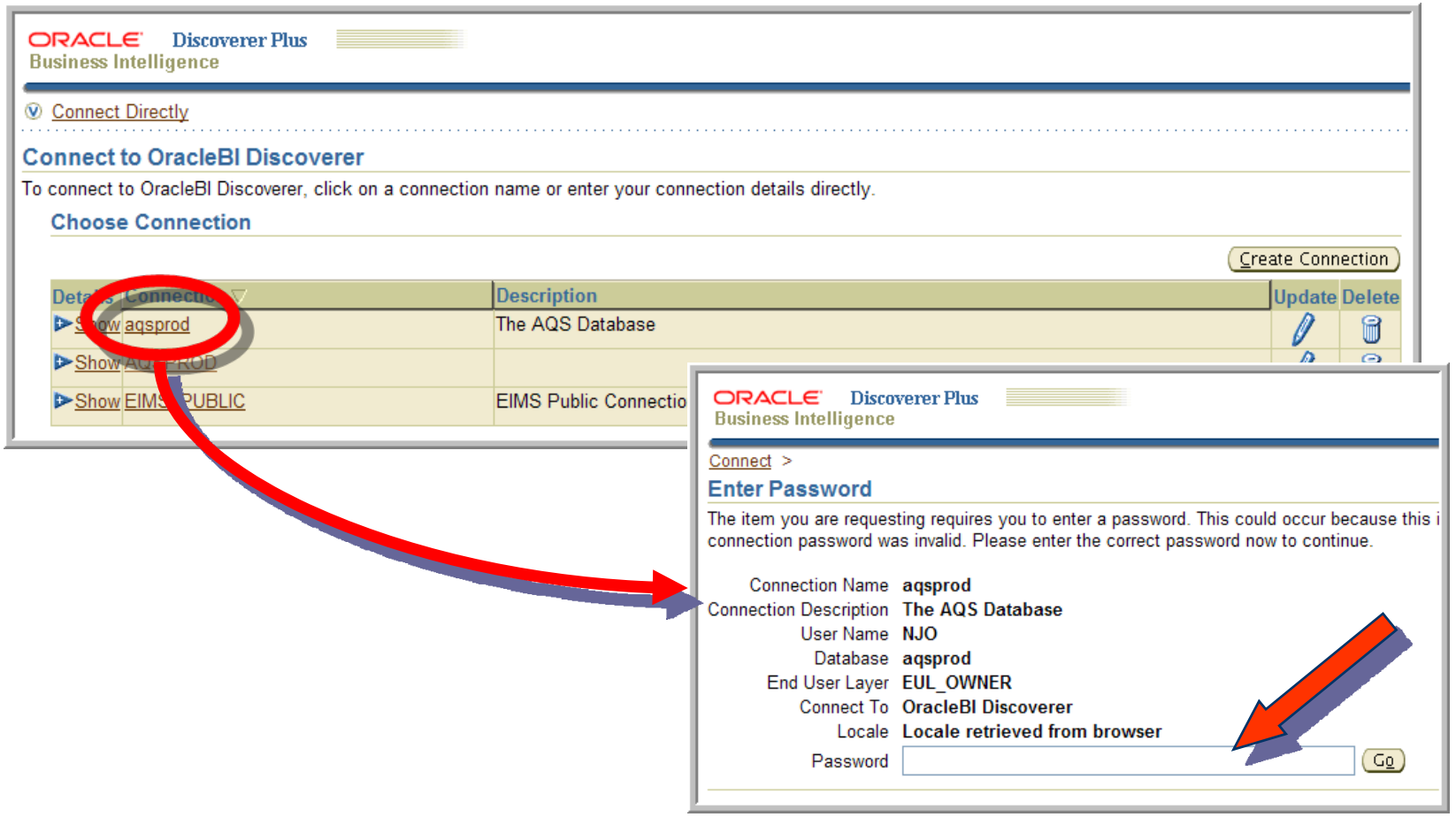
Making a Connection



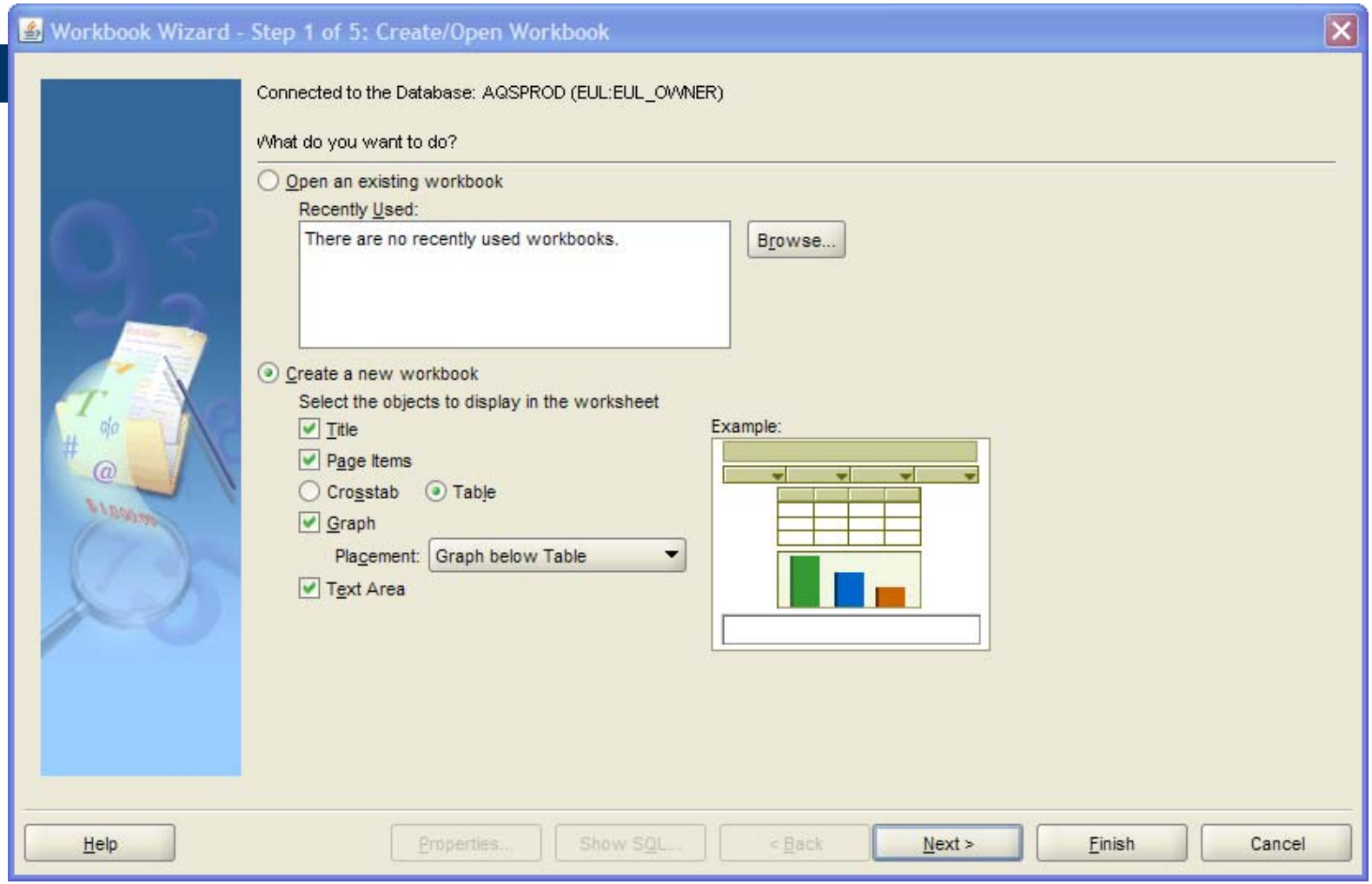
Setup: Fill in the Blanks and Apply



Click, Type, and Go

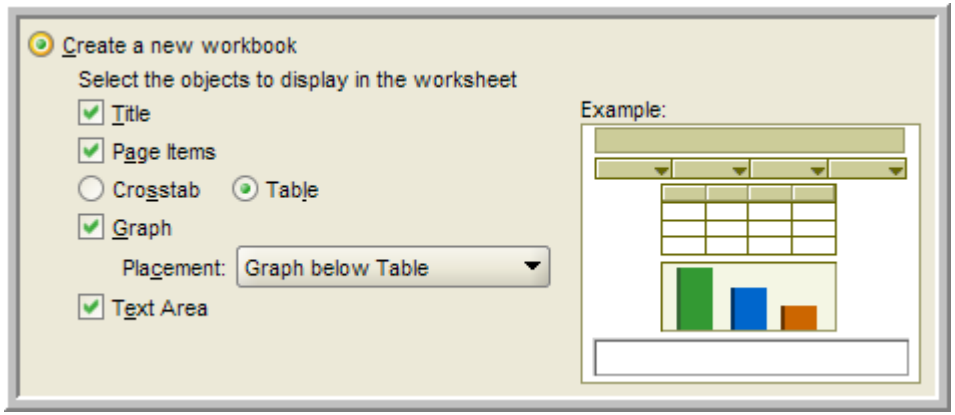


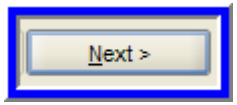
Successful Connection / Create Workbook



Creating a New Workbook

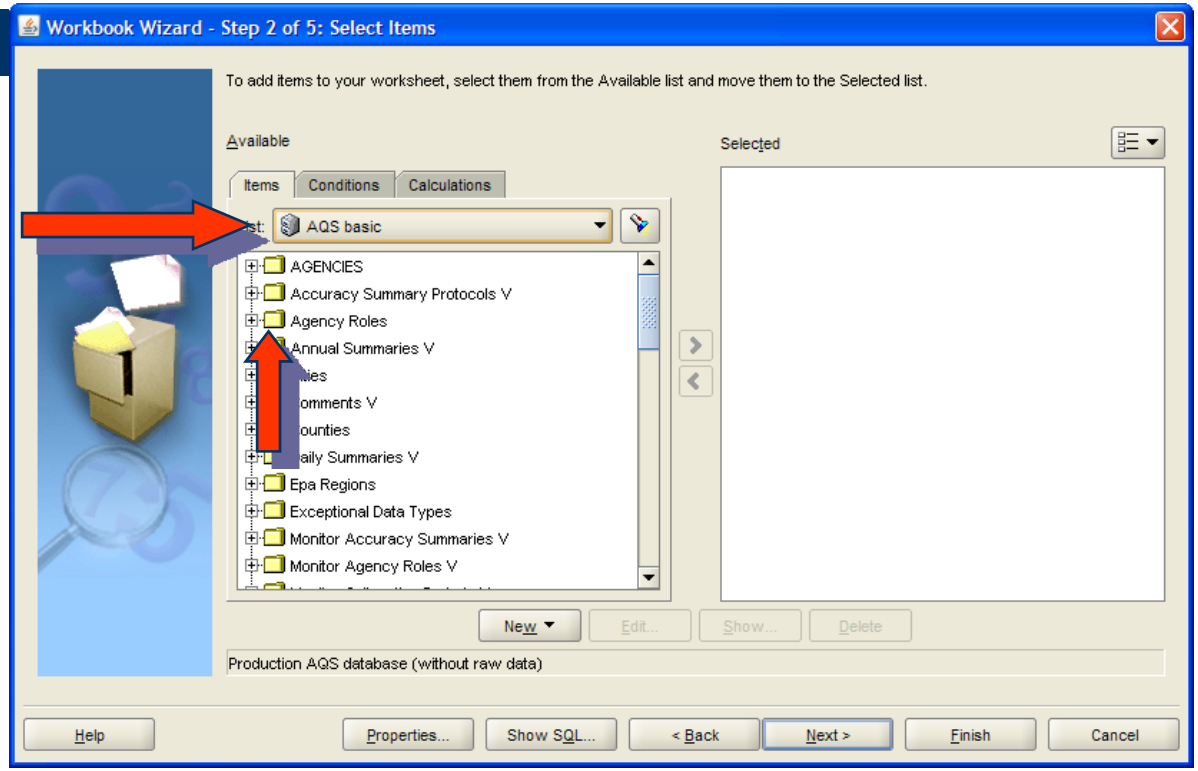
- Use the Workbook Wizard
- Start with a simple goal in mind – e.g., **Find the states in your region**





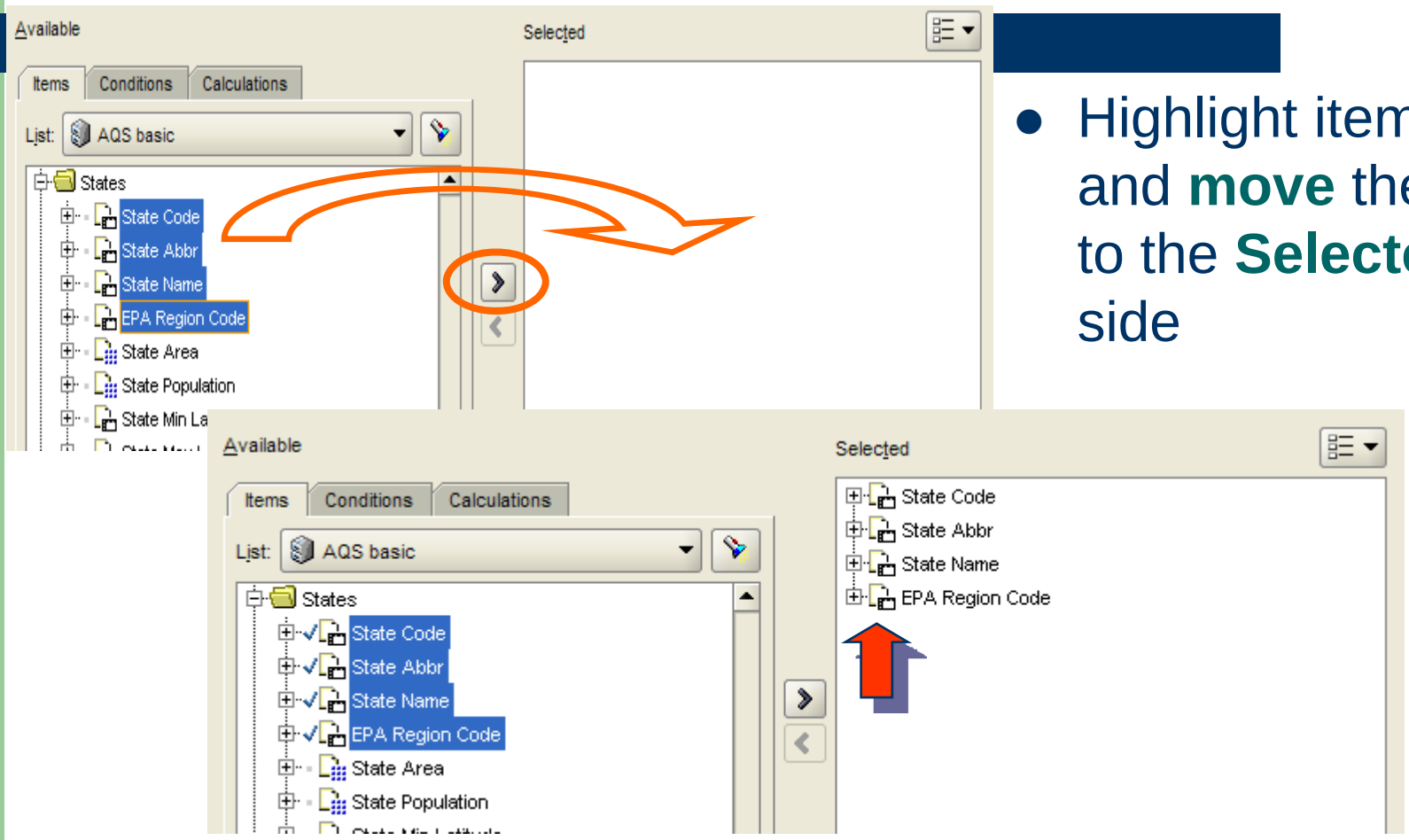
Selecting Items

- **AQS prod** is the only Business Area for you
- Scroll through list of folders
- Expand a folder to see its items



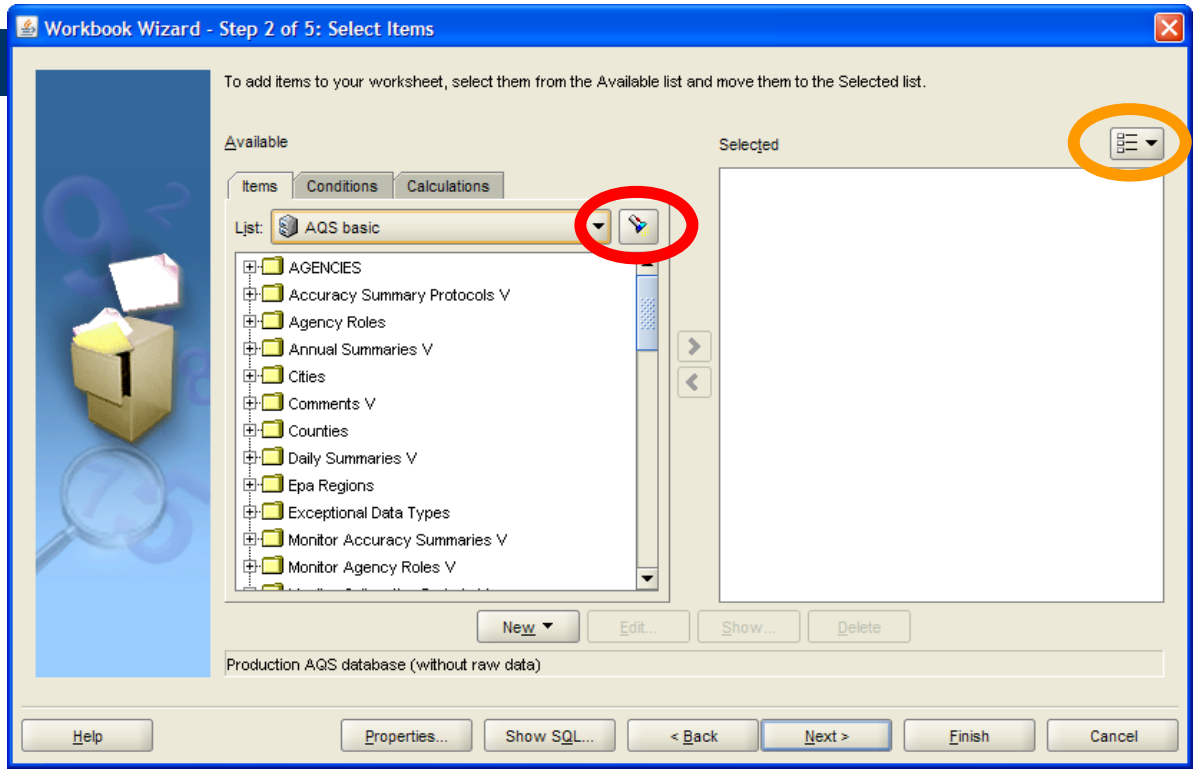
Terminology: …V = view

Selecting Items, cont'd



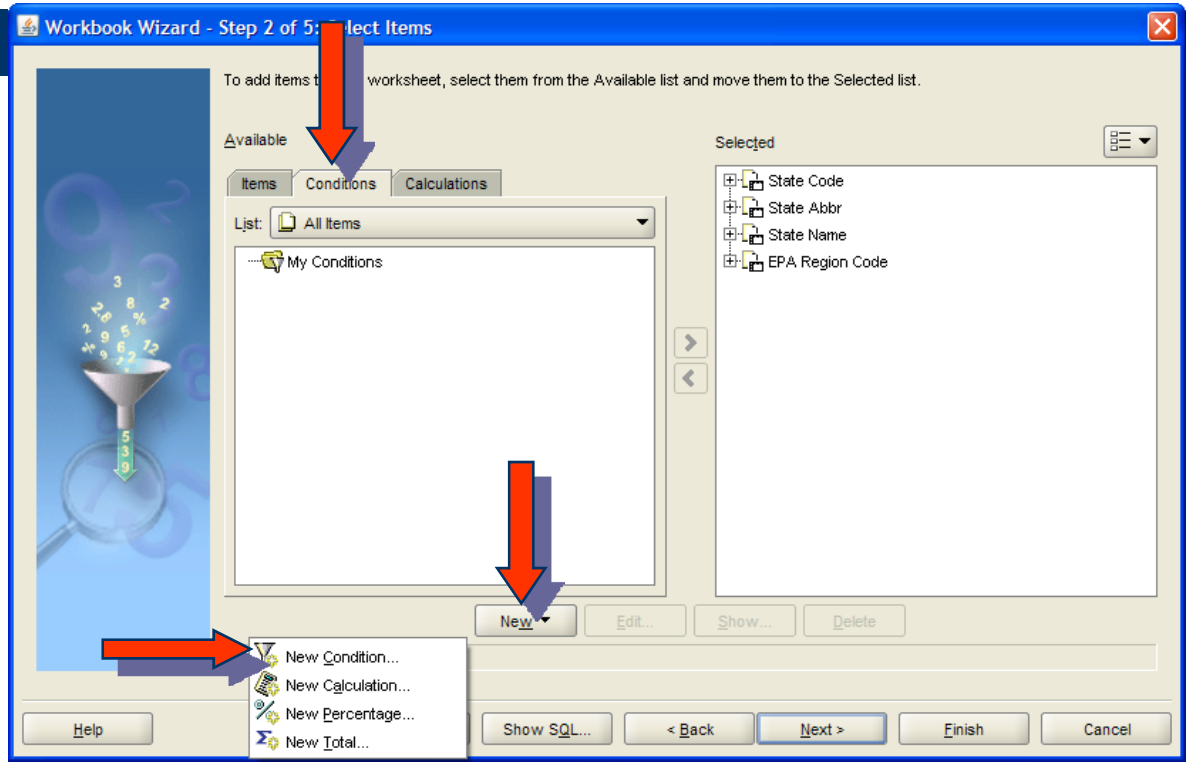
- Highlight items and **move** them to the **Selected** side

Searching for Items and Showing Pedigree



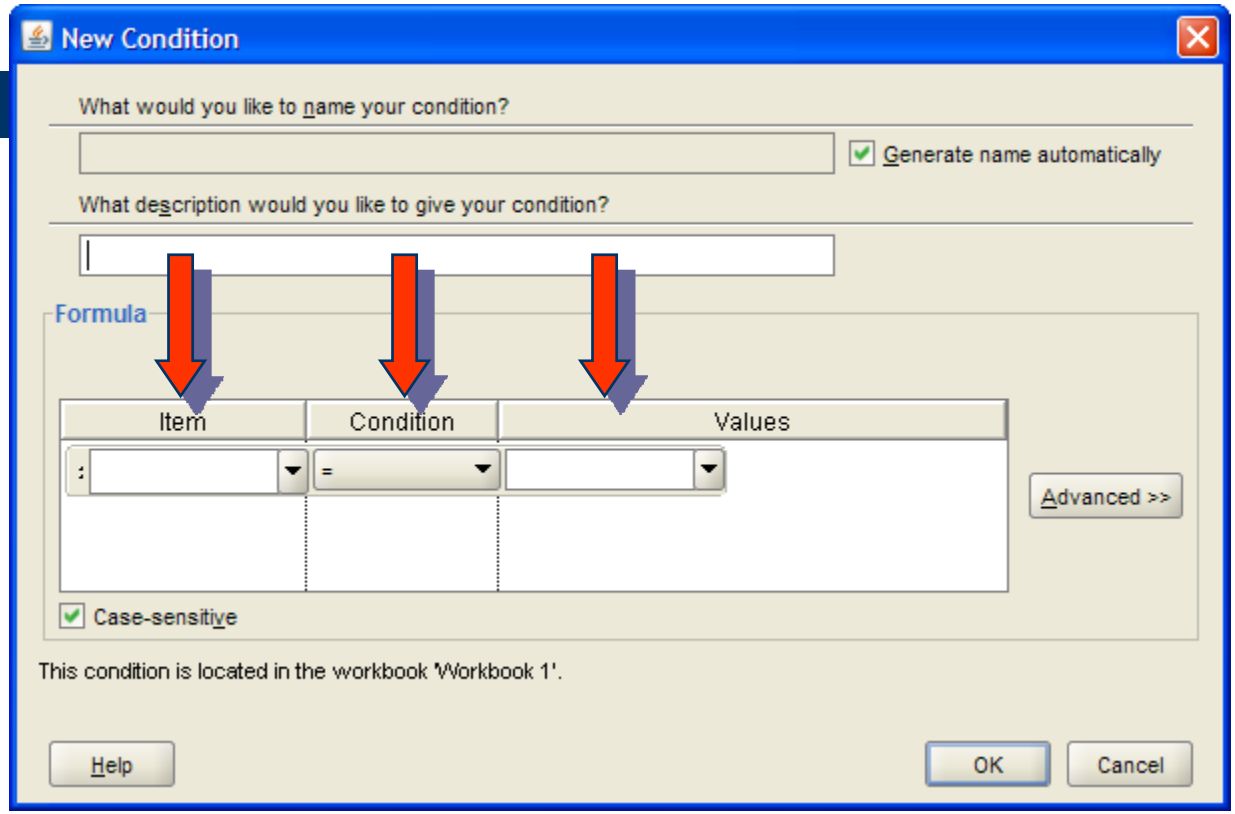
Selecting Conditions

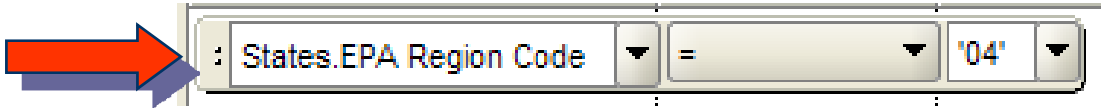
- Go to **Conditions** tab
- Select New
- Select New Condition



Selecting Conditions, cont'd

- Enter formula (other things optional)
  - Item
  - Condition
  - Value
    - Type in or pull down





Tablename . Fieldname

Will autoformat value (w/hint)

Possible Conditions

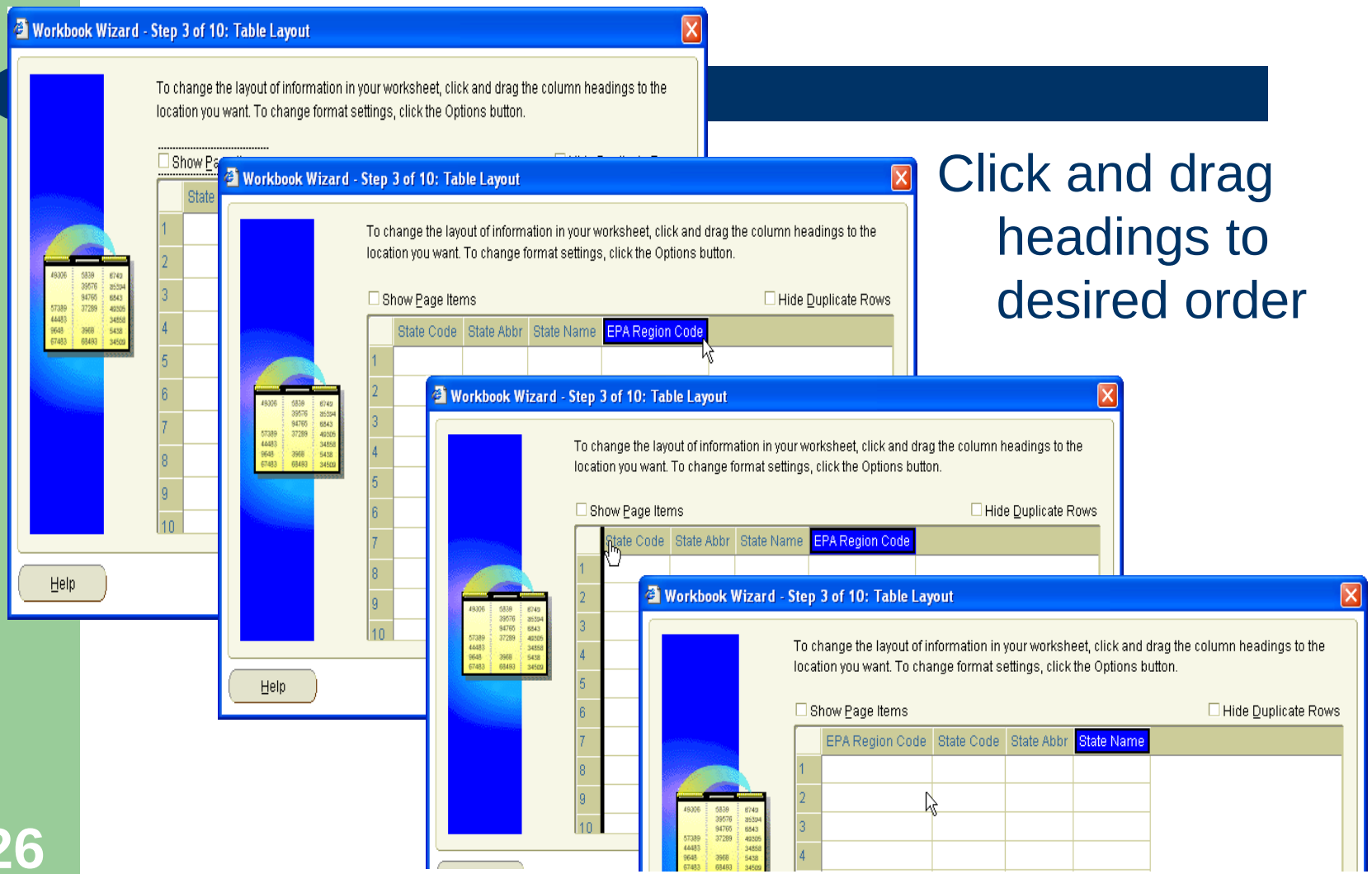
- We're spending a lot of time on conditions – they might be important!
- Possible Conditions

=	IS NULL
<> (not equal)	IS NOT NULL
>	NOT IN
<	BETWEEN
<=	NOT BETWEEN
>=	NOT LIKE
	!= (not equal)

LIKE LIKE '%OLORAD%' (% = *, _ = 1 char, CaSe SensiTivE)

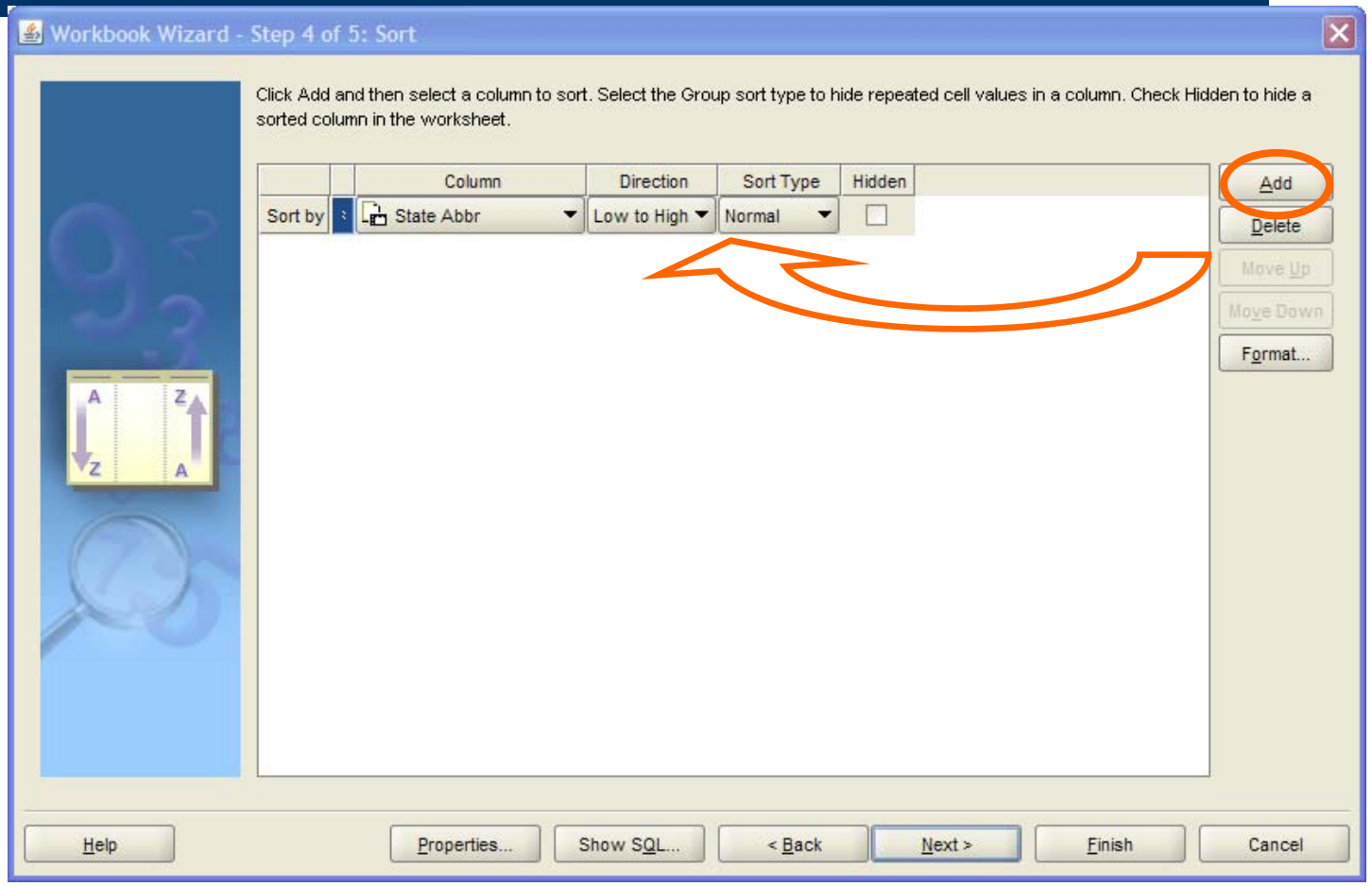
IN IN ('88101', '88501', '88502')

Table Layout

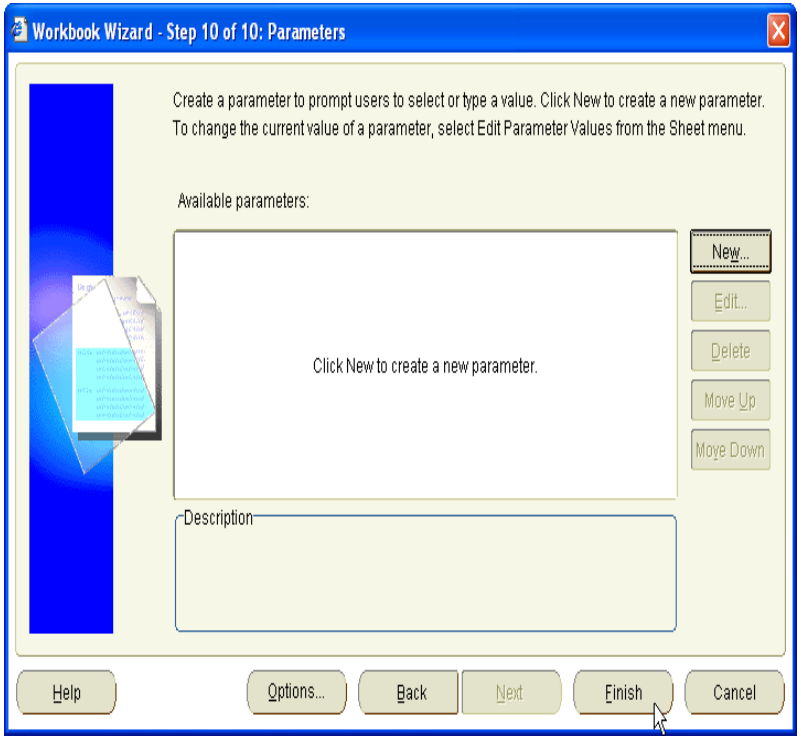


Click and drag headings to desired order

Sorts


Workbook Wizard - Step 4 of 5: Sort
Click Add and then select a column to sort. Select the Group sort type to hide repeated cell values in a column. Check Hidden to hide a sorted column in the worksheet.
Column
Direction
Sort Type
Hidden
Sort by
State Abbr
Low to High
Normal
Add
Delete
Move Up
Move Down
Format...
Help
Properties...
Show SQL...
< Back
Next >
Finish
Cancel

Parameters and Results

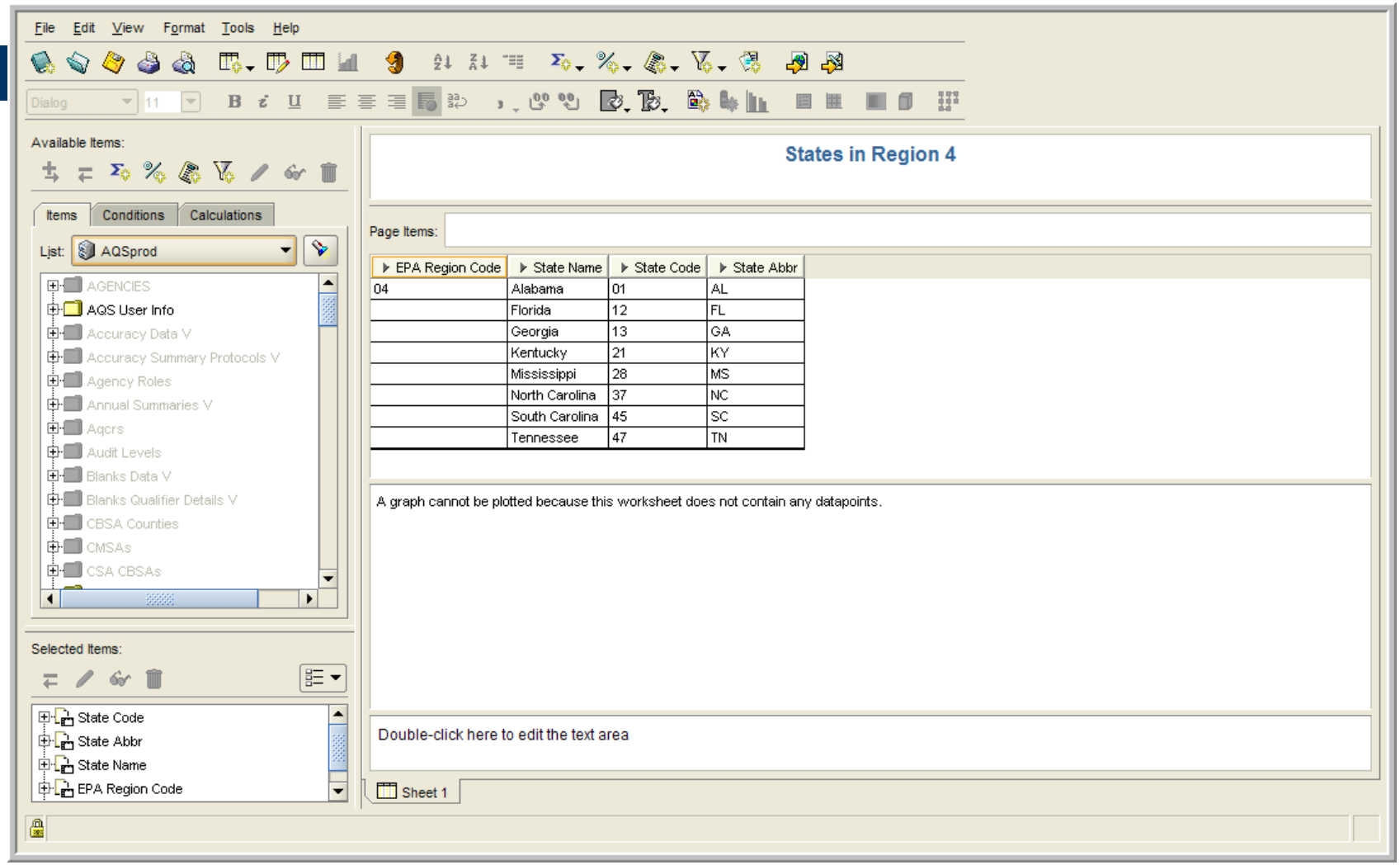


EPA Region Code	State Name	State Code	State Abbr
04	Alabama	01	AL
04	Florida	12	FL
04	Georgia	13	GA
04	Kentucky	21	KY
04	Mississippi	28	MS
04	North Carolina	37	NC
04	South Carolina	45	SC
04	Tennessee	47	TN

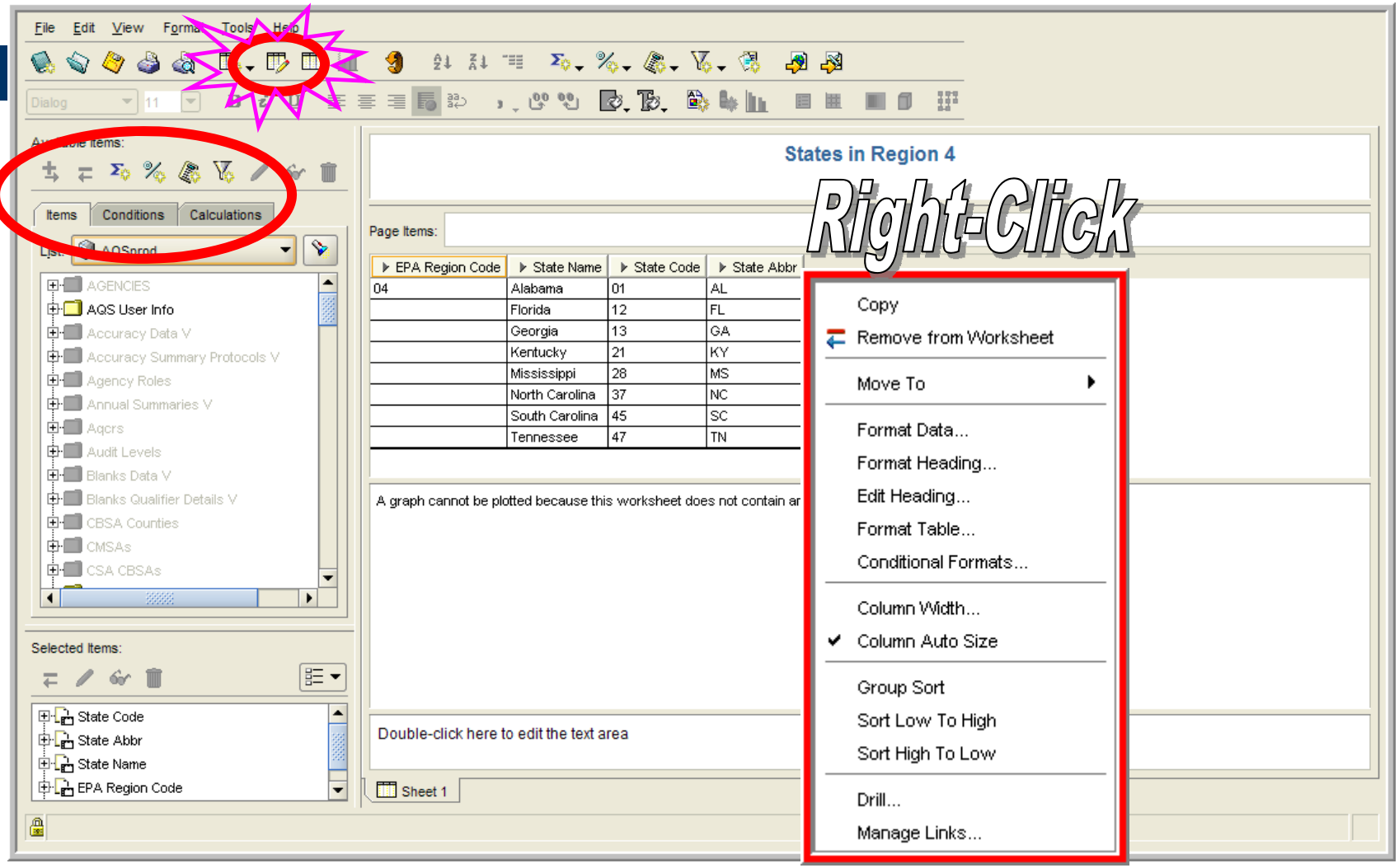
Exercise 1 - Introduction

1. Launch a web browser and go to Discoverer (http://www.epa.gov/ttn/airs/airsaqs)
2. Create your connection and Connect to aqsprod
3. Create a new workbook using the AQS basic business area to show all states within your EPA region
4. Adjust the table layout so that EPA Region is in column 1, State Abbreviation in column 2, State Name in column 3, State Code in column 4
5. Sort by EPA Region, then State Name. Change to a group sort on EPA Region.
6. Add a title
7. Extra Credit – Create another query to list the Pollutant Standards

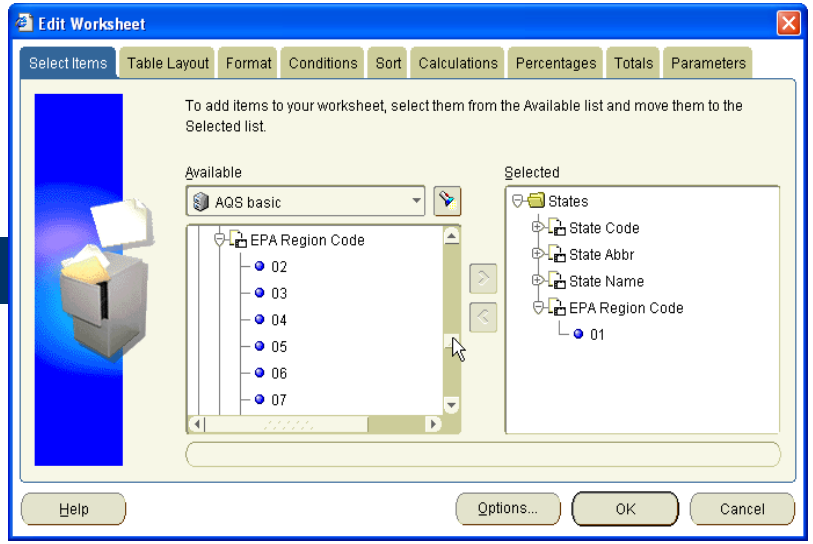
Exercise 1: Sample Results



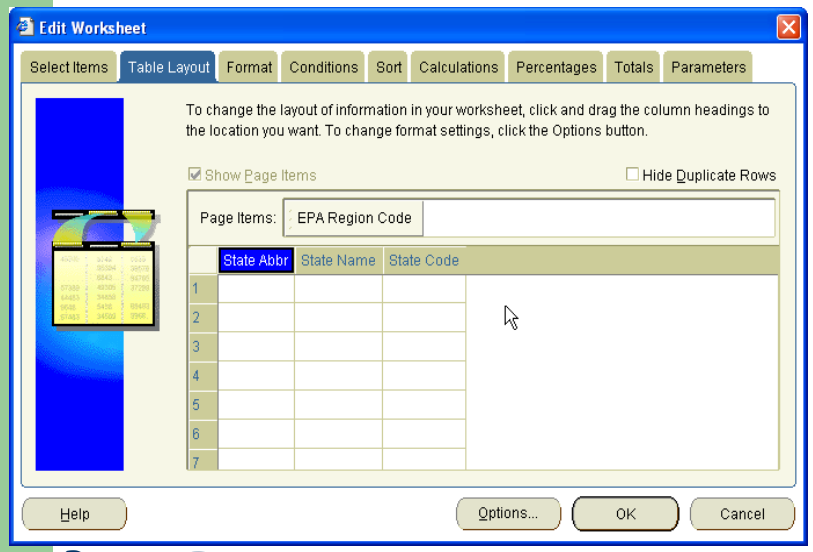
Important Screen Navigation Items



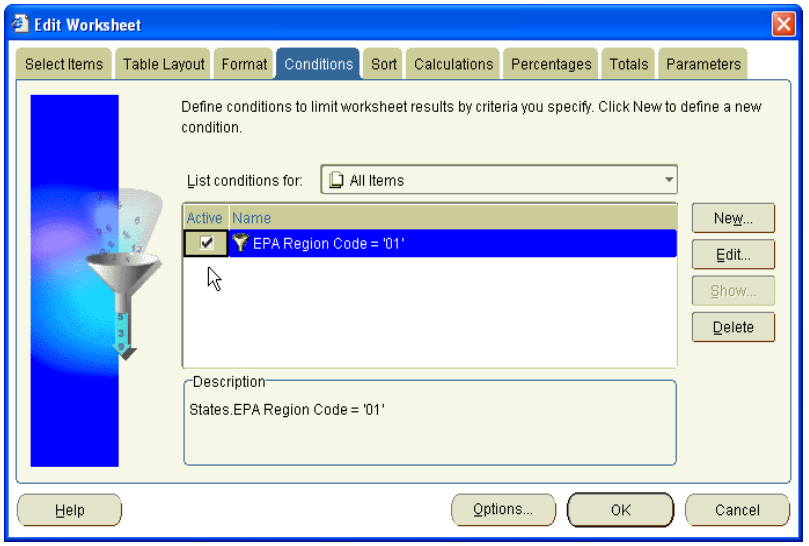
Did you notice?



Lists of Values (LOVs)

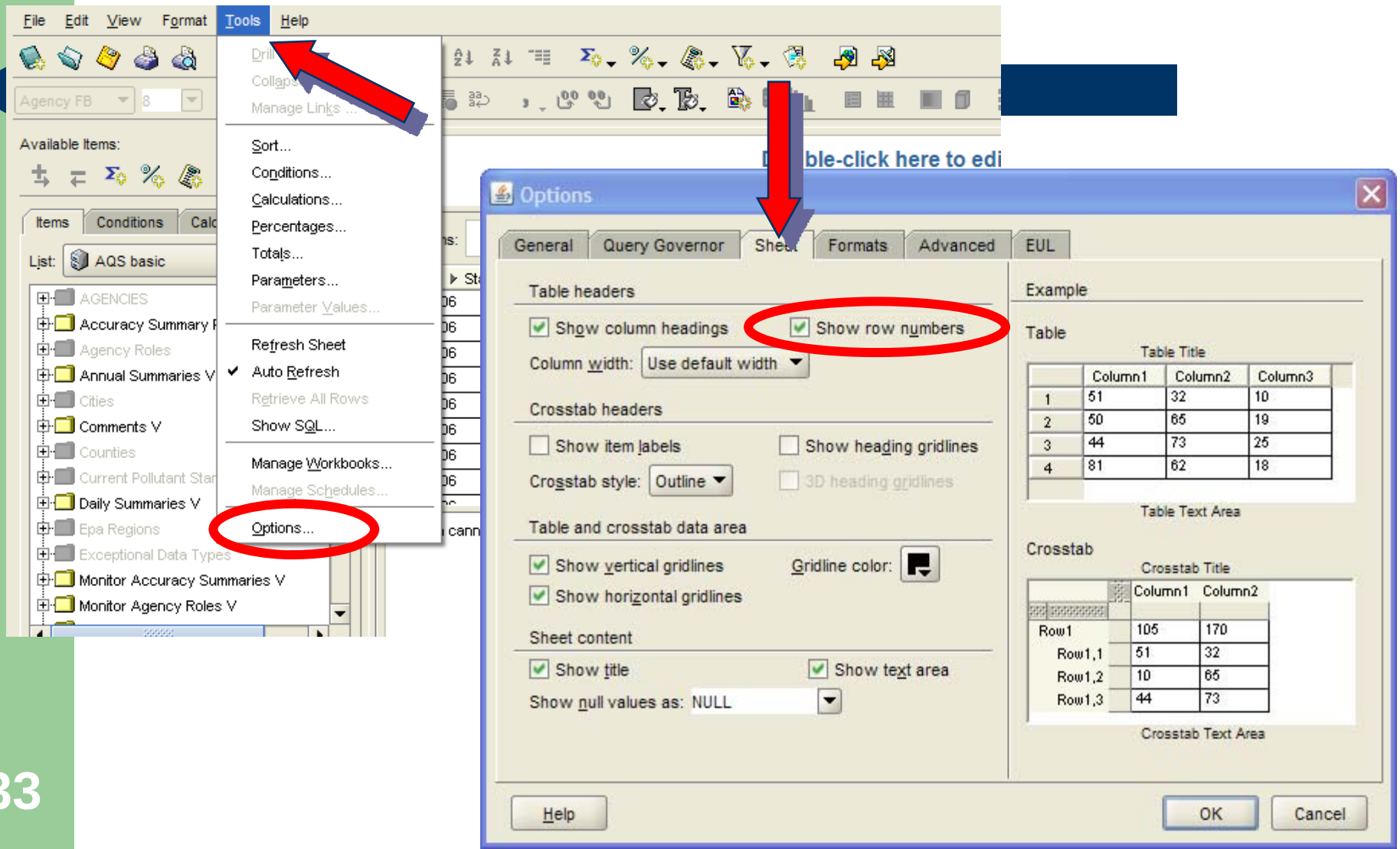


Page Items



Conditions may be inactive

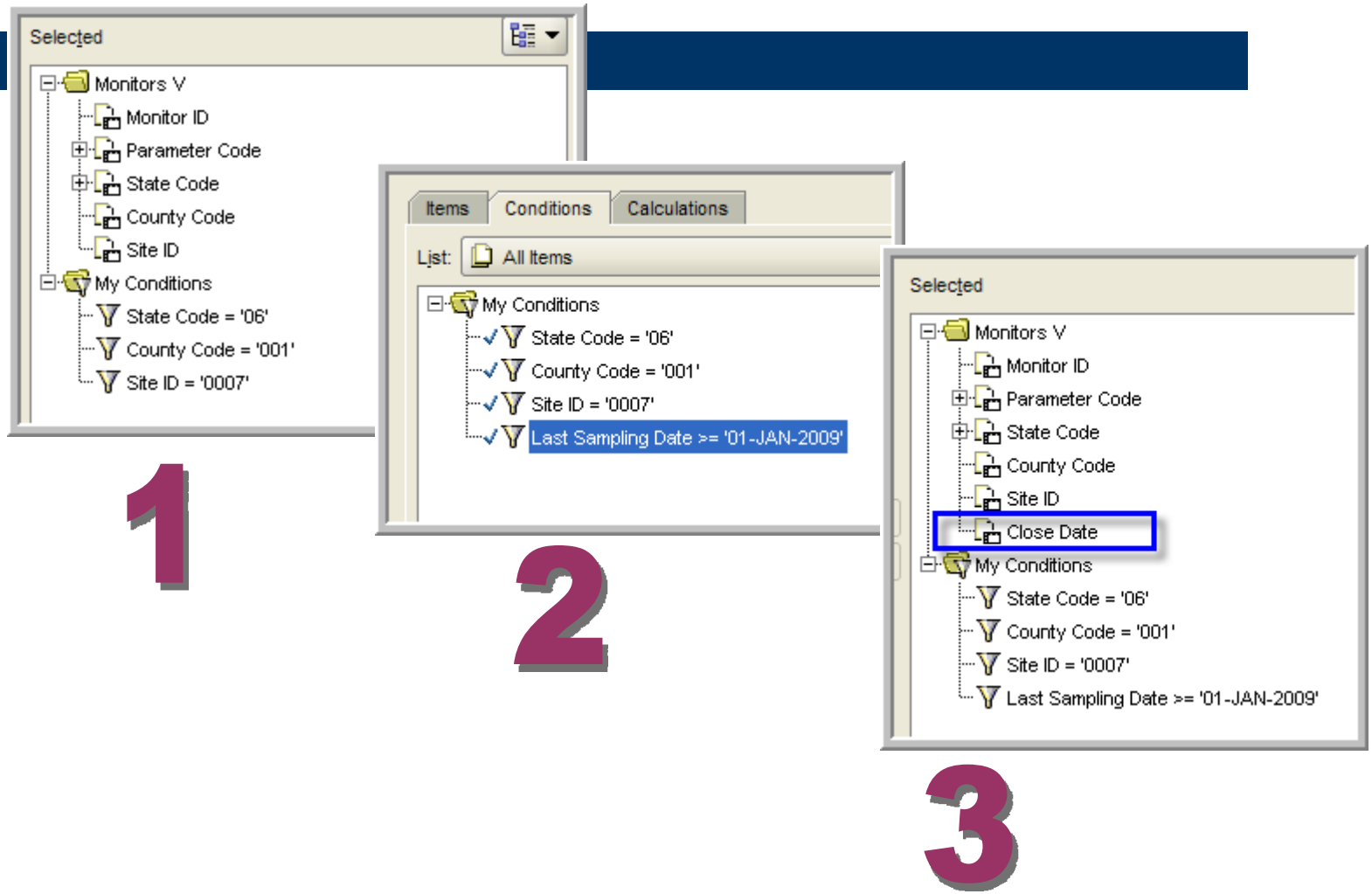
Changing Your Defaults



Exercise 2 – Edit Worksheet

1. Create a new worksheet that lists all the parameters (pollutants) measured at site 06-001-0007
2. Modify this sheet to include only parameters being sampled since 2009
3. Modify this sheet to show last sample date and close date (sampling end date)

Exercise 2 – How To



Exercise 2 Results

All parameters at site (116)

	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	42603
5	06	001	0007	43101
6	06	001	0007	43102
7	06	001	0007	43201
8	06	001	0007	43207
9	06	001	0007	43218
10	06	001	0007	43372
11	06	001	0007	43551
12	06	001	0007	43552
13	06	001	0007	43802
14	06	001	0007	43803
15	06	001	0007	43804
16	06	001	0007	43811
17	06	001	0007	43814
18	06	001	0007	43815
19	06	001	0007	43817
20	06	001	0007	43824

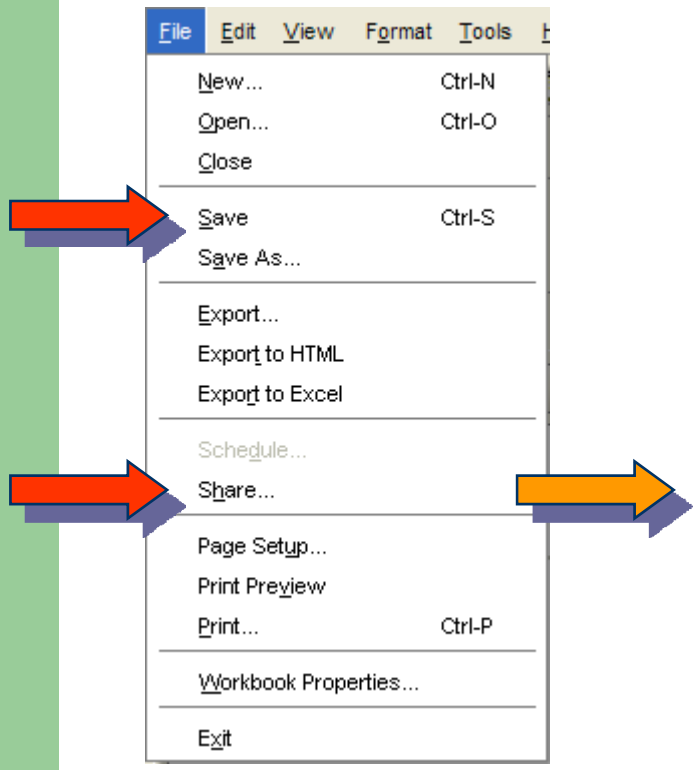
All parms since 2009 (37)

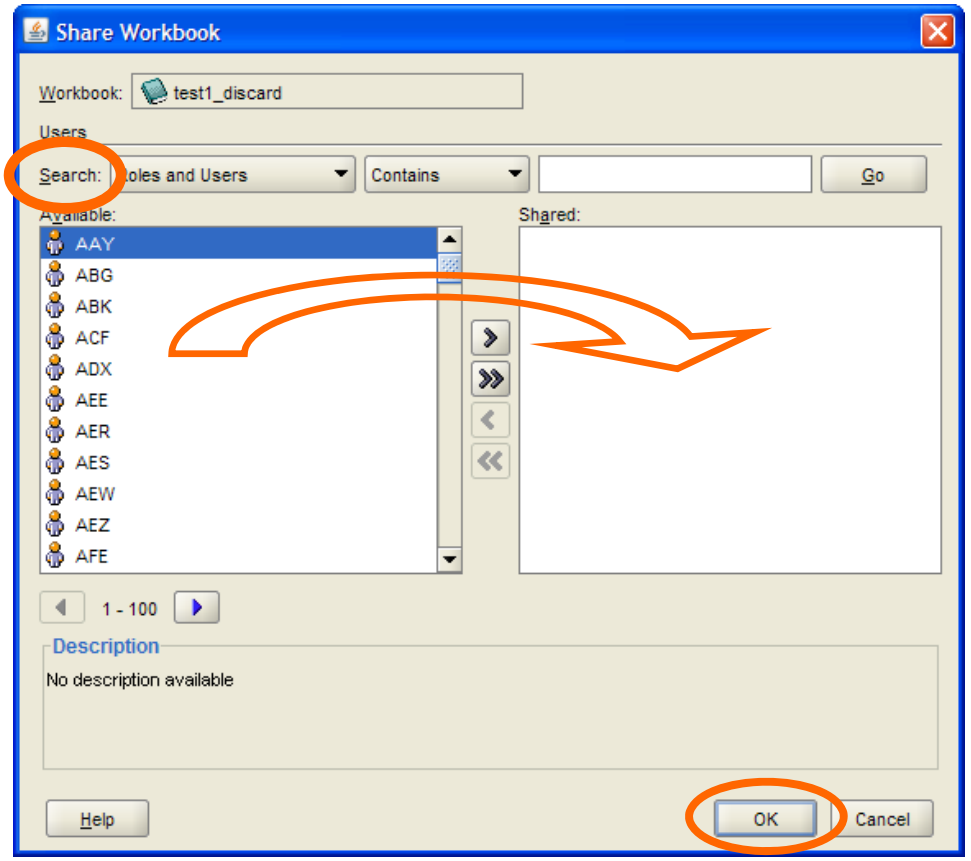
	State Code	County Code	Site ID	Parameter Code
1	06	001	0007	42101
2	06	001		
3	06	001		
4	06	001		
5	06	001		
6	06	001		
7	06	001		
8	06	001		
9	06	001		
10	06	001		
11	06	001		
12	06	001		
13	06	001		
14	06	001		
15	06	001		
16	06	001		
17	06	001		
18	06	001		
19	06	001		
20	06	001		

All parms w/end dates

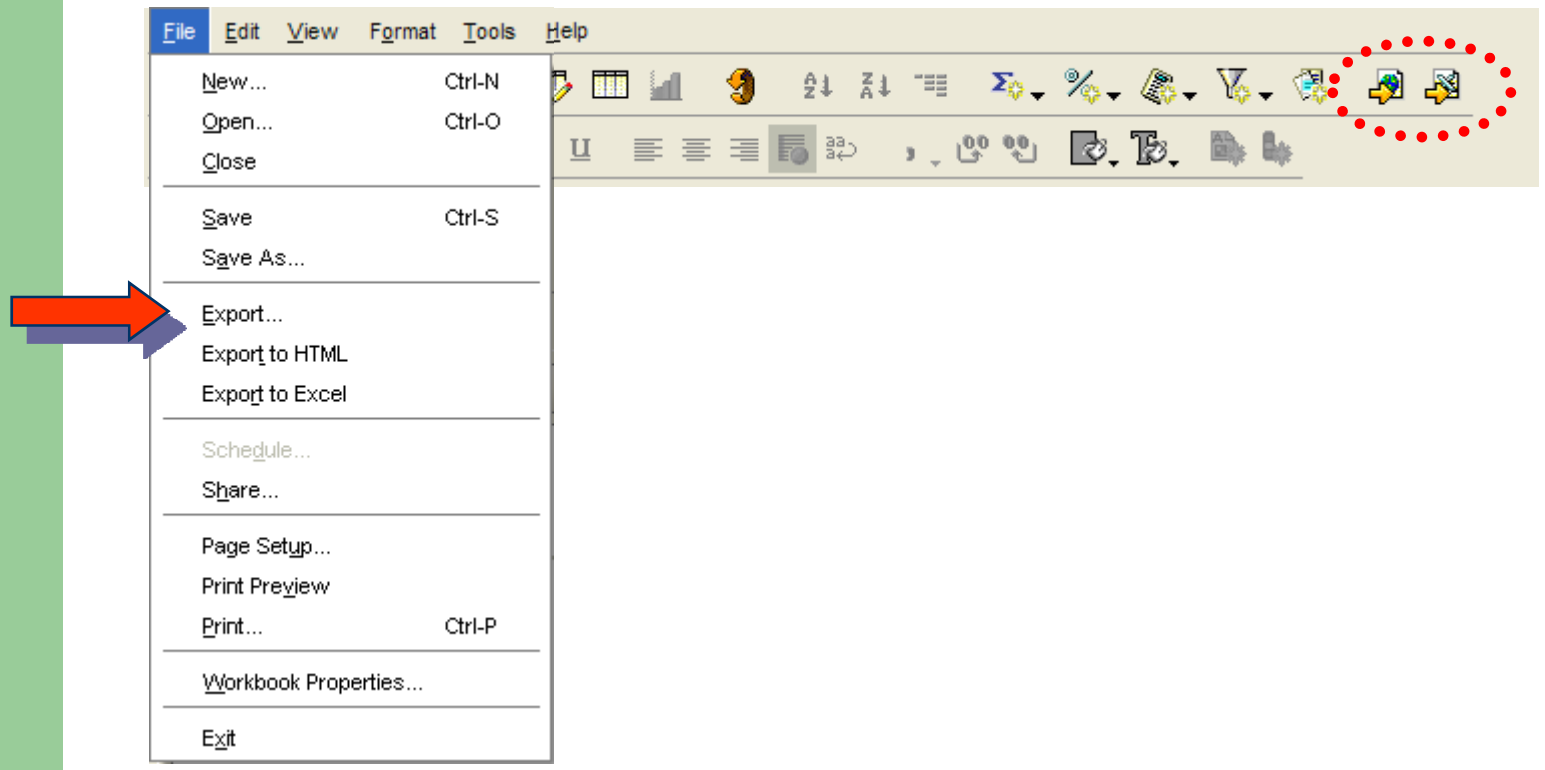
	State Code	County Code	Site ID	Parameter Code	Close Date	Last Sampling Date
29	06	001	0007	68101	NULL	31-DEC-2009
30	06	001	0007	68102	NULL	31-DEC-2009
31	06	001	0007	68103	NULL	31-DEC-2009
32	06	001	0007	68104	NULL	31-DEC-2009
33	06	001	0007	68105	NULL	31-DEC-2009
34	06	001	0007	68106	NULL	31-DEC-2009
35	06	001	0007	68107	NULL	31-DEC-2009
36	06	001	0007	68108	NULL	31-DEC-2009
37	06	001	0007	68109	NULL	31-DEC-2009
38	06	001	0007	81102	31-MAY-2000	31-MAY-2000
39	06	001	0007	81102	28-JUL-2008	29-JUN-2008
40	06	001	0007	82203	28-JUL-2008	29-JUN-2008
41	06	001	0007	82301	28-JUL-2008	29-JUN-2008
42	06	001	0007	82306	28-JUL-2008	29-JUN-2008
43	06	001	0007	82403	28-JUL-2008	29-JUN-2008
44	06	001	0007	85101	28-JUL-2008	29-JUN-2008
45	06	001	0007	88101	NULL	31-DEC-2009
46	06	001	0007	88102	NULL	27-SEP-2008
47	06	001	0007	88103	NULL	27-SEP-2008

Saving and Sharing

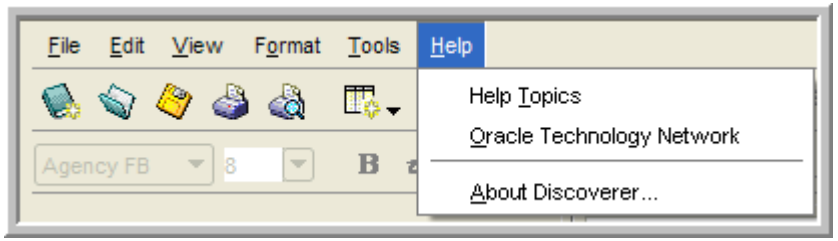




Exporting


File
Edit
View
Format
Tools
Help
New... Ctrl-N
Open... Ctrl-O
Close
Save Ctrl-S
Save As...
Export...
Export to HTML
Export to Excel
Schedule...
Share...
Page Setup...
Print Preview
Print... Ctrl-P
Workbook Properties...
Exit

Getting Help

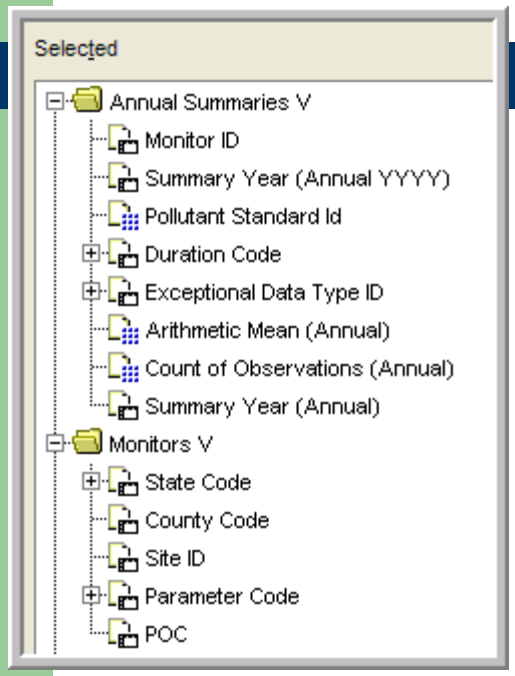


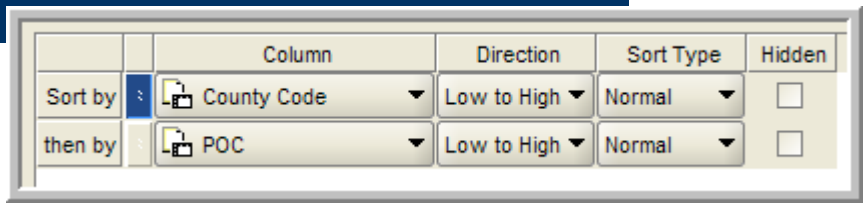
- On-line help
  - Note: the OTN info is for v11, we're still at v10
- Call the EPA Help Desk
  - 866-411-4372
  - "AQS Discoverer"

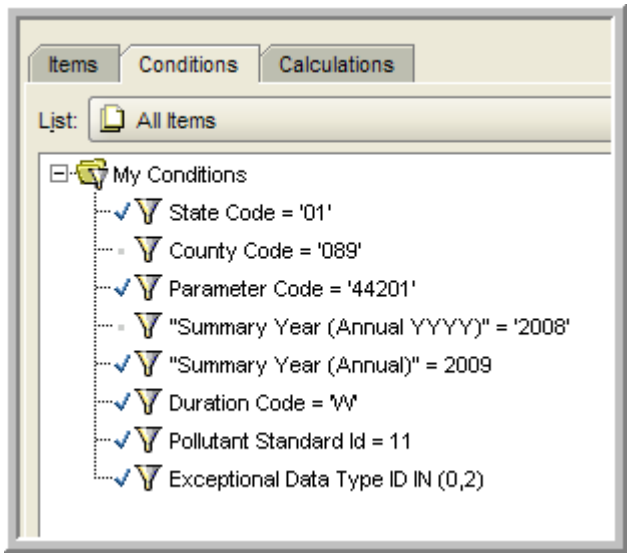
Getting Data from Multiple Tables

- Once you select an item, folders that don't connect to it are grayed out (EUL in action!)
- You have to build a bridge to the items you need by selecting intermediate folders
- A "fan trap" error means you've connected items in an ambiguous way that Discoverer cannot interpret – use a different connection path

Annual Summary for 1 State, 1 Yr, 1 Parm







Notice what happens when an annual summary item is selected

Should Look Something Like This…

	Monitor ID	Summa	State Code	County Code	Site ID	Parameter Code	POC	Dur
1	01-003-0010-44201-1	2009	01	003	0010	44201	1	W
2	01-033-1002-44201-1	2009	01	033	1002	44201	1	W
3	01-051-0001-44201-1	2009	01	051	0001	44201	1	W
4	01-055-0011-44201-1	2009	01	055	0011	44201	1	W
5	01-069-0004-44201-1	2009	01	069	0004	44201	1	W
6	01-073-1010-44201-1	2009	01	073	1010	44201	1	W
7	01-073-2006-44201-1	2009	01	073	2006	44201	1	W
8	01-073-6002-44201-1	2009	01	073	6002	44201	1	W
9	01-073-1003-44201-1	2009	01	073	1003	44201	1	W
10	01-073-1005-44201-1	2009	01	073	1005	44201	1	W
11	01-073-1009-44201-1	2009	01	073	1009	44201	1	W
12	01-073-5002-44201-1	2009	01	073	5002	44201	1	W

Build a Bridge to Another Table

- Now, add the county population
  - Via states?
    - Monitor has state, states has counties…
    - Did it work?
  - Via ???
- Fan traps are caused by how the data tables (and EUL) are organized, not how physical reality is organized

Calculations and Totals

- Add some math
  - Observations per person
  - Mean of annual means
- Calculations operate on columns
  - Generally an operator: + - * / ||
  - Example: (0.5 * ALT_MDL) - SAMPLE_VALUE
  - PARAMETER_CODE || ' (' || PARAMETER_DESC || ')'
  - Adds a new column
- Totals operate on rows
  - Generally a function: min, max, sum, avg, stddev, count, etc.
  - Example: MIN(OBS_PERCENT)
  - Adds a new row
- Can also be done by exporting

Calculation and Total Example

Arithmetic Mean (Annual)	County Population	Obs per capita
.03671	662047	0.01
.03895	662047	0.01
.03951	223510	0.03
.03921	49756	0.12
.03353	14798	0.39
.03706	88787	0.06
.04371	276700	0.03
.03952	399843	0.01
.03913	399843	0.01
.04102	111064	0.05
.03998	143293	0.04
.03219	164875	0.03
.03769	140415	0.04
Average: .03869		

Exercise 3 – Data from Multiple Tables

- Repeat the lecture example (except for county population items) selecting data from more than one table – explore the data related to the annual summary data
- Items
  - Monitors Table: state, county, site, parameter, POC
  - Annual Summary Table: monitor, year, PS, duration, EDT, mean, obs
- Conditions
  - State = 01
  - Parameter = 44201
  - Year = 2008
  - DEP = W, (0,2), 11
- Sort
  - County
  - POC

Extra Credit: Add some totals or calculations

Crosstabs (Pivot Tables)

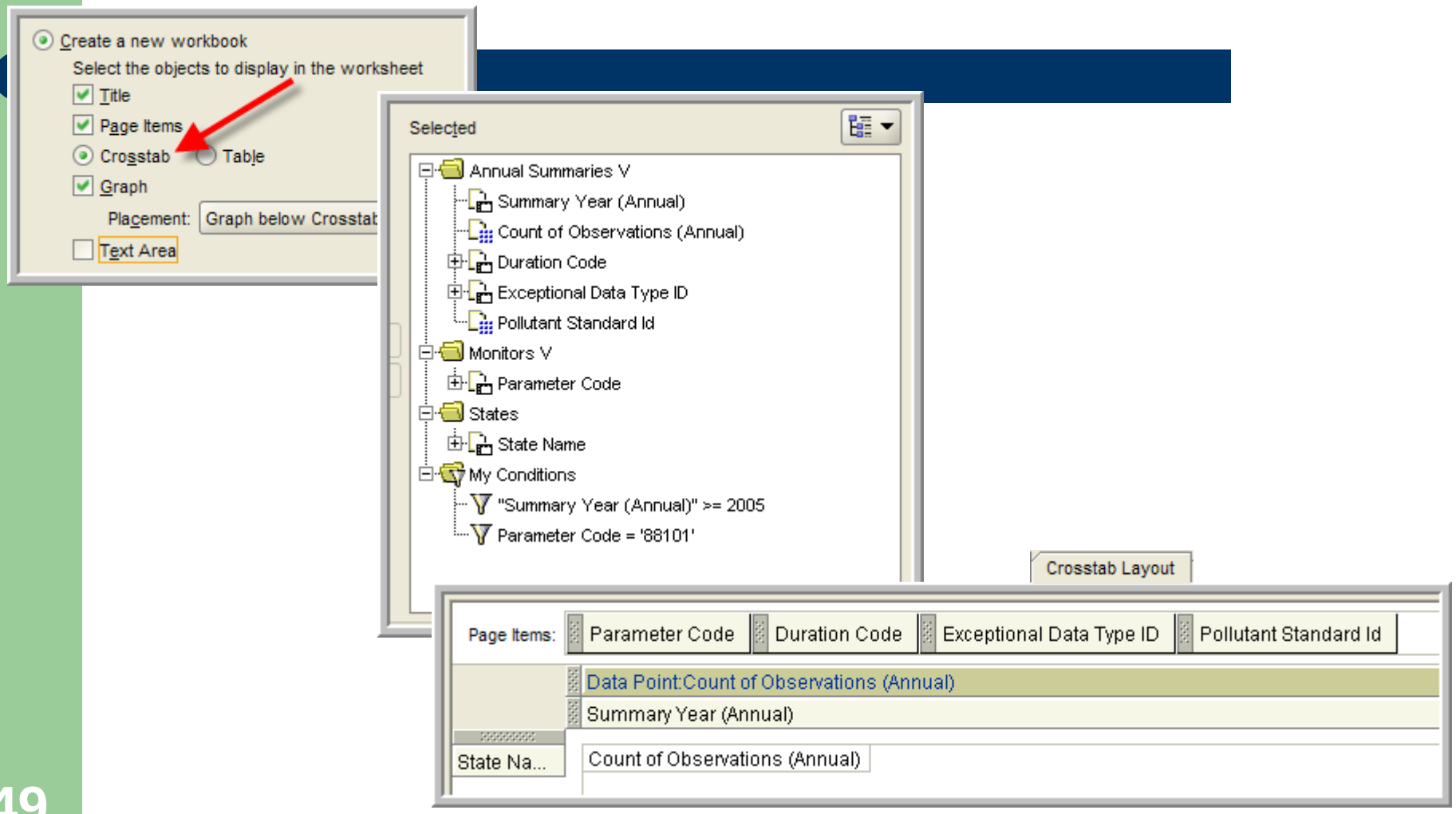
- Normal tables have selected items across the top
- Crosstabs have items across the top and items down the side
  - Values are shown in the grid
  - Some kind of aggregate (sum, count, etc.) if necessary
- Page Items are also useful for filtering data

Crosstab Example

Page Items: Parameter Code: 88101 | Duration Code: 7 | Exceptional Data Type ID: 0 | Pollutant Standard Id: 16

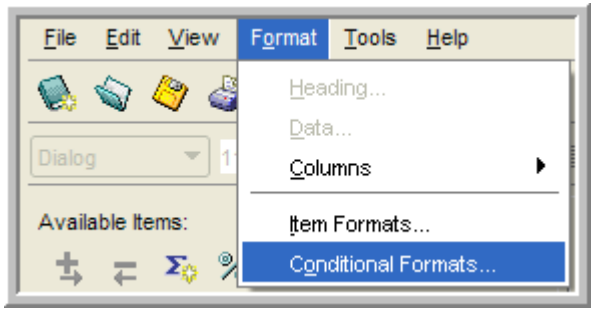
	Count of Observations (Annual)					
	2005	2006	2007	2008	2009	2010
Alabama	4094	4024	436	4500	4110	604
Alaska	369	407	559	578	365	NULL
Arizona	1629	1740	1516	1405	1089	NULL
Arkansas	2320	2418	1364	2532	2539	630
California	11084	9434	6710	6883	9112	193
Canada	99	51	NULL	NULL	NULL	NULL
Colorado	2525	2577	2220	2347	546	712
Connecticut	1867	1972	2520	2669	2670	NULL
Delaware	1108	1055	1067	1033	1054	178
District Of Columbia	863	529	846	875	888	218
Florida	5114	4846	745	2142	6793	1340
Georgia	3427	3823	675	4026	2474	311
Hawaii	164	NULL	467	243	105	NULL
Idaho	1385	958	955	1034	1114	30
Illinois	3403	3471	3464	3462	4086	983
Indiana	4848	5122	NULL	5831	5233	1357
Iowa	2671	2542	3610	3768	4418	1081
Kansas	1588	1436	1255	1259	927	NULL
Kentucky	NULL	1597	16	3189	2985	412
Louisiana	2950	2735	2523	2526	2322	653
Maine	1056	976	919	1149	1129	109
Maryland	2679	2381	2457	1886	2571	NULL

Building the Crosstab Workbook

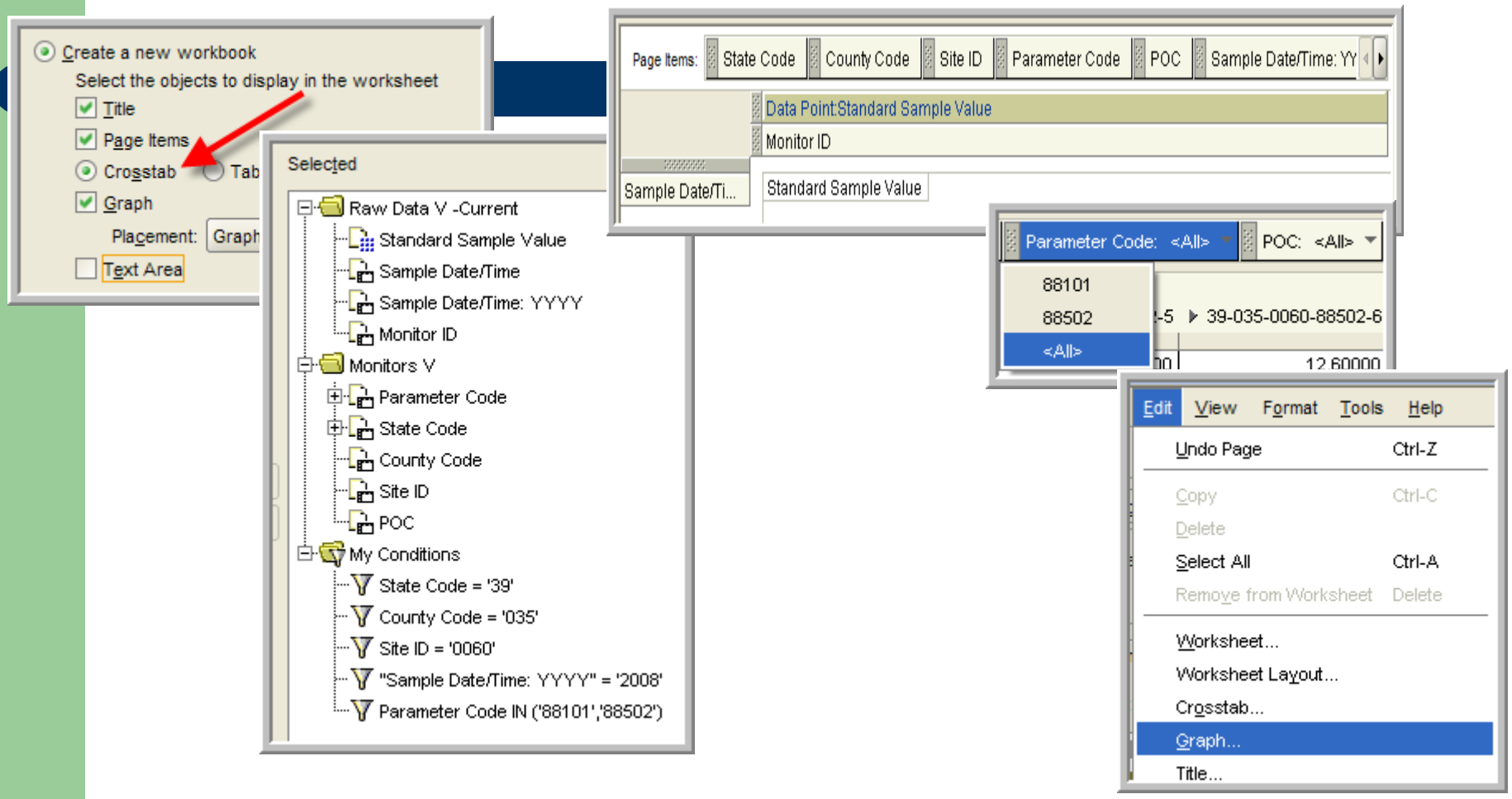


Exercise 4 – Crosstab with Graph

- Open a new worksheet and select a crosstab query
- Query Raw Data - Current data for :
  - site 39-035-0060
  - Year = 2008
  - Parameters in 88101, 88502
- Move
  - Date to right side
  - Monitor ID to top
  - Standard Sample value into grid
  - All else into Page Items
- Finish
- From Page Items: All params and POCs
- Edit Graph format to line
- Optional: conditionally format all samples over 35 with red background



Exercise 4 – How To

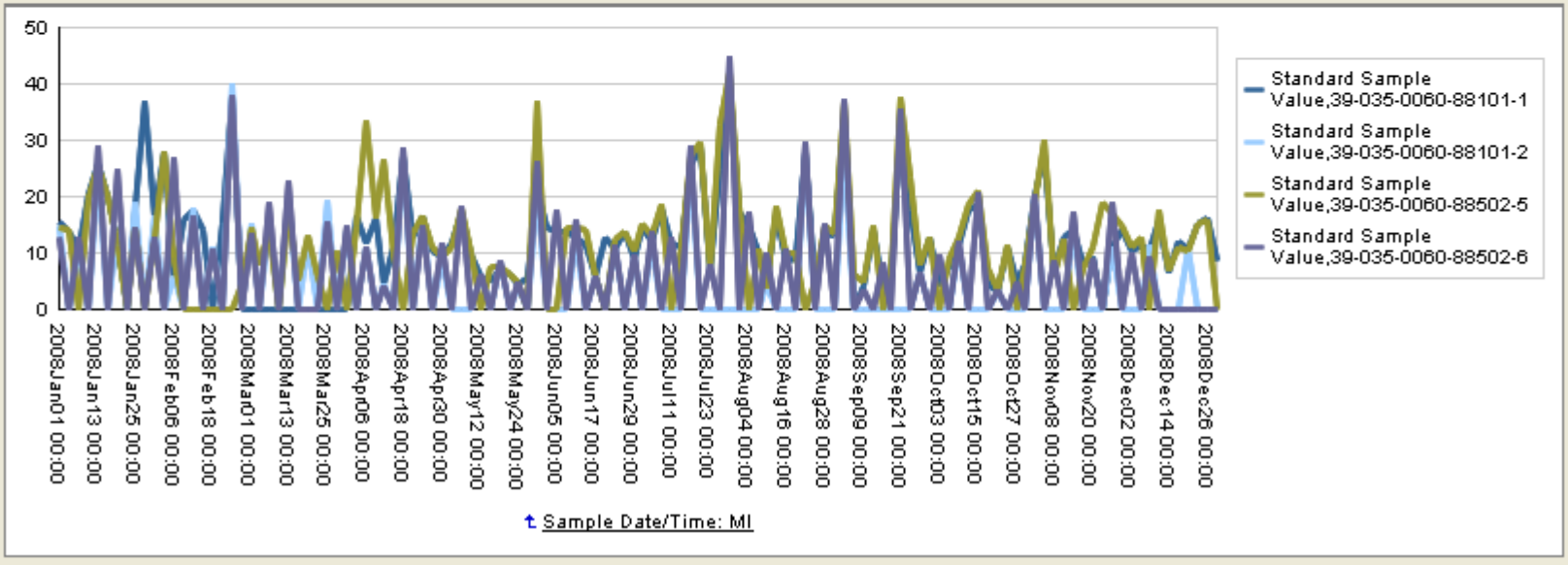


Exercise 4 – Crosstab with Graph

Collocated PM2.5 Monitors

Page Items: State Code: 39 | County Code: 035 | Site ID: 0060 | Parameter Code: <All> | POC: <All> | Sample Date/Time: YYYY: 20

	Standard Sample Value			
	39-035-0060-88101-1	39-035-0060-88101-2	39-035-0060-88502-5	39-035-0060-88502-6
2008Jan01 00:00	15.60000	14.80000	14.60000	12.60000
2008Jan04 00:00	13.90000	NULL	14.30000	NULL
2008Jan07 00:00	10.80000	9.70000	NULL	12.70000
2008Jan10 00:00	20.40000	NULL	19.80000	NULL
2008Jan13 00:00	25.70000	24.60000	25.50000	28.90000
2008Jan16 00:00	19.90000	NULL	20.00000	NULL



Hints, Tips and Other Good Stuff

- Discoverer is accessing the live database (same as the AQS application)
- Your Discoverer ID and password are the same as your AQS ID and password
- Only registered AQS users can access the AQS database via Discoverer
- If you delete your cookies, you'll have to recreate your connection
- Your Connection must be created on each different PC you use
- There is one EUL for AQS; and 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
- 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

More Hints, Tips, and Good Stuff

- Selection has search function (flashlight icon)
- Max time for Discoverer queries = 60 mins
- Max number of rows= 65,000
- 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
- Filter builder has auto-format
- Online Help is pretty good
- 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)
- You can change the directory for exported worksheets:
  - File | Export… (use wizard instead of shortcut)
- 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)

There’s lots to do!

Discover the data

with ***DISCOVERER***!

You are now a ***DISCOVERERER***?

Appendix: Syllabus for Class

- AQS Database basics
- Starting Discoverer
- Basic query
- Grouping data
- Queries with multiple tables
- Editing an existing query
- Calculations / Totals
- Saving queries
- Sharing queries
- Exporting data
- Getting Help
- Refreshing data
- Crosstab queries
- Graphs

Appendix: Key Terminology

Interface to database = End User Layer (EUL)
Grouping of Tables = Business Area
Group of worksheets = Workbook

Tables	⇨	**Folders**
Columns	⇨	**Items**
Rows	⇨	**Records**
Queries	⇨	**Worksheets**
Filters	⇨	**Conditions, Criteria**