

# Instructions for accessing and analyzing SDWIS/FED drinking water data in Excel PivotTables

<b>1</b>	<b>What's a PivotTable?</b> .....	<b>1</b>
	A few terms.....	1
<b>2</b>	<b>What's in the Annual Trends PivotTables?</b> .....	<b>3</b>
2.1	Inventory.....	3
2.2	Violation type .....	3
2.3	Rule.....	4
2.4	Contaminant.....	4
2.5	GPRA.....	4
<b>3</b>	<b>How to use PivotTables</b> .....	<b>6</b>
3.1	Just the basics.....	6
3.1.1	Adding or removing dimensions and measures.....	6
3.1.2	Filtering data members .....	7
3.1.3	Rearranging the dimensions and measures to facilitate analysis .....	8
3.1.4	Drilling down.....	9
3.1.5	Saving PivotTables .....	9
3.2	Organizing data, printing, graphing .....	10
3.2.1	Adding and removing dimension subtotals .....	10
3.2.2	Adding or removing row or column totals.....	10
3.2.3	Changing labels .....	11
3.2.4	Sorting data in ascending or descending order .....	11
3.2.5	Organizing data using the Page area.....	12
3.2.6	Printing .....	12
	Graphing.....	13
3.3	Some additional functions not covered here .....	14
<b>4</b>	<b>So, what do we make of all these numbers?</b> .....	<b>15</b>
4.1	General tips and mistakes to avoid .....	15
4.1.1	Don't forget to select the 'Sum of...' measures each time you insert them into a PivotTable .....	15
4.1.2	Don't refresh your data unless you have the original data tables on your computer	15
4.1.3	Don't put measures in a Column, Row or Page area and don't put dimensions in the Data area .....	15
4.1.4	Don't forget about filtered or "hidden" data members .....	15
4.1.5	Don't add the number of systems in violation or population affected across violation types or contaminant types.....	15
4.1.6	Be careful when interpreting population served data, especially when looking for trends	16
4.1.7	Organizing 'State' and 'PSource' data .....	16
4.2	Tips by PivotTable.....	16
4.2.1	Inventory PivotTable .....	16
4.2.2	Violation type PivotTable.....	16
4.2.3	Contaminant type PivotTable .....	17
4.2.4	GPRA PivotTable .....	17
<b>5</b>	<b>Contact information</b> .....	<b>17</b>
	<b>Appendix—details on the Annual Trends PivotTables</b> .....	<b>18</b>

# Instructions for accessing SDWIS/FED drinking water data in MS Excel PivotTables

These instructions will get you started using MS Excel PivotTables that contain EPA Safe Drinking Water Information System (SDWIS/FED) data. No prior knowledge of spreadsheets or of MS Excel is required.

Examples from the Annual Trends PivotTables are used in these instructions. These PivotTables include:

1. Inventory—contains summary inventory data on water systems
2. Violation type—contains violations data by violation type
3. Rule—contains violations data by contaminant types and rules
4. Contaminant—contains the number of MCL and MRDL violations, by contaminant
5. GPRA—contains data on health-based violations, and can be used to calculate GPRA (Government Performance and Results Act) data

These instructions are organized as follows: Section 1 describes what PivotTables are and defines some terms. Section 2 lists what data are contained in the Annual Trends PivotTables, and Section 3 tells you how to use them. Some mistakes to avoid are listed in Section 4.

And to make things even easier, each PivotTable contains a full description, brief instructions, and tips/mistakes to avoid.

In addition, several additional PivotTables have been built. You can download PivotTables from:

[www.epa.gov/safewater/data/getdata.html](http://www.epa.gov/safewater/data/getdata.html)

## 1 What's a PivotTable?

A PivotTable is a Microsoft Excel tool. To stoop to “techie” terms, they are multidimensional databases (MDBs) that provide online analytical processing, or OLAP. This means they enable you to quickly summarize, cross-tabulate, and analyze large amounts of data. You can pivot, or rotate, rows and columns to see different summaries of the source data, filter the data, and drill-down to the details in the underlying source data.

For example, you can easily find the number of systems reporting a certain type of violation, in a certain year, for a certain state. To see the data in context, you can compare the results to those of other states, or to other states in that region. To look for trends you can build a graph to compare the results across several years. To investigate an unusually high or low result you can see if the problem stems from a certain contaminant or rule, within a certain system type, size category, or primary source. Using the PivotTables described in these instructions, you could probably perform this analysis in less than 15 minutes.

In addition, you can easily build PivotTables to organize and analyze additional data, and to greatly simplify your SDWIS/FED queries.

### A few terms

PivotTables are primarily made up of “Dimensions” and “Measures.”

**Measures** contain the core information or source data—the numbers, the facts. These include the number of systems and population served by them, as well as the number of violations reported, the number of systems reporting violations, and the population served by these systems.

**Dimensions** organize or categorize the measures. Dimensions include state, water system type, size category, fiscal year, etc.

Each dimension contains various “**Members.**” The SizeCat5 dimension, for example, includes 5 members—one for each size category. The State dimension includes individual states as members.

A sample PivotTable is shown below. It contains inventory data on water systems.

				SizeCat5				
PWSType	GwSw	Data	1_Very_Small	2_Small	3_Medium	4_Large	5_Very_Large	
CWS	gw	TotSys	26,282	10,209	2,658	1,467	70	
		TotPop	4,234,070	13,929,151	15,134,120	37,550,411	17,949,091	
	sw	TotSys	3,372	3,911	2,090	2,179	316	
		TotPop	690,026	6,119,272	12,380,594	65,161,431	108,355,716	
NTNCWS	gw	TotSys	15,917	2,548	85	11		
		TotPop	2,207,480	2,508,977	449,898	275,638		
	sw	TotSys	428	157	17	6		
		TotPop	75,007	198,234	107,844	229,277		
TNCWS	gw	TotSys	81,631	2,584	86	15	2	
		TotPop	7,127,312	2,486,784	459,247	424,914	850,000	
	sw	TotSys	1,700	132	25	8	2	
		TotPop	169,919	170,345	139,259	179,299	2,144,000	

Here the measures are ‘TotSys’ and ‘TotPop’. Dimensions including ‘SizeCat5’, ‘PWSType’, and ‘GwSw’ categorize these measures. The members in each dimension are also listed. For example, ‘CWS’, ‘NTNCWS’, and ‘TNCWS’ are members in the ‘PWSType’ dimension.

## 2 What's in the Annual Trends PivotTables?

Below is an overview of the types of data contained in the Annual Trends PivotTables, which are used as examples in these instructions.

These tables are updated each January, but do not include Chemical MR violations, which are added the following quarter, in April.

Additional details are provided in the Appendix. It includes a list, and description, of Measures and Dimensions included in each PivotTable, and a table that describes which combinations of contaminant codes and violation codes apply to which drinking water rules.

### 2.1 Inventory

The Inventory PivotTable contains summary statistical information on the number of water systems and their population served. Only currently active systems are included (some water systems are seasonal and may be inactive at the time the database is frozen).

The following measures—

- TotSys—total # of water systems and/or the
- TotPop—total population served by water systems

—can be categorized by any combination of the following dimensions:

- EPA Region
- Geography Type
- State
- Primary County served
- Water system type (this is explained in the table below)
- Primary source of water
- Three different population size categories
- Owner type (this is explained in the table below)

### 2.2 Violation type

The Violation type PivotTable contains reported violations by violation type for the last ten fiscal years. Violation types include MCL, MRDL, TT, M/R, Other, and Total (these acronyms are spelled-out in the table below).

The following measures—

- Viols—# of violations reported and/or the
- SysInViol—# of water systems reporting violations and/or the
- PopInViol—population served by water systems reporting violations

—can be categorized by any combination of the following dimensions:

- Fiscal year
- Violation type
- EPA Region
- Geography Type
- State
- Water system type
- Primary source of water

- Three different population size categories

### **2.3 Rule**

The Rule PivotTable contains reported violations for certain contaminants and rules for the last ten fiscal years. These include TCR, SWTR, Interim and LT1 Enhanced SWTR, Stage 1 DBP, TTHM pre-Stage 1 and other VOCs, SOCs, Nitrates, Arsenic and other IOCs, Radionuclides, and LCR (these acronyms are spelled-out in the table below). These are further organized by violation type: MCL, TT and M/R.

The following measures—

- Viols—# of violations reported and/or the
- SysInViol—# of water systems reporting violations and/or the
- PopInViol—population served by water systems reporting violations

—can be categorized by any combination of the following dimensions:

- Contaminant type or rule
- Violation type
- Fiscal year
- EPA Region
- Geography type
- State
- Water system type
- Primary source of water
- Three different population size categories

### **2.4 Contaminant**

The Contaminant PivotTable contains reported violations for federally-regulated MCL and MRDL violations for the last nine fiscal years.

The following measure—# of violations reported—can be categorized by any combination of the following dimensions:

- Contaminant type or rule
- Violation type
- Fiscal year
- EPA Region
- Geography type
- State
- Water system type
- Primary source of water
- Three different population size categories

### **2.5 GPRA**

The GPRA PivotTable contains data on systems reporting health-based violations (MCL, MRDL and TT violations) for fiscal years since 1993.

The following measures—

- Viols—# of health-based violations reported and/or the
- SysInViol—# of water systems reporting health-based violations and/or the
- PopInViol—population served by water systems reporting health-based violations

- TotSys—total # active water systems
- TotPop—total population served by active water systems

—can be categorized by any combination of the following dimensions:

- Fiscal year
- EPA Region
- Geography type
- State
- Water system type
- Primary source
- Three different population size categories

With the GPRA pivot tables you can double-click any data cell to get a list of the underlying water systems reporting health-based violations.

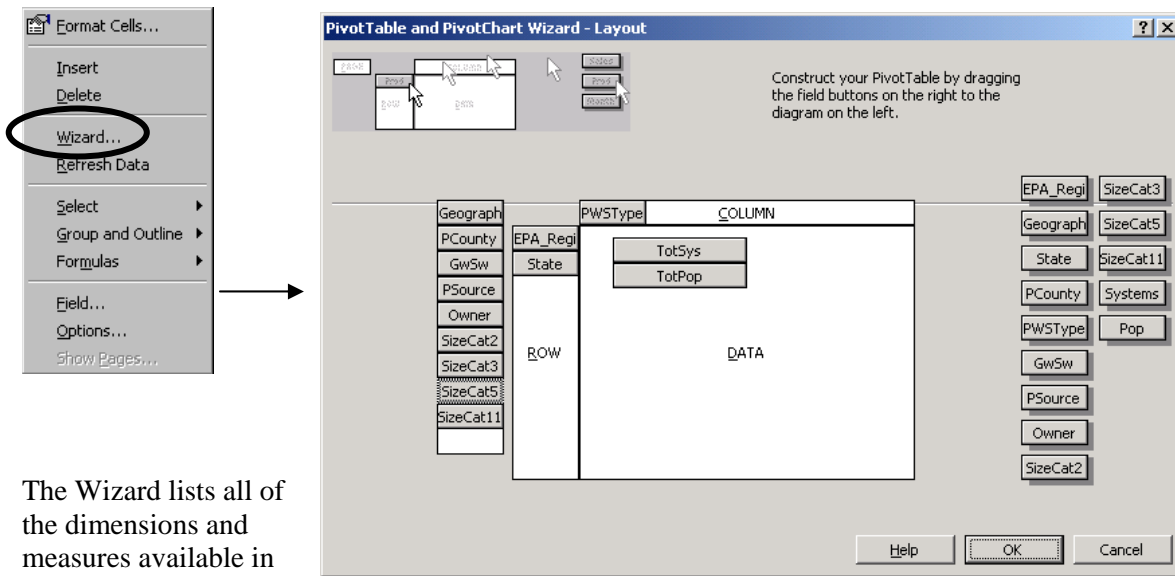
### 3 How to use PivotTables

This section covers basic information to get you started using PivotTables. More information is available in the help menus in Excel. Also, feel free to give me a call if you have any questions.

#### 3.1 Just the basics

##### 3.1.1 Adding or removing dimensions and measures

Use the Wizard to add or remove dimensions and measures in a PivotTable. To get to the Wizard, select any cell inside the PivotTable and right-click on it. Then select 'Wizard...'.



The Wizard lists all of the dimensions and measures available in the PivotTable (which in this case is the Inventory PivotTable) on the right side. The measures (in this case “TotSys” and “TotPop”) are listed last, by convention.

The dimensions and measures placed in the “Row area” “Column area” and “Data area” on the left side of the Wizard reflect the structure of the PivotTable. The Wizard shown above underlies the following PivotTable:

	Geography_type	(All)			
4	Geography_type	(All)			
5	PCounty	(All)			
6	GwSw	(All)			
7	PSource	(All)			
8	Owner	(All)			
9	SizeCat2	(All)			
10	SizeCat3	(All)			
11	SizeCat5	(All)			
12	SizeCat11	(All)			
13					
14			PWSType		
15	EPA_Region	State	Data	CWS	NTNCWS
16	01	01	TotSys	1	2
17			TotPop	41,000	60
18		CT	TotSys	581	636
19			TotPop	2,886,573	124,288
20		MA	TotSys	521	251
21			TotPop	8,942,825	69,791
22		ME	TotSys	396	377
23			TotPop	618,693	72,033
24		NH	TotSys	708	449
25			TotPop	821,209	91,438
26		RI	TotSys	84	77
27			TotPop	976,078	26,929
28		VT	TotSys	436	241
29			TotPop	447,464	43,983
30	01	TotSys		2,727	2,033
31	01	TotPop		14,533,842	426,522
32	02	02	TotSys	8	3
33			TotPop	7,650	300

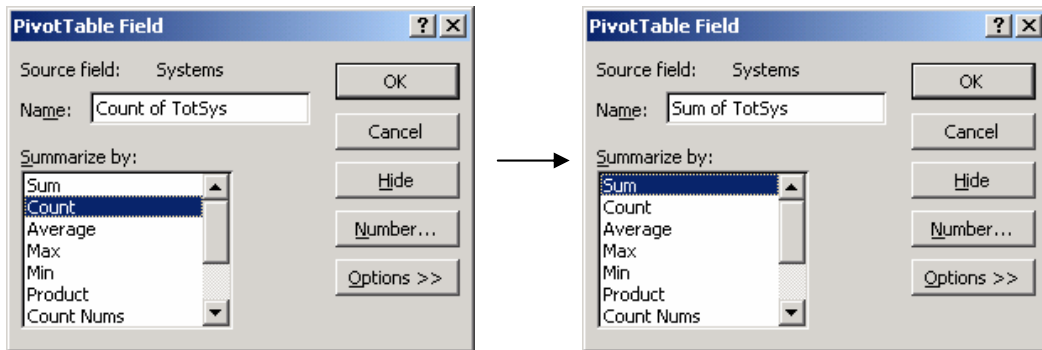
As reflected in the Wizard:

- 'PWSType' is selected in the Column area
- 'EPA\_Region' and 'State' are selected in the Row area
- both measures—"TotSys" and "TotPop"—are selected in the Data area

You can construct or alter a PivotTable by dragging dimensions and measures to Row, Column, or Data areas. These areas are highlighted in the copy of the Wizard above. Make sure you don't place dimensions fields in the Data area, or measures in the Row, Column, or Page areas—you could end up with meaningless results.

For any dimension placed in the Page area, you can view one dimension member at a time. For example, if you placed 'Region' in the Page area you could look at the data for any single EPA Region at a time. If '<all>' Regions are selected then all Regions are included.

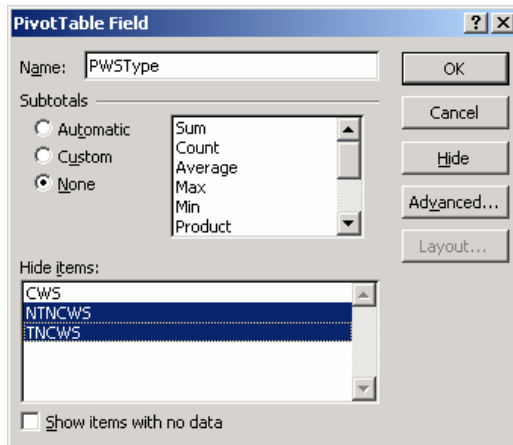
It's important to remember that whenever you drag a measure onto the Data area, you'll need to specify that the 'Sum' of the values is used. The default is 'Count'. To correct this, double-click on the measure; the following screen will appear:



Select 'Sum'.

### 3.1.2 Filtering data members

You can filter-out members within any dimension that's currently used in a PivotTable. Select a dimension (which are colored gray) and either right-click or double-click on it. The screen below will appear. Then select which members you'd like to hide:



In the example above, we've hidden NTNCWSs and TNCWSs from the 'PWSType' dimension. Only CWSs will be included.



Be careful: it's easy to forget you've hidden members. It's a good idea to frequently right-click on a dimension and double-check if any members are hidden. Also be aware that if you move a dimension out of the PivotTable area and bring it back in, any members you'd previously hidden will remain hidden. However, the data from all members will be counted if the dimension is removed from your PivotTable.

### 3.1.3 Rearranging the dimensions and measures to facilitate analysis

You can rearrange dimensions and measures in a PivotTable "on the fly" without using the Wizard. To do this, point to a dimension you're interested in moving, press and hold the mouse button, and move it around. You can move it from the Row area to the Column area or the Page area. You can place it before or after another dimension in the Row or Column area, or move it to or from the Page area. In addition, you can remove a dimension from your PivotTable by simply dragging it out of the PivotTable area. Fiddle around with this for a while to see how it works.

I usually build a PivotTable by selecting all the required dimensions and measures I'll need; then I organize it "on the fly" by dragging them around. Below is a simple example.

		PWSType	Data							
		CWS			NTNCWS			TNCWS		
FY	VType_pt	Viols	SysInViol	PopInViol	Viols	SysInViol	PopInViol	Viols	SysInViol	PopInViol
2005	a_MCL	9,739	4,472	19,228,465	1,980	1,207	352,038	5,043	3,977	542,026
	b_MRDL	4	4	424,859	1	1	110			
	c_TT	3,036	1,625	16,879,450	853	550	164,644	502	215	86,648
	d_MR	124,605	14,403	62,511,599	54,561	5,591	1,668,159	37,505	17,556	2,264,085
	e_Other	18,414	10,258	23,442,800	717	491	122,177	4,800	2,284	344,411
	f_Total	155,798	22,902	91,626,483	58,112	6,822	2,044,368	47,850	21,258	2,816,995
2004	a_MCL	5,562	3,455	15,060,424	1,381	1,033	288,280	5,032	3,974	549,673
	b_MRDL	6	4	3,617						
	c_TT	2,168	1,409	16,588,301	681	474	130,242	415	195	51,915
	d_MR	96,225	13,693	51,043,156	37,514	4,963	1,402,090	38,705	17,334	2,207,974
	e_Other	14,459	8,947	17,833,120	538	356	88,003	3,028	1,697	234,291
	f_Total	118,420	21,200	79,058,089	40,114	5,999	1,713,877	47,180	20,680	2,708,036

In the (Violation type) PivotTable above, the measures are listed in a columnar orientation. The default is a row orientation. In the second PivotTable, I moved the 'PWSType' field from the Column area to the Row area and placed before 'FY'. In the last PivotTable, I moved 'VType' before the 'PWSType' and 'FY' dimensions.

		PWSType	FY	VType_pt	Data		
		CWS			Viols	SysInViol	PopInViol
CWS	2005	a_MCL	9,739	4,472	19,228,465		
		b_MRDL	4	4	424,859		
		c_TT	3,036	1,625	16,879,450		
		d_MR	124,605	14,403	62,511,599		
		e_Other	18,414	10,258	23,442,800		
		f_Total	155,798	22,902	91,626,483		
CWS	2004	a_MCL	5,562	3,455	15,060,424		
		b_MRDL	6	4	3,617		
		c_TT	2,168	1,409	16,588,301		
		d_MR	96,225	13,693	51,043,156		
		e_Other	14,459	8,947	17,833,120		
		f_Total	118,420	21,200	79,058,089		
NTNCWS	2005	a_MCL	1,980	1,207	352,038		
		b_MRDL	1	1	110		
		c_TT	853	550	164,644		
		d_MR	54,561	5,591	1,668,159		
		e_Other	717	491	122,177		
		f_Total	58,112	6,822	2,044,368		
NTNCWS	2004	a_MCL	1,381	1,033	288,280		
		b_MRDL					
		c_TT	681	474	130,242		
		d_MR	37,514	4,963	1,402,090		
		e_Other	538	356	88,003		
		f_Total	40,114	5,999	1,713,877		
TNCWS	2005	a_MCL	5,043	3,977	542,026		
		c_TT	502	215	86,648		
		d_MR	37,505	17,556	2,264,085		
		e_Other	4,800	2,284	344,411		
		f_Total	47,850	21,258	2,816,995		
		TNCWS	2004	a_MCL	5,032	3,974	549,673
c_TT	415			195	51,915		
d_MR	38,705			17,334	2,207,974		
e_Other	3,028			1,697	234,291		
f_Total	47,180			20,680	2,708,036		

		VType_pt	PWSType	FY	Data		
		a_MCL	CWS		Viols	SysInViol	PopInViol
a_MCL	2005	CWS	9,739	4,472	19,228,465		
		NTNCWS	1,980	1,207	352,038		
		TNCWS	5,043	3,977	542,026		
		CWS	5,562	3,455	15,060,424		
		NTNCWS	1,381	1,033	288,280		
		TNCWS	5,032	3,974	549,673		
b_MRDL	2005	CWS	4	4	424,859		
		NTNCWS	1	1	110		
		TNCWS	6	4	3,617		
		CWS	2,168	1,409	16,588,301		
		NTNCWS	681	474	130,242		
		TNCWS	2,168	1,409	16,588,301		
c_TT	2005	CWS	3,036	1,625	16,879,450		
		NTNCWS	853	550	164,644		
		TNCWS	3,036	1,625	16,879,450		
		CWS	124,605	14,403	62,511,599		
		NTNCWS	54,561	5,591	1,668,159		
		TNCWS	124,605	14,403	62,511,599		
d_MR	2005	CWS	18,414	10,258	23,442,800		
		NTNCWS	717	491	122,177		
		TNCWS	18,414	10,258	23,442,800		
		CWS	96,225	13,693	51,043,156		
		NTNCWS	37,514	4,963	1,402,090		
		TNCWS	96,225	13,693	51,043,156		
e_Other	2005	CWS	14,459	8,947	17,833,120		
		NTNCWS	538	356	88,003		
		TNCWS	14,459	8,947	17,833,120		
		CWS	155,798	22,902	91,626,483		
		NTNCWS	58,112	6,822	2,044,368		
		TNCWS	155,798	22,902	91,626,483		
f_Total	2005	CWS	155,798	22,902	91,626,483		
		NTNCWS	58,112	6,822	2,044,368		
		TNCWS	155,798	22,902	91,626,483		
		CWS	118,420	21,200	79,058,089		
		NTNCWS	40,114	5,999	1,713,877		
		TNCWS	118,420	21,200	79,058,089		

In short, these data can be displayed in any combination.

### 3.1.4 Drilling down

You can get to the core data in any PivotTable data cell by double-clicking on the data in any data cell. Excel will automatically create another worksheet containing the core data.

As an example of how to drill down, suppose you were interested in finding out more about CWSs served primarily by GUP (purchased ground water under the influence of surface water) in Region 6. A PivotTable which contains the required dimensions and measures is shown below.

		PSource						
EPA_Region	Data	GU	GUP	GW	GWP	SW	SWP	
01	TotSys	29	1	10,064	117	332	247	
	TotPop	154,883	400	4,004,856	41,279	8,768,924	2,883,093	
02	TotSys	236	2	12,751	208	636	838	
	TotPop	186,855	1,797	8,878,511	289,255	20,706,349	5,844,639	
03	TotSys	173	16	16,640	110	622	661	
	TotPop	165,505	23,713	5,000,728	83,195	16,665,457	4,882,517	
04	TotSys	104	13	17,995	430	719	1,366	
	TotPop	315,319	14,300	23,893,971	1,141,830	21,506,927	8,146,406	
05	TotSys	34	2	43,642	1,105	481	1,168	
	TotPop	210,850	5,000	18,661,911	641,889	17,209,861	9,918,797	
06	TotSys	56	5	9,102	527	760	1,697	
	TotPop	72,156	5,084	13,246,325	458,127	19,478,180	6,691,381	
07	TotSys	33	36	5,911	389	201	571	
	TotPop	635,991	60,613	4,531,750	281,015	5,243,769	1,335,934	
08	TotSys	115	12	5,623	301	461	550	
	TotPop	71,166	1,158	3,491,746	264,785	5,752,115	1,274,273	
09	TotSys	92	5	9,005	115	870	327	
	TotPop	278,745	5,058	14,562,628	126,371	19,429,870	12,842,647	
10	TotSys	50	2	9,364	166	618	202	
	TotPop	193,033	1,965	5,215,164	771,757	4,052,355	1,079,750	

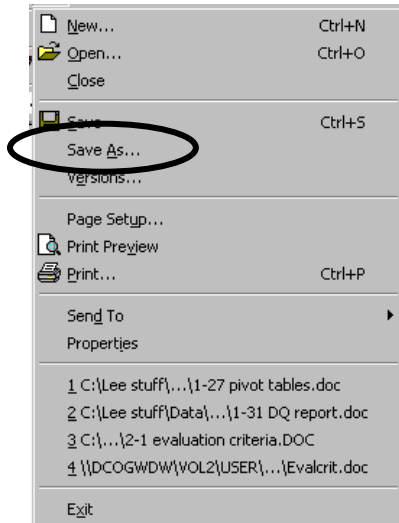
When you double-click on a data cell (in this example, the data cell that contains 5 systems), Excel will create the following worksheet:

EPA_Region	Geography_type	State	PCounty	PWSType	GwSw	PSource	Owner	SizeCat2	SizeCat3	SizeCat5	SizeCat11	Systems	Pop
06	State	TX	Palo Pinto	CWS	sw	GUP	Local_Govt	a_<10,000	a_<=3,300	2_Small	c_501-1,000	1	530
06	State	TX	Harris	CWS	sw	GUP	Private	a_<10,000	a_<=3,300	1_Very_Small	b_101-500	1	465
06	State	TX	Fort Bend	CWS	sw	GUP	Local_Govt	a_<10,000	b_3,301-50,000	3_Medium	e_3,301-10,000	1	3429
06	State	OK	Adair	CWS	sw	GUP	Local_Govt	a_<10,000	a_<=3,300	1_Very_Small	b_101-500	1	316
06	State	AR	Newton	CWS	sw	GUP	Local_Govt	a_<10,000	a_<=3,300	1_Very_Small	b_101-500	1	344

Unfortunately, the core data underlying these Annual Trends PivotTables doesn't go all the way down to the water system level due so we can keep the size of the file within reason. However, you can double-click the GPRAs pivot tables and view the systems reporting health-based violations in any year. In addition, you can also download inventory PivotTables that includes information at the water system level. Due to a row limitation in Excel, you can use view only 65,000 rows at a time.

### 3.1.5 Saving PivotTables

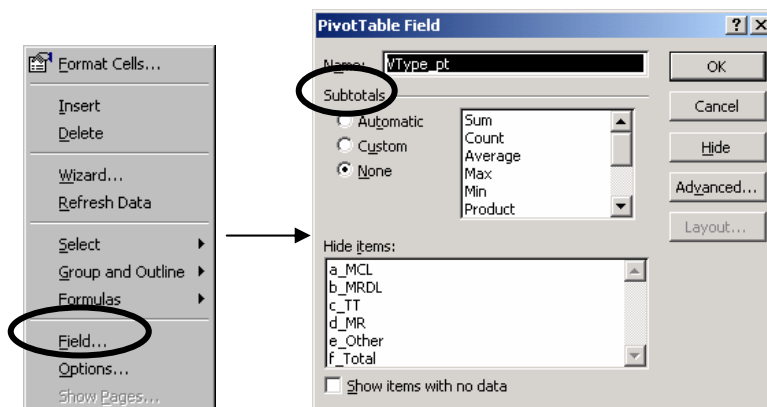
The PivotTables are read-only, which means they will appear in the same format each time that you open them. If you want to save changes to a PivotTable you've modified, just save it under a different name. To do this, press the 'File' menu and select 'Save As...' and save it under a different name.



## 3.2 Organizing data, printing, graphing

### 3.2.1 Adding and removing dimension subtotals

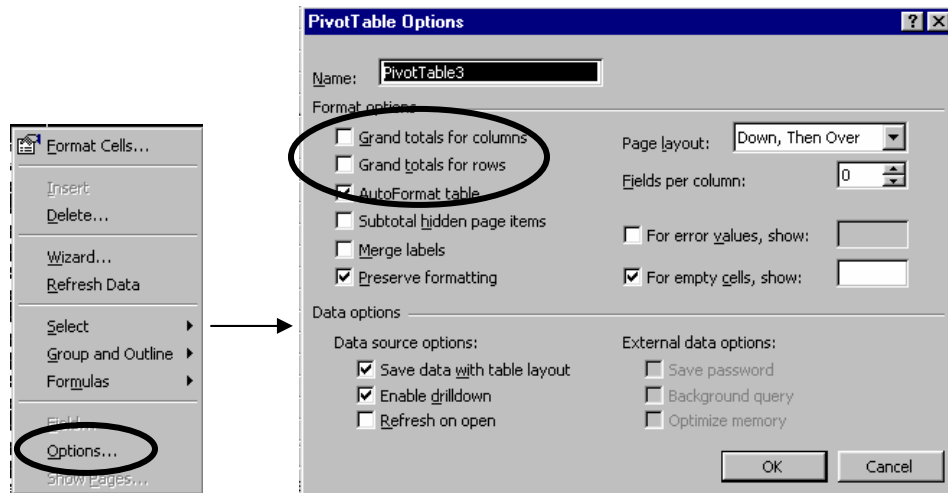
Subtotals of data items in a dimension can be very helpful. To add or remove them, first point your cursor on the dimension of interest. Next, right-click your mouse, and then select 'Field...', as shown below:



Be careful when using subtotals and row or column totals. It's easy to sum things that shouldn't be summed, such as across violation types or contaminants/rules. For example, if you simply added the number of systems having MCL, M/R, and Other violation types, one system having one or more of each type would be counted three times instead of once. In the example above, you should select 'None' as the subtotal of the 'VType' dimension.

### 3.2.2 Adding or removing row or column totals

To add or remove row or column totals right-click on any data cell in the PivotTable, and then select 'Options...'. This is shown below:



You can also remove row or column totals by selecting them in the PivotTable, right-clicking your mouse, and selecting 'Delete...'

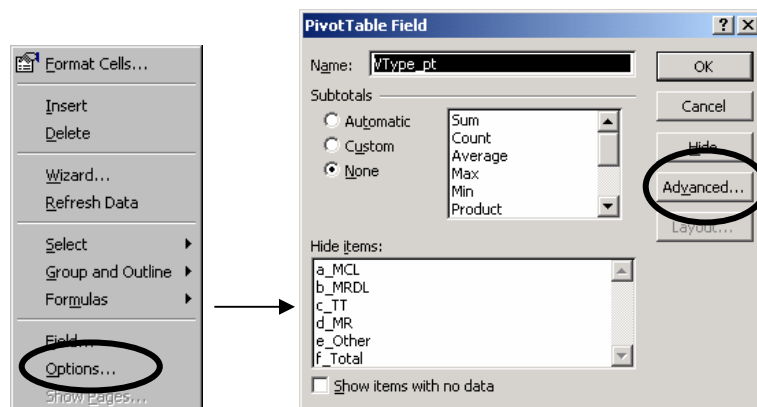
As mentioned above, be careful when using totals by row or column. Summing across violation types or contaminants/rules can yield erroneous results.

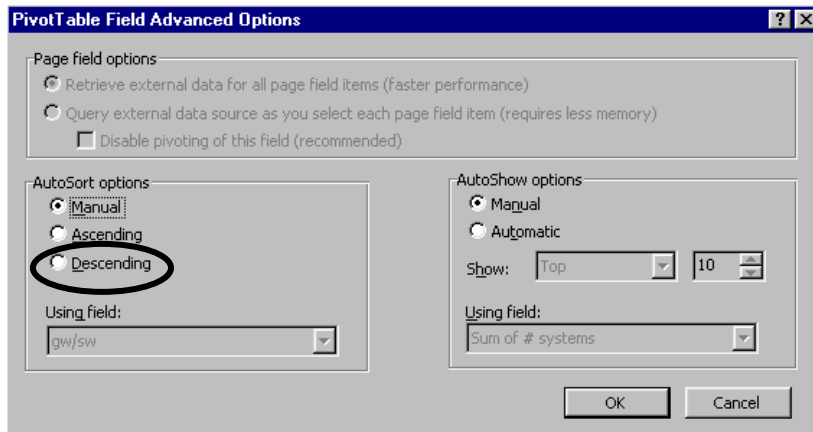
### 3.2.3 Changing labels

To make your PivotTable labels more understandable, you can simply click on the label and type over them. For example, instead of listing State "01" you might want to type something like "Tribal systems in Region 1." These changes will be lost if you remove the field from the PivotTable and later restore it.

### 3.2.4 Sorting data in ascending or descending order

You can choose whether members in Row/Column fields are listed in ascending or descending order. For example, when listing FY in the Row area, my personal preference is to list the most recent year first. To sort dimension members in descending order double-click on the field, select 'Advanced...' and then 'Descending' under 'AutoSort options'.

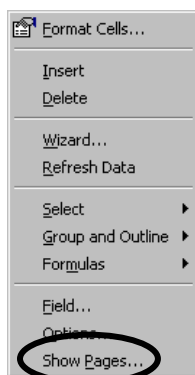




### 3.2.5 Organizing data using the Page area

As discussed above, using dimensions in the Page area (which is above the table and to the right), you can quickly filter-out (hide) all but a single member in a dimension. For example, if you're primarily interested in data for a certain state, simply place 'State' in the Page area and select that individual state from the drop-down toolbar. This is quicker and easier than placing 'State' in the Row or Column area, right-clicking your mouse on the 'State' dimension, selecting 'Field...' and hiding all the other states one at a time.

The Page area can also help you organize large amounts of data. With a dimension in the Page area, you can make a separate PivotTable for each member in that dimension. For example, with the 'Region' dimension listed in the Page area you can make a separate PivotTable in a separate worksheet for each region. To do this, point to a cell in the PivotTable, and right-click your mouse. The following screen will appear.

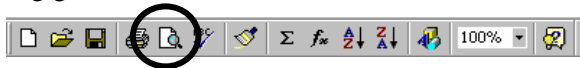


Select 'Show Pages' to make a separate PivotTable for each member in the dimension.

### 3.2.6 Printing

It can be very easy to make large or even huge PivotTables®, and so difficult to organize them to print in a logical way. Here are a few quick tips.

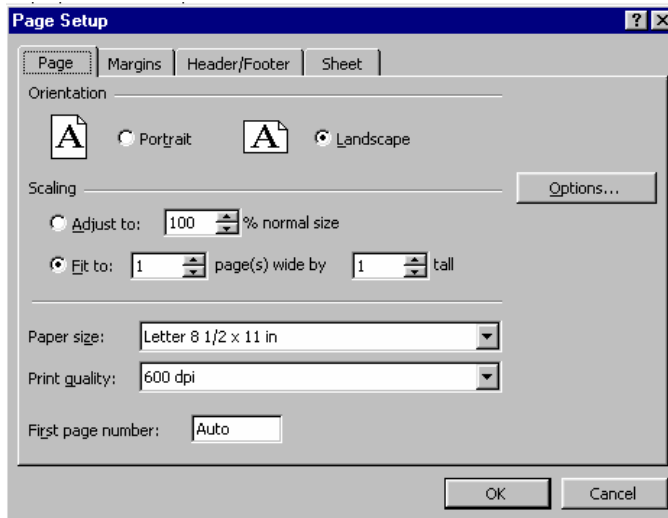
Before printing, press the 'Print Preview' menu button. The button looks like a white page with a magnifying glass:



You can also get there from the main menu bar by selecting 'File' and then 'Print Preview'. It will show you how your PivotTable will set up on the page, as well as how many pages it will take to print. The top menu bar is shown below:



From here press the "Setup..." menu button to alter how the spreadsheet will print. The 'Page Setup' screen will appear:



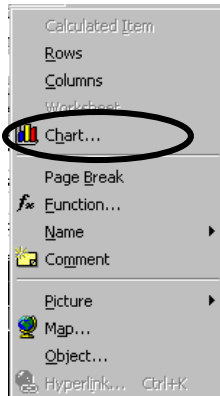
You can see whether your PivotTable will print better in either Landscape or Portrait orientation. To fit the most on a page, select Landscape and then Portrait and see how small the data will print. Try not go below 60% or it will be difficult to read.

For large PivotTables® that will take several pages to print, the 'Page Break Preview' feature will enable you to select where each page break will occur. You can get there by pressing the 'Print Preview' button and then the 'Page Break Preview' button. It's also usually a good idea to print row and/or column headings at the top of each page. You can specify these by selecting the 'Sheet' button from the 'Page Setup' screen.

## Graphing

Excel has powerful graphing capabilities. There are 'Wizards' to help you make graphs of your data, and a little experimenting will get you far. Here's a few quick tips to get you started:

- Select only one measure at a time
- Remove all subtotals and totals.
- With your mouse, select any data field in the PivotTable.
- Under the 'Insert' menu, select 'Chart...'



From there, a ‘Wizard’ will guide you. Once you’ve made a chart you can change the underlying data and it will graph automatically. For example, suppose you make a chart using the Violation type PivotTable and graph the number of violations by FY (remember, you’ve got to include the ‘VType’ field when using this PivotTable). You can replace the ‘Viols’ measure with ‘SysInViol’ or ‘PopInViol’ measures and Excel will automatically change the chart.

### **3.3 Some additional functions not covered here**

There are numerous elements of PivotTables not covered here. If you would like to learn more about them, here are a couple of functions you might want to look into.

One of them is the PivotTable toolbar. Using it will enable you to go directly to the Wizard or to format a dimension or measure without right-clicking. There are additional functions there as well. I haven’t discussed how to use the toolbar here since some of the functions are slightly more advanced. However, fiddling will get you far: there are excellent help menus available, and Excel is very well thought-out—learning how to do additional things is almost intuitive.

Another powerful tool is the ability of Excel to calculate measures. Right-click on a cell in the PivotTable and select ‘formulas’, then ‘Calculated field...’ to get started. The GPRA PivotTable contains some simple calculated measures.

## 4 So, what do we make of all these numbers?

Now you can easily extract and analyze large amounts of data from SDWIS/FED. Unfortunately, there's still a few ways to get incorrect or meaningless numbers. Please spend some time to understand this section.

### 4.1 General tips and mistakes to avoid

#### 4.1.1 Don't forget to select the 'Sum of...' measures each time you insert them into a PivotTable

The default is 'Count of...'. To correct this, double-click the measure, and select "Sum." See 3.1.1.

#### 4.1.2 Don't refresh your data unless you have the original data tables on your computer

The core data tables are not included with the Annual Trends PivotTables. If you select a cell in the PivotTable and right-click your mouse, you'll see the 'Refresh Data' button. If you select this and you don't have the core data tables on your computer, the program will crash and you'll have to reboot.

#### 4.1.3 Don't put measures in a Column, Row or Page area and don't put dimensions in the Data area

You'll usually end up with meaningless results. One example where this could work is if you've got data on individual water systems, which include the number of violations reported by each system. By placing this measure in the Column, Row or Page area, you could find the number of systems reporting different numbers of violations.

#### 4.1.4 Don't forget about filtered or "hidden" data members

As discussed above, it's easy to forget you've hidden data members from a dimension. It's especially easy when dimensions are placed in the Page area. You can check by double-clicking the dimension(s) in question.

#### 4.1.5 Don't add the number of systems in violation or population affected across violation types or contaminant types.

Since systems can have many different violation types and/or contaminant/rule violations, the number of systems in violation (and population served) cannot simply be added. For example, if you added the number of systems having MCL, MRDL, M/R, and Other violation types, one system having one or more of each type would be counted three times instead of once.

There are ways to get accurate counts of systems and population affected across violation types:

- The GPRA PivotTable can list the number of systems reporting health-based violations (MCL, MRDL and/or TT violations), and the population affected.
- The Violation type PivotTable contains the field "f\_Total," that lists the number of systems, and population served, by water systems reporting any type of violation.



#### 4.1.6 Be careful when interpreting population served data, especially when looking for trends

When analyzing population served by systems having a violation, be careful of reading too much into the data, or to draw conclusions looking only at this parameter. Especially if there's only a small number of systems in violation, and/or those in violation tend to be smaller systems, one or a few very large systems can significantly skew the results. For example, there would appear to be a serious problem if the total population affected by a contaminant, which has been roughly 500,000 a year, suddenly jumps to 10,000,000. This increase, by a factor of 20, may indeed indicate a serious health problem. However, it may also be due to one very large system having a violation, and is therefore not indicative of a trend.

When we developed the Trends Report we wanted to cite trends primarily using population affected. However, the findings were generally unreliable and could even be misleading given the fact that one or a few very large systems could significantly skew the findings, as discussed above. Many of the erratic jumps in population served estimates are due to what we call "noise" in the data.

The number of systems in violation provides a more reliable measure of compliance when assessing trends, as it is not susceptible to bias by large systems. Looking at the total number of violations can also be subject to bias by one or a few systems having several violations, but this measure is still significantly less volatile than the population served numbers.

The higher the population affected by systems in violation (which comes with more systems in violation), the less susceptible will be the results to a very large system, and the more confidently you can use the data. So be careful not to parse the data too far (by state, by water system type, etc.), as this lessens the total population served and increases the influence of very large systems on the results.

#### 4.1.7 Organizing 'State' and 'PSource' data

When using the 'State' dimension, if you also specify the 'EPA\_Region' dimension, place it before the 'State' dimension, and subtotal it, you'll get results by Region and State in the same PivotTable. In the same way, you can place the 'GwSw' dimension before 'PSource' and subtotal it.

### 4.2 *Tips by PivotTable*

#### 4.2.1 Inventory PivotTable

- Since several states can have the same county names, don't use the 'PCounty' dimension unless it's preceded by the 'State' dimension.
- Don't add population served across system types. Since all these types of water systems can serve an individual there's a lot of double-counting in the totals. For this reason, OGWDW usually lists the total population served by PWSs as those served by CWSs only.

#### 4.2.2 Violation type PivotTable

- Don't make a PivotTable without including the 'Vtype' dimension in the Row or Column area. If you do, the sum of MCL+MRDL+TT+M/R+Other+Total will be listed, which will result in double the number of violations.

- Don't use subtotals or grand totals across 'VType'. These are already contained in Item 'f\_Total'.

#### 4.2.3 Contaminant type PivotTable

- Don't make a PivotTable without specifying the 'CType' dimension in the Row or Column area.
- Don't use subtotals or totals to sum the number of systems in violation or population affected across CType. You can use the 'Violation type' PivotTable to sum by violation type, or overall. In addition, you can use the 'GPRA' PivotTable if you want to get the total number of systems and population affected by systems reporting any MCL and/or TT violations.

#### 4.2.4 GPRA PivotTable

- For the reasons discussed in 4.1.6, it's not a good idea to parse the GPRA numbers too far. In my opinion, you can get more reliable estimates for analyzing trends using the number of systems reporting violations. Each system reporting a certain violation is counted equally, as opposed to estimates of the population affected, where a Very Large system can easily skew the results.
- In the PivotTables® GPRA estimates are included for all water system types. Remember, however, that the official GPRA measure only includes CWSs.

## 5 Contact information

PivotTables®, such as those described in this paper, are one piece of a continuum of products being developed to make SDWIS/FED data more accessible and usable. Your feedback will help us improve these products and better assess your SDWIS/FED data needs. We can also make PivotTables® to meet your specific needs, and show you how to make your own.

Contact Lee Kyle to provide feedback, ask questions, or to learn more about some of the other PivotTables available. (202) 564-4622 or [kyle.lee@epa.gov](mailto:kyle.lee@epa.gov).

You can download copies of the latest PivotTables and instructions from the web at:

<http://www.epa.gov/safewater/data/getdata.html>

## Appendix—details on the Annual Trends PivotTables

The next three pages include a list of Measures and Dimensions used in the Annual Trends PivotTables. This information is also included in the Annual Trends spreadsheet.

Pivot table:								<b>Measures and dimensions included in the pivot tables</b> <b>Notice:</b> EPA is aware of inaccuracies and underreporting of some data in the Safe Drinking Water Information System. We are working with the states to improve the quality of		
Inventory	Violation type	Rule	Contaminant	GPRA	GPRA ST01 --rules by 12/2001	GPRA ST2--rules 01/2002+				
<b>Measures:</b>										
TotSys	X			X	X	X	Number of current, active systems (or, # systems subject to a GPRA measure)			
TotPop	X			X	X	X	Population served by current, active systems (or, pop served by systems subject to a GPRA measure)			
Viols		X	X	X	X	X	X	X	# of violations reported	
SysInViol		X	X		X	X	X	X	# of systems reporting violations	
PopInViol		X	X		X	X	X	X	Population served by systems reporting violations	
GPRA-%sys					X	X	X		% of systems not reporting any health-based violations (GPRA applies to CWSs only)	
GPRA-%pop					X	X	X		% of population served by systems not reporting any health-based violations (GPRA applies to CWSs only)	
<b>Dimensions:</b>										
						<b>Members:</b>		Description of members:		
EPA_Region	all					I	CT, ME, MA, NH, RI, VT			
						II	NJ, NY, PR (Puerto Rico), VI (Virgin Islands)			
						III	DE, DC, MD, PA, VA, WV			
						IV	FL, FL, GA, KY, MS, NC, SC, TN			
						V	IL, IN, MI, MN, OH, WI			
						VI	AR, LA, NM, OK, TX			
						VII	IA, KS, MO, NE			
						VIII	CO, MT, ND, SD, UT, WY			
						IX	AZ, CA, HI, NV, AS (American Samoa), GU (Guam), MP (Northern Marianas Islands), PW (Palau), NN (Navajo Nation)			
						X	AK, ID, OR, WA			
Geography type	all					State				
						Territory	AS (Am. Samoa), GU (Guam), MP (N. Marianas Islands), PR (Puerto Rico), PW (Palau), VI (Virgin Islands)			
						Tribal	NN (Navajo Nation) plus numeric states (01 to 10), which refer to tribal systems in that respective region.			
State		all				Numeric states (01 to 10) refer to tribal systems in that respective region.				
PCounty	X			X	X	X	Primary County served by a water system			
PWSType	all					CWS	Community Water System			
						NTNCWS	Non-Transient Non-Community Water System			
						TNCWS	Transient Non-Community Water System			
Gw/sw	all					SW	includes SW, SWP, GU, GUP			
						GW	includes GW and GWP			
PSource	all					Primary source of water:				
						SW	surface water			
						SWP	purchased surface water			
						GU	ground water under the influence of surface water			
						GUP	purchased ground water under the influence of surface water			
						GW	ground water			
Owner	X					F	Federal government			
						S	State government			
						L	Local government			
						M	Mixed public/private			
						N	Native American			
						P	Private			
SizeCat2	all					Size category, based on average daily population served:				
						a	<10,000			
						b	≥10,000			

Pivot table:		Inventory	Violation type	Rule	Contaminant	GPRAs	GPRAs ST01 --rules by 12/2001	GPRAs ST2 --rules 01/2002+	<h3>Measures and dimensions included in the pivot tables</h3> <p><b>Notice:</b> EPA is aware of inaccuracies and underreporting of some data in the Safe Drinking Water Information System. We are working with the states to improve the quality of</p>												
SizeCat5	all								Size category, based on average daily population served: a Very small 25-500 b Small 501-3,300 c Medium 3,301-10,000 d Large 10,001-100,000 e Very large >100,000												
SizeCat11	all								Size category, based on average daily population served: a <=100 b 101-500 c 501-1,000 d 1,001-3,300 e 3,301-10,000 f 10,001-50,000 g 50,001-100,000 h 100,001-250,000 i 250,001-500,000 j 500,001-1,000,000 k >1,000,000												
FY		X	X	X			X	X	Most include ten years of data (GPRAs from FY1993 on). A fiscal year (FY) runs from October 1st to September 30th.												
Vtype		X	X	X					Violation type: MCL Maximum Contaminant Level violation MRDL Maximum Residual Disinfectant Level TT Treatment Technique violation MR Monitoring or reporting violation Other Other violation, including public notice violation Total All violations												
Ctype					X	X			Contaminant type/Rule: a TCR Total Coliform Rule b1 SWTR Total Trihalomethanes, a Volatile Organic Chemical (VOC) b2 LLT1_IESWTR Interim Enhanced Surface Water Treatment Rule and the LT1 (future) Enhanced SWTR e ST1_DBP Stage 1 Disinfectants By-Product Rule g TTHM_pre-ST1 TTHM Rule violations, which was replaced by the ST1 DBP Rule h VOC Other Volatile Organic Chemicals i SOC Synthetic Organic Chemicals j1 Nitrates Nitrates j2 Arsenic j3 Other_IOC Other Inorganic Chemicals k Rads Radionuclides l LCR Lead and Copper Rule												
CName					X				Contaminant name												
CCode					X				Contaminant number												
PWSID						X	X	X	Water system ID#												
PWSName						X	X	X	Water system name												
<b>Frozen database tables:</b>																					
FY06Q1 was frozen in April 2006																					
FY05Q4 was frozen in January 2006																					
FY05Q1 was frozen in April 2005																					
FY04Q4 was frozen in January 2005																					
FY04Q1 was frozen in April 2004																					
FY03Q4 was frozen in January 2004																					
FY03Q1 was frozen in April 2003																					
FY02Q4 was frozen in January 2003																					
FY02Q1 was frozen in early April 2002																					
FY01Q4 was frozen in January 2002																					
FY01Q1 will be frozen in early April 2001																					
FY00Q4 was frozen in early January 2001																					
etc.																					

Below is a description of the coding we use to sort violation codes and contaminant codes into different rules (CType). Only federally-regulated contaminants will get sorted by rule. For example, an MCL for contaminant code 2043, which is regulated for monitoring only, will get a CType of 'not regulated', and the violation will not be included in GPRA.

VCode	VName	Vtype	Applicable rules and contaminant codes (CCodes)							
1	MCL, Single Sample	MCL			Other VOC	SOC 2005/	Nitrates	Other IOC	Rads	Coliform (Pre-TCR)
2	MCL, Average	MCL	DBP 1009,	TTHM pre-'02	2378/ 80,	10/ 15/ 20/	1038,	1005/ 10/ 15/	4000/ 06/	3000
3	Monitoring, Regular	MR	1011, 2456,	2941/ 42/ 43/	2955/ 64/ 68/	31/ 32/ 33/	1040,	20/ 24/ 25/ 35/	10, 4100/	
4	Monitoring, Check/Repeat/Confirmation	MR	2950	44, 2950	69/ 76/ 77/	34/ 35/ 36/	1041	36*/ 45/ 74/	01/02/ 74	
5	Notification, State	Other			79/ 80/ 81/	37/ 39/ 40/		75/ 85/ 94		
6	Notification, Public	Other			82/ 83/ 84/	41/ 42/ 43*/				
7	Treatment Techniques	Other			85/ 87/ 89/	44*/ 46/ 47*/				
8	Variance/Exemption/Other Compliance	Other			90/ 91/ 92/	50/ 51/ 63/				
9	Record Keeping	Other	FBR 0500	IESWTR 0300	96	65/ 67,				
10	Operations Report	Other				2105/ 10,				
11	Non-Acute MRDL	MRDL	DBP 0999, 1006/ 08			2274/ 98,				
12	Treatment Technique No Certif. Operator	TT	DBP 0400			2306/ 26/				
13	Acute MRDL	MRDL	DBP 1008			83/ 88/ 90/				
21	MCL, Acute (TCR)	MCL				92/ 94/ 96/				
22	MCL, Monthly (TCR)	MCL				98, 2400,				
23	Monitoring, Routine Major (TCR)	MR	TCR 3100			2931/ 46/ 59				
24	Monitoring, Routine Minor (TCR)	MR								
25	Monitoring, Repeat Major (TCR)	MR								
26	Monitoring, Repeat Minor (TCR)	MR								
27	Monitoring and Reporting Stage 1	MR	DBP 0400, 0999, 1006/ 08/ 09/ 11, 2456, 2920, 2950							
28	Sanitary Survey (TCR)	Other	SS, TCR 3100							
29	M&R Filter Profile/CPE Failure	MR	IESWTR 0300							
31	Monitoring, Routine/Repeat (SWTR-Unfilt)	MR	SWTR 0200							
36	Monitoring, Routine/Repeat (SWTR-Filter)	MR								
37	Treatment Technique State Prior Approval	TT	IESWTR	DBP 0400						
38	M&R Filter Turbidity Reporting	MR	0300							
39	M&R (FBRR)	MR	FBR 0500							
40	Treatment Technique (FBRR)	TT								
41	Treatment Technique (SWTR)	TT	SWTR 0200							
42	Failure to Filter (SWTR)	TT								
43	Treatment Technique Exceeds Turb 1 NTU	TT	IESWTR							
44	Treatment Technique Exceeds Turb 0.3 NTU	TT	0300							
46	Treatment Technique Precursor Removal	TT	DBP 2920							
47	Treatment Technique Uncovered Reservoir	TT	IESWTR							
51	Initial Tap Sampling for Pb and Cu	MR								
52	Follow -up and Routine Tap Sampling	MR								
53	Initial Water Quality Parameter WQP M&R	MR								
54	Follow -up & Routine E.P. WQP M&R (deleted)	MR	X							
55	Follow -up & Routine Tap WQP M&R (deleted)	MR	X							
56	Initial, Follow -up, or Routine SOWT M&R	MR								
57	OCCT Study Recommendation	TT								
58	OCCT Installation/Demonstration	TT								
59	WQP Entry Point Non-Compliance	TT								
60	WQP Entry Point Non-Compliance (deleted)	TT	X							
61	SOWT Recommendation (deleted)	TT	X							
62	SOWT Installation (deleted)	TT	X							
63	MPL Non-Compliance	TT	LCR 1022,1030							
64	Lead Service Line Replacement (LSLR)	TT	LCR 5000							
65	Public Education	TT								
71	CCR Complete Failure to Report	Other	CCR 7000							
72	CCR INADEQUATE REPORTING	Other								
75	PN Violation for NPDWR Violation	Other	PN 7500							
76	Other Non-NPDWR Potential Health Risks	Other								

\* codes in red are required for monitoring only

Violations counted in a fiscal year have a compliance period overlapping some part of that year. That is, violations must have a start on or prior to the end of the fiscal year (sept 30th) and an end date on or after the beginning of the fiscal year (Octo

last updated 9/2005