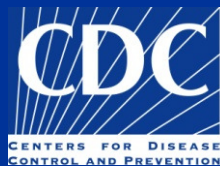
Using the Chart & Statistics Tool and Groups

Steven Stroika

April 2011



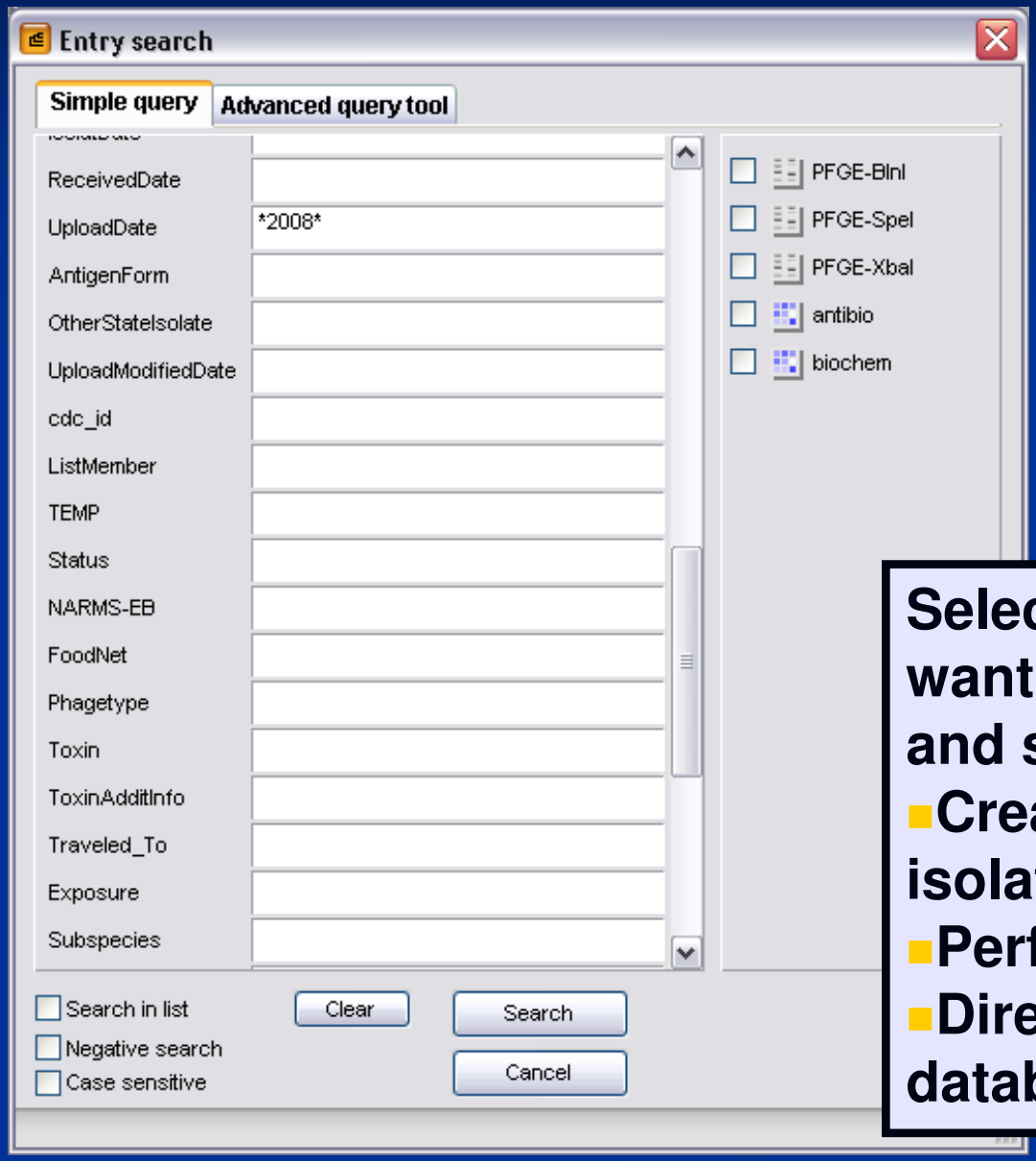
Overview

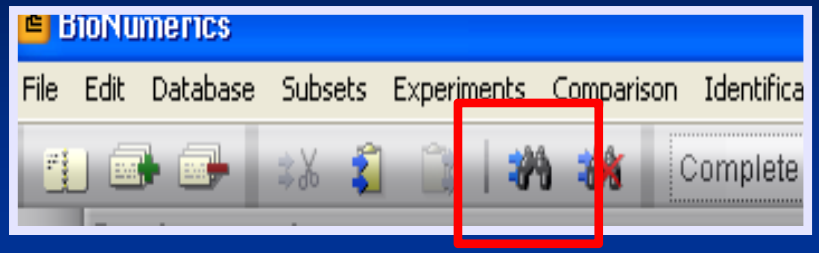
- Using the Chart and Statistics Tool
- Utility of Graphs in Cluster Detection and Reporting
- Utility of Groups

Chart and Statistics Tool: Functionality

- The utility of the Chart and Statistics Tool:
  - Generate database statistics
  - Produce different types of graphs
  - Identify discrepancies in data format

Chart & Statistics

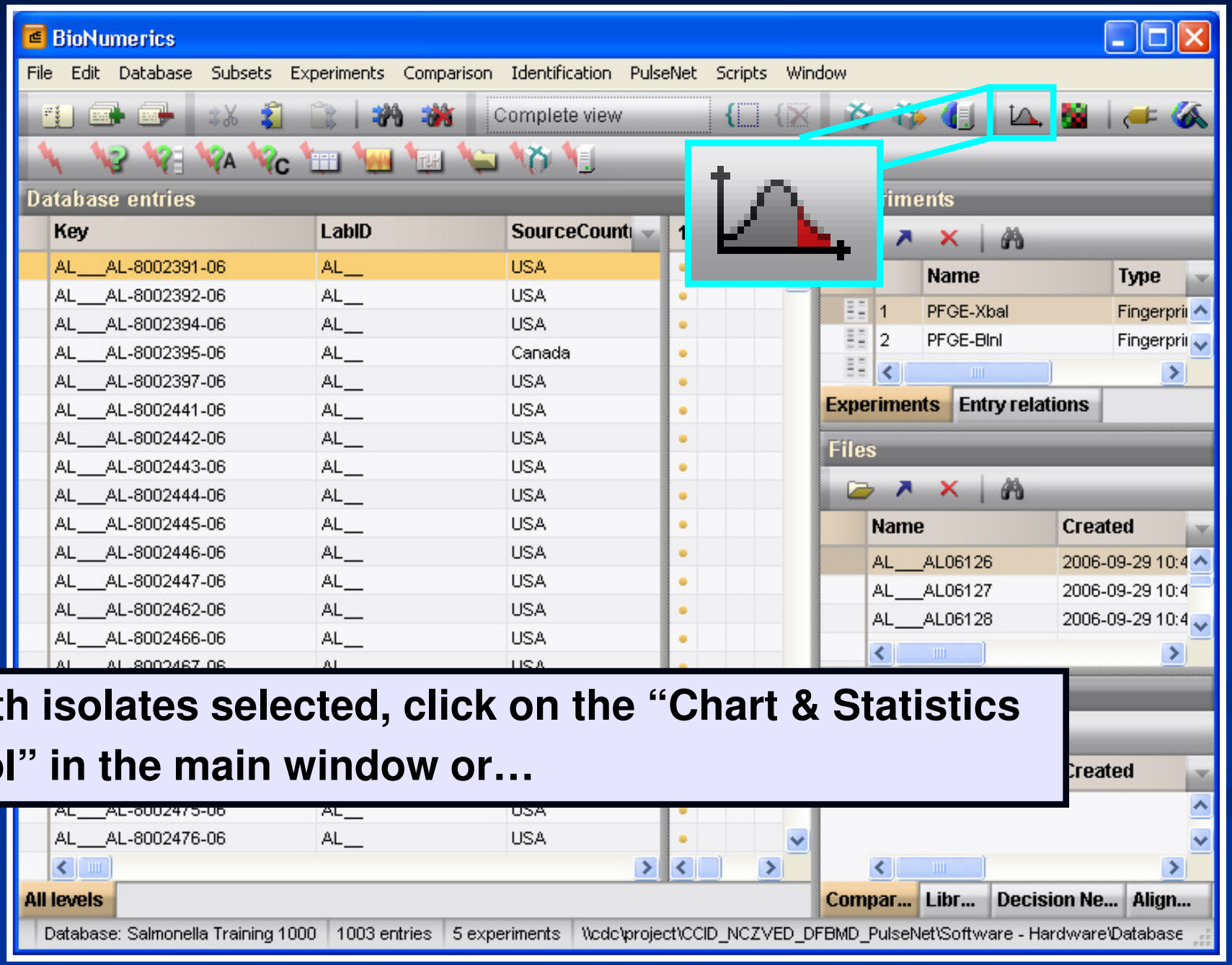




Select the group of isolates you want to analyze using the chart and statistics tool

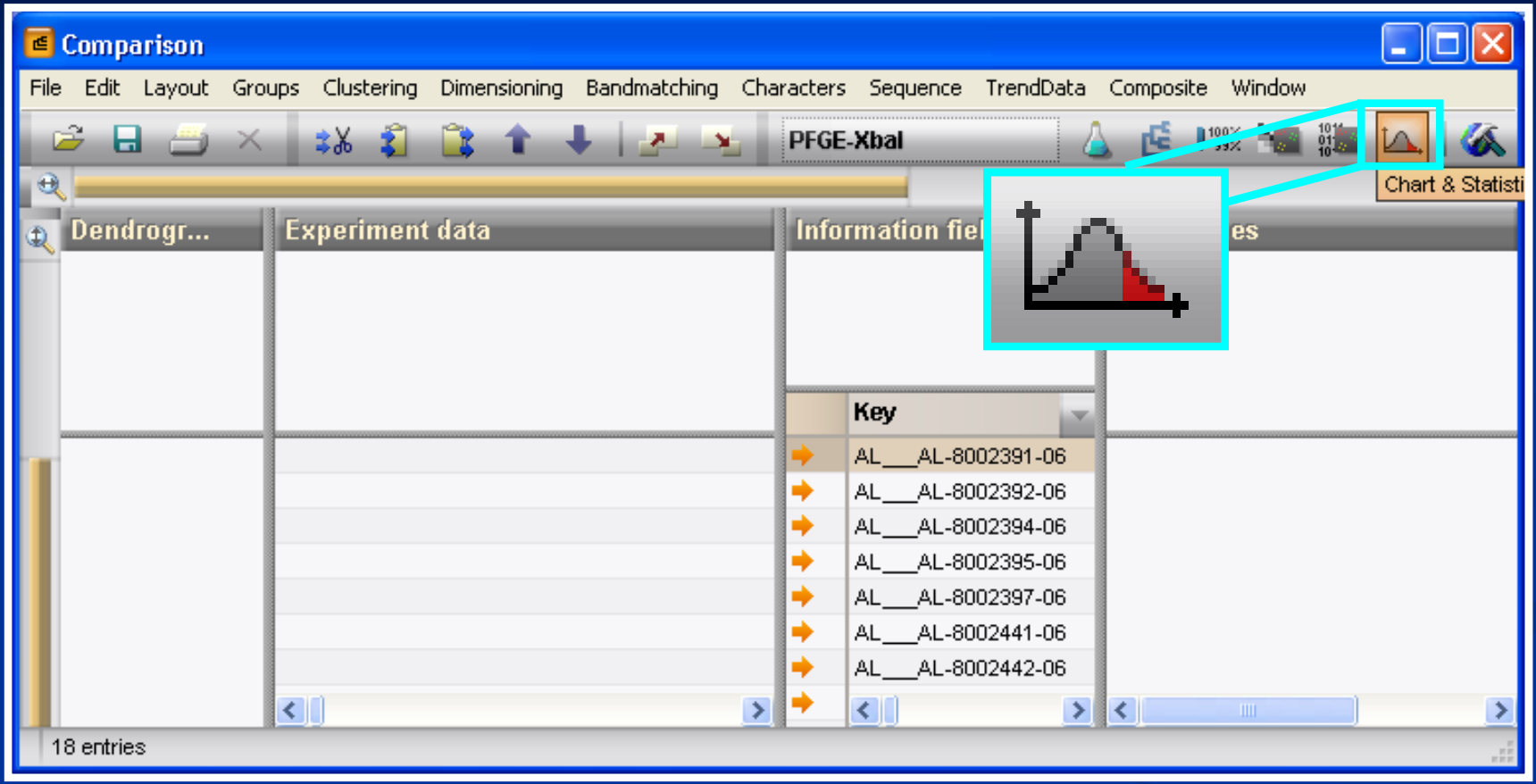
- **Create a comparison of these isolates**
- **Perform a query or**
- **Directly select isolates in the database**

Chart & Statistics



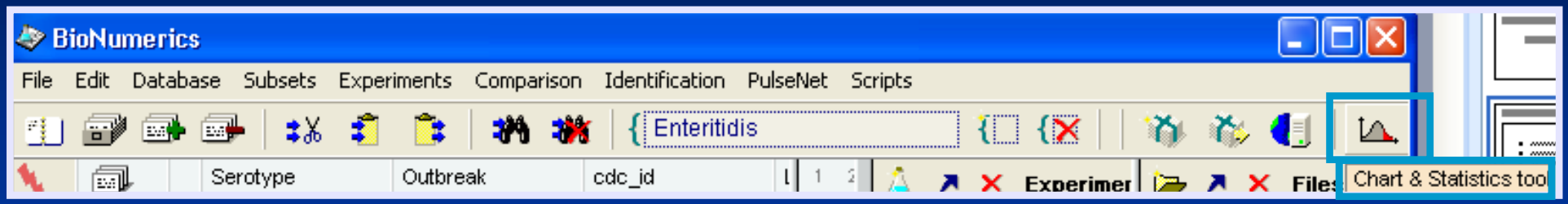
With isolates selected, click on the "Chart & Statistics tool" in the main window or…

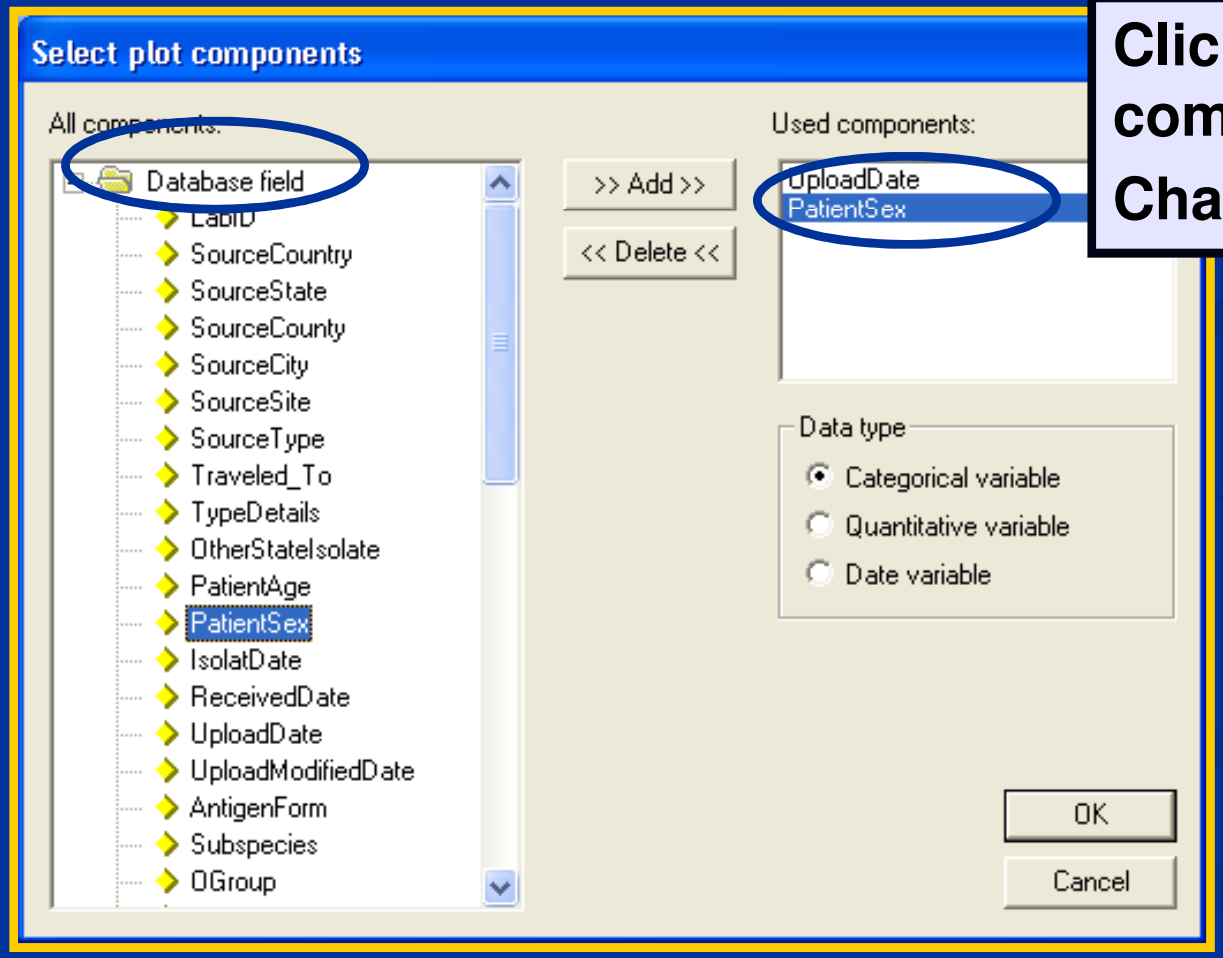
Chart & Statistics



…from a comparison window

Chart & Statistics Plot Components

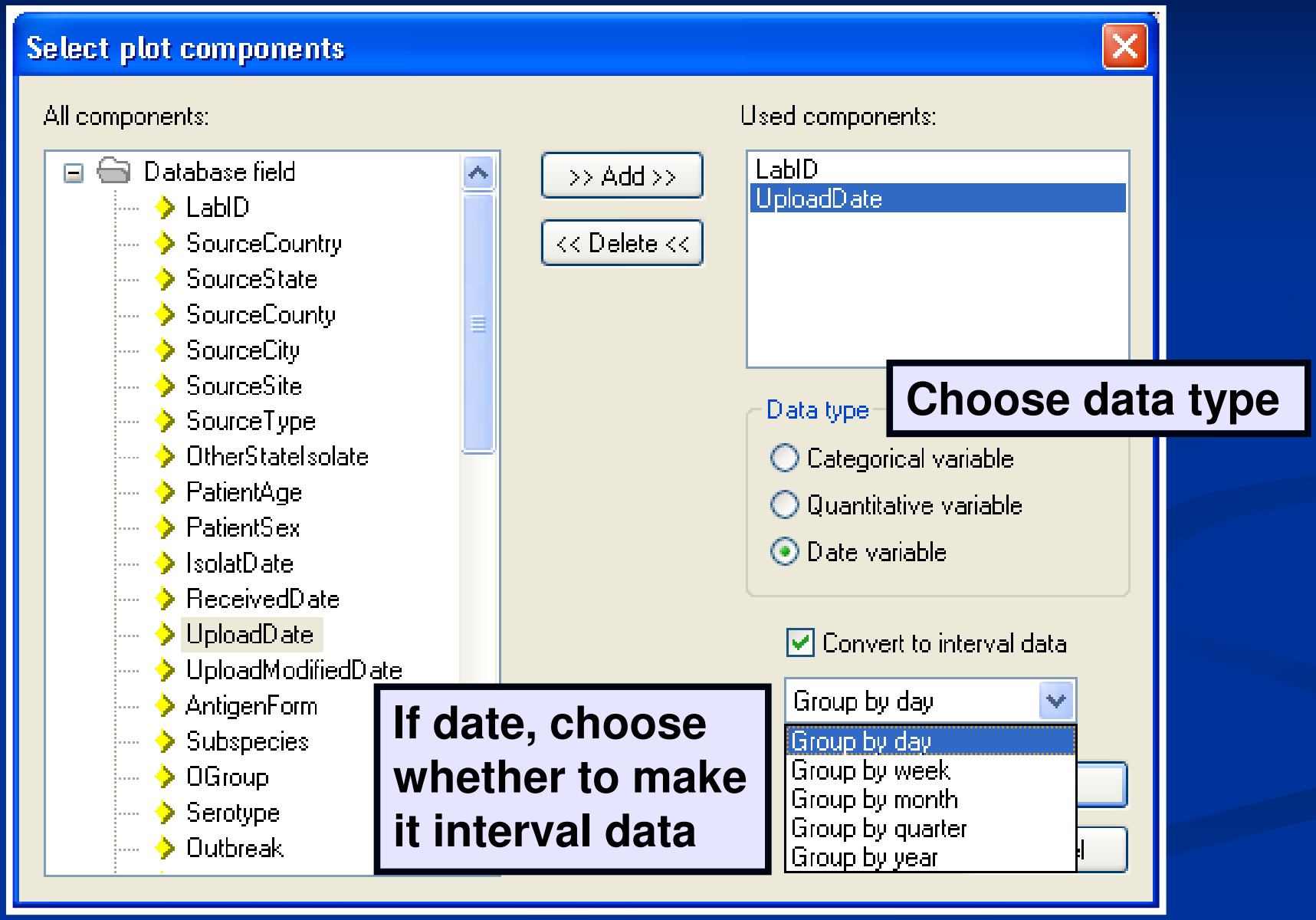




Click the tool in the main or comparison window to use Chart and Statistics tool

Choose the database components (fields) that you want to graph

Chart & Statistics Plot Components



Choose data type

If date, choose whether to make it interval data

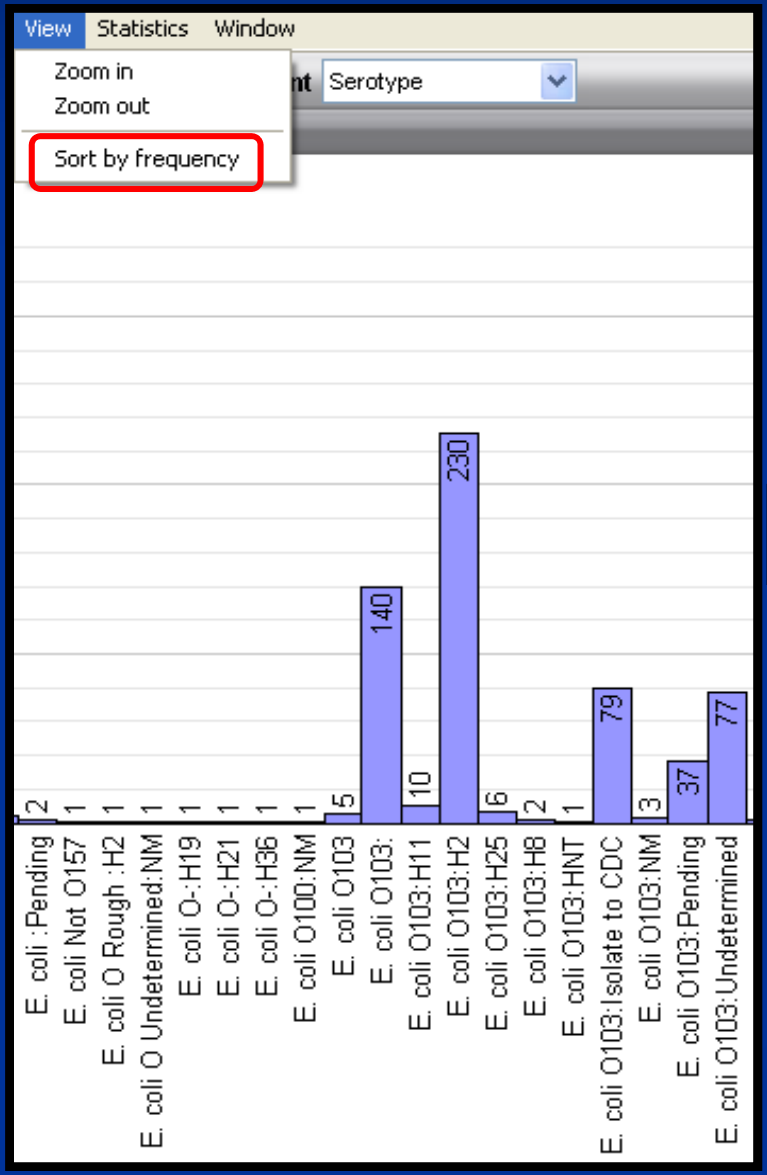
Chart & Statistics Plot Components

Data types:

- Categorical variable: descriptive variable, ex. serotype
- Quantitative variable: numerical variable, ex. age
- Date variable: can be converted into interval data (categorical or quantitative)
  - group by day, week, month, quarter or year

Chart & Statistics: Types of Graph

View →Sort by frequency to arrange the bars from highest to lowest



Bar Graph

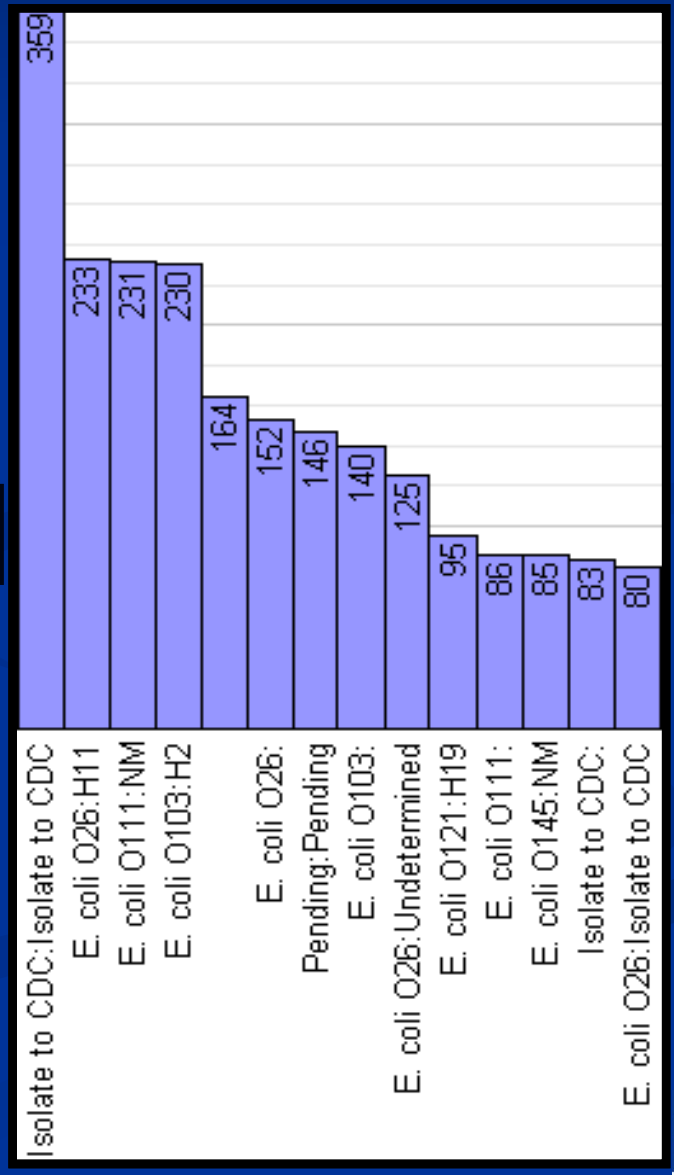


Chart & Statistics: Types of Graphs

Chart and Statistics

File Plot View Statistics Window

X component: Serotype Y component: UploadDate

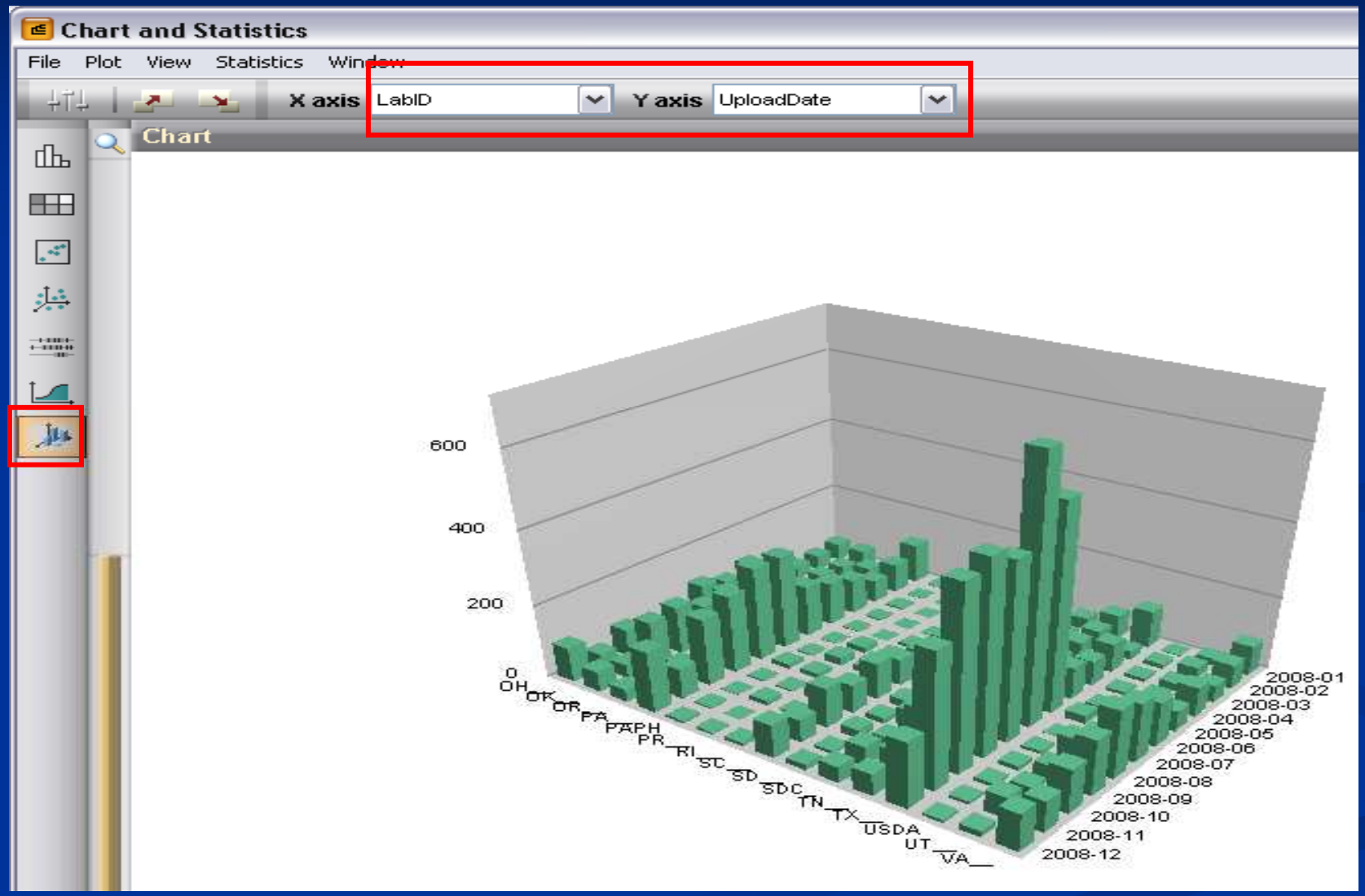
Chart

Use toolbar to change which components on which axis

	Undetermined:H11	Undetermined:H16	Undetermined:H19	Undetermined:H2	Undetermined:H21	Undetermined:H28	Undetermined:H45	Undetermined:H49	Undetermined:H6	Undetermined:H7	ined:Isolate to CDC	Undetermined:NM	Undetermined:Pending	Undetermined:Undetermined
2010-01	1	1	0	2									0	1
2010-02	0	0	0	1	0	0	0	0	0	0	0	2	0	0
2010-03	0	0	0	0	0	0	0	0	0	0	1	2	0	1
2010-04	0	0	0	2	0	1	1	0	0	0	1	0	0	2
2010-05	0	0	0	0	1	0	0	0	0	0	9	2	0	0
2010-06	0	1	1	0	0	1	0	0	0	0	4	4	0	2
2010-07	0	0	1	1	0	0	0	0	0	0	3	7	0	2
2010-08	1	0	0	0	0	0	0	0	0	0	1	0	1	0
2010-09	0	0	0	0	0	0	0	0	0	0	6	1	1	3
2010-10	0	0	0	0	0	0	0	0	0	0	2	0	0	1
2010-11	0	0	0	1	0	0	0	1	0	2	2	3	0	2
2010-12	0	0	1	3	0	0	0	0	1	0	0	2	2	0

2D Contingency Table

Chart & Statistics: Types of Graphs



3D Bar graph

Chart & Statistics Tool: Example

- Lets say you want to do some database cleaning for 2009
  - Select all 2009 isolates and pull into new comparison
  - Go to Chart and Statistics tool icon
  - Expand "Database fields" in the "All components" pane and select "Source State"
    - Repeat this for every BioNumerics field that you want to ensure correct data entry

Chart & Statistics Tool: Example

Choose database components

Click "Add"

Select plot components

All components:

- Database fields
  - SourceState
  - SourceCounty
  - PatientAge
  - PatientSex
  - IsolatDate
  - ReceivedDate
  - Serotype
  - LabID
  - SourceCountry
  - SourceCity
  - SourceSite
  - SourceType
  - Traveled_To
  - Exposure
  - TypeDetails
  - OtherStateIsolate
  - UploadDate
  - UploadModifiedDate

>> Add >>

<< Delete <<

Used components:

SourceState

Data type

- Categorical variable
- Quantitative variable
- Date variable

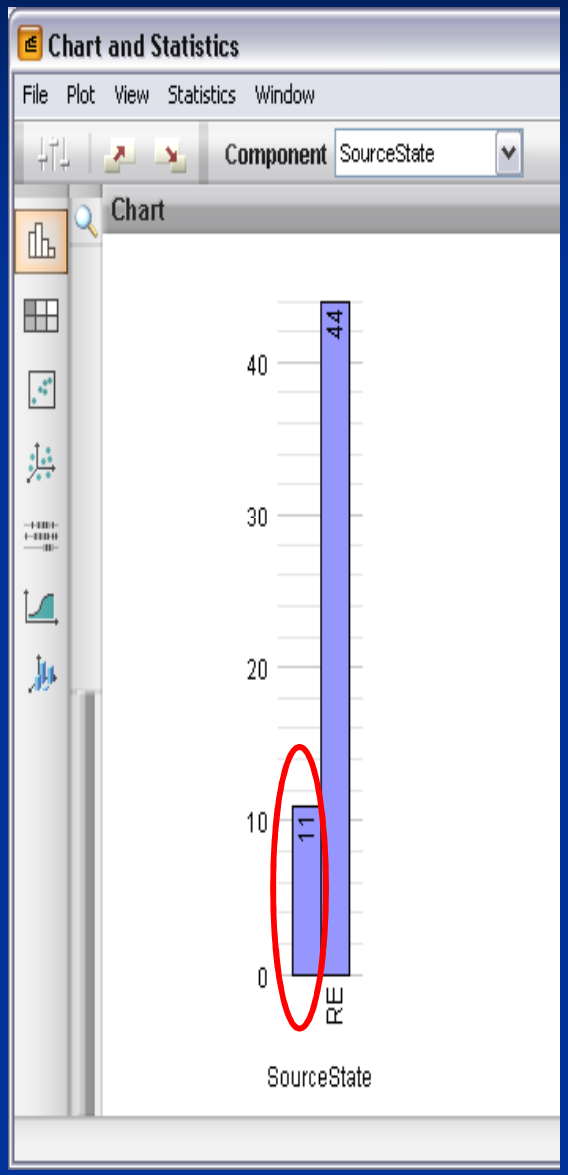
OK

Cancel

Select data type: categorical

Click OK

Chart & Statistics Tool: Example



Notice the amount of isolates without a source state

To select these isolates, go to the main BioNumerics window, deselect all isolates without closing the graph. (The bars should now be gray)

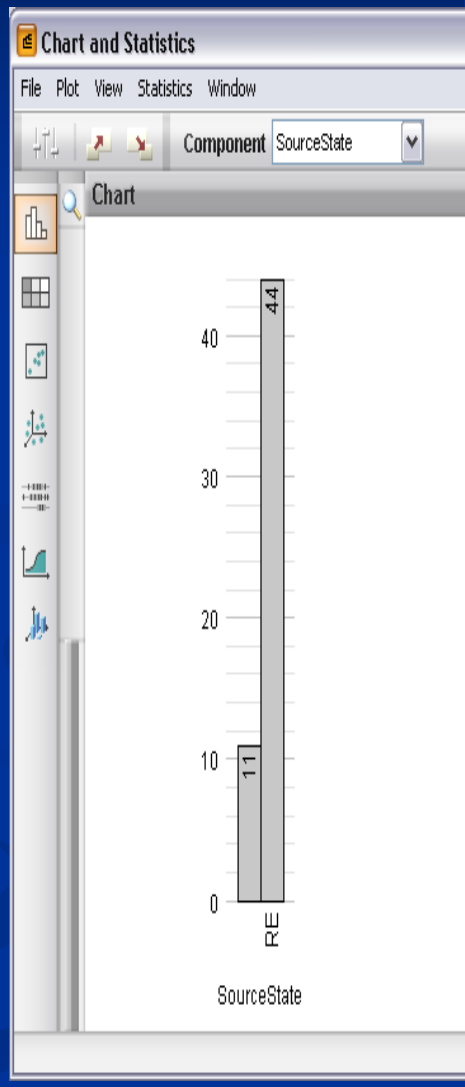
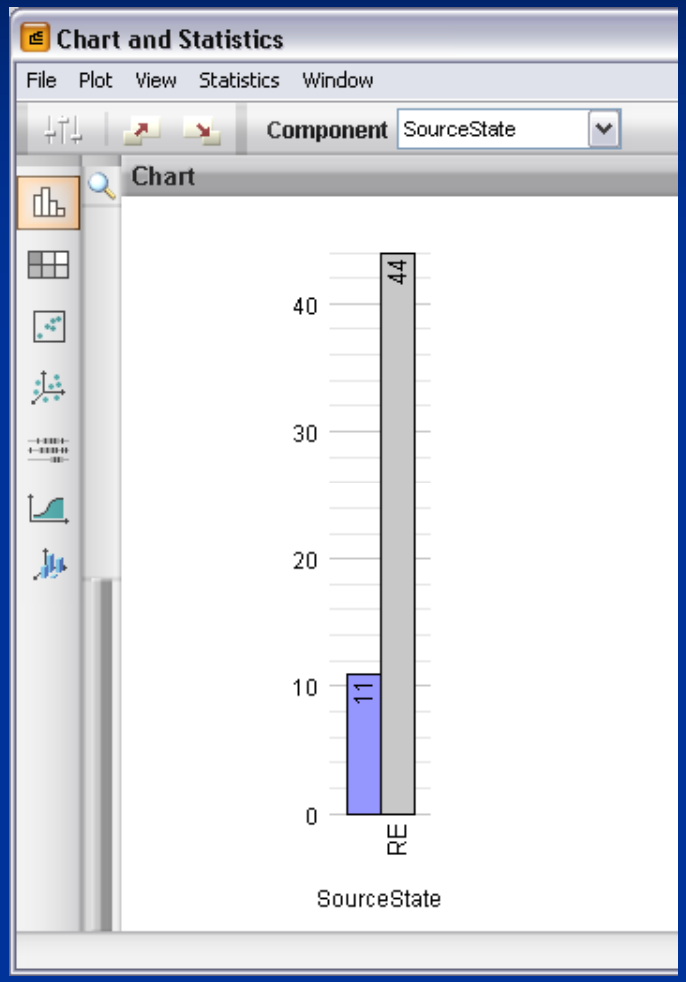


Chart & Statistics Tool: Example



Ctrl + click on the "blank state" bar to highlight isolates in the database (The bar should now purple)

Pull into a new comparison, investigate these isolates and fill in the appropriate source state if known

Correct Data Format for BioNumerics Fields

- **Patient Sex**–FEMALE, MALE, or UNKNOWN
- **Source State** –two letter postal code where the isolate sample was taken
- **Source Type** –Animal, Environment(al), Human, Food, or Unknown
- **Source Site** –Blood, Stool, Urine, CSF, etc
- **Dates** –YYYY-MM-DD

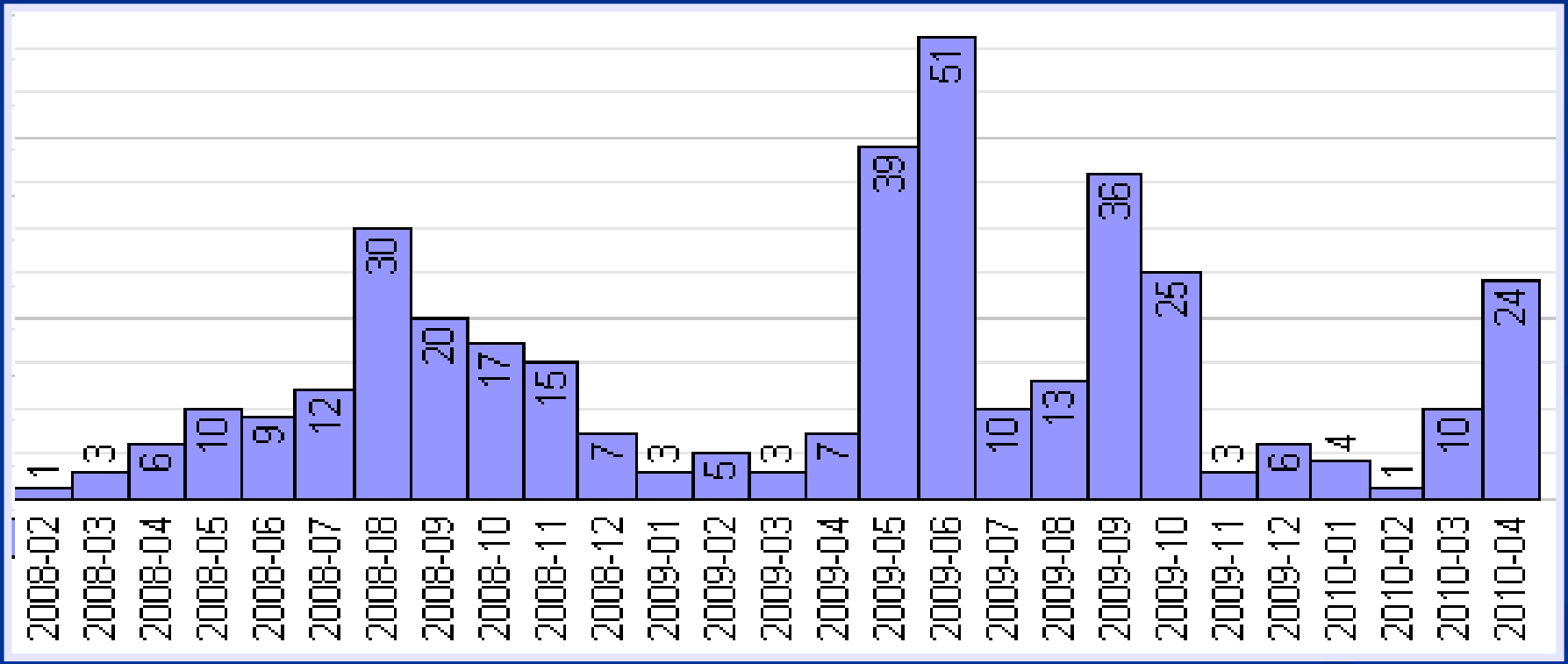
Utility of Graphs in Cluster Detection and Reporting

- When you notice 2+ isolates with the same pattern name, search for all isolates with the pattern designation
- Graph by upload date, convert to interval data, and group by month/ week
- Evaluate the past 60 days of submission to determine if there is an increase over baseline
  - If there is an increase, post onto CDC team
  - If there is an epi link, post onto CDC team
  - Follow your lab protocol

Utility of Graphs in Cluster Detection and Reporting

- You've noticed some clusters, 2+ recent isolates with the same PFGE pattern name
  - What should you do?
  - Step 1: go to your main screen and search for all isolates assigned the pattern name of interest
  - Step 2: select chart and statistics and create a bar graph by upload data and select "date variable" by week or month

Utility of Graphs in Cluster Detection and Reporting



How can I represent just the past 60 days in this graph?

Utility of Graphs in Cluster Detection and Reporting

BioNumerics

File Edit Database Subsets Experiments Comparison Identification Pulse

Complete view

Unselect all entries (F4)

Database entries

Index			Serotype	PFGE-XbaI-patt...	P
7			Enteritidis	JEGX01.0004	T
8			Enteritidis	JEGX01.0004	T
9			Enteritidis	JEGX01.0004	T
10			Enteritidis	JEGX01.0004	T

Deselect all entries *without* closing your graph

BioNumerics

File Edit Database Subsets Ex

Database entries

Index	Key
7	TN___043793
8	
9	
10	
11	
12	
13	
14	TN___045394
15	TN___050256
16	TN___050354
17	TN___050422
18	TN___050938
19	TN___050982
20	TN___051256

Query recent uploads

Perform a Hot List search of the past 60 days

Enter number of days before now

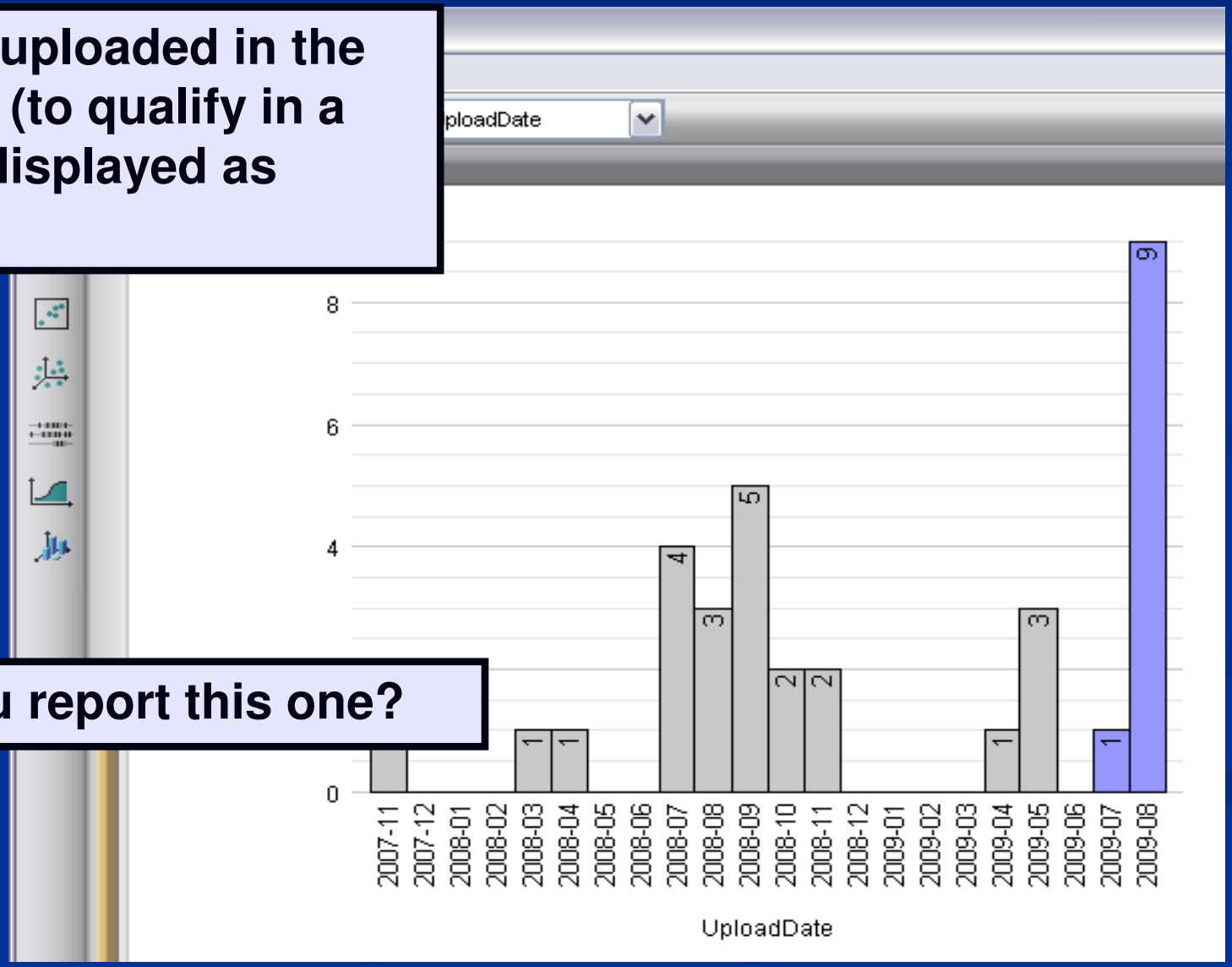
60

OK

Cancel

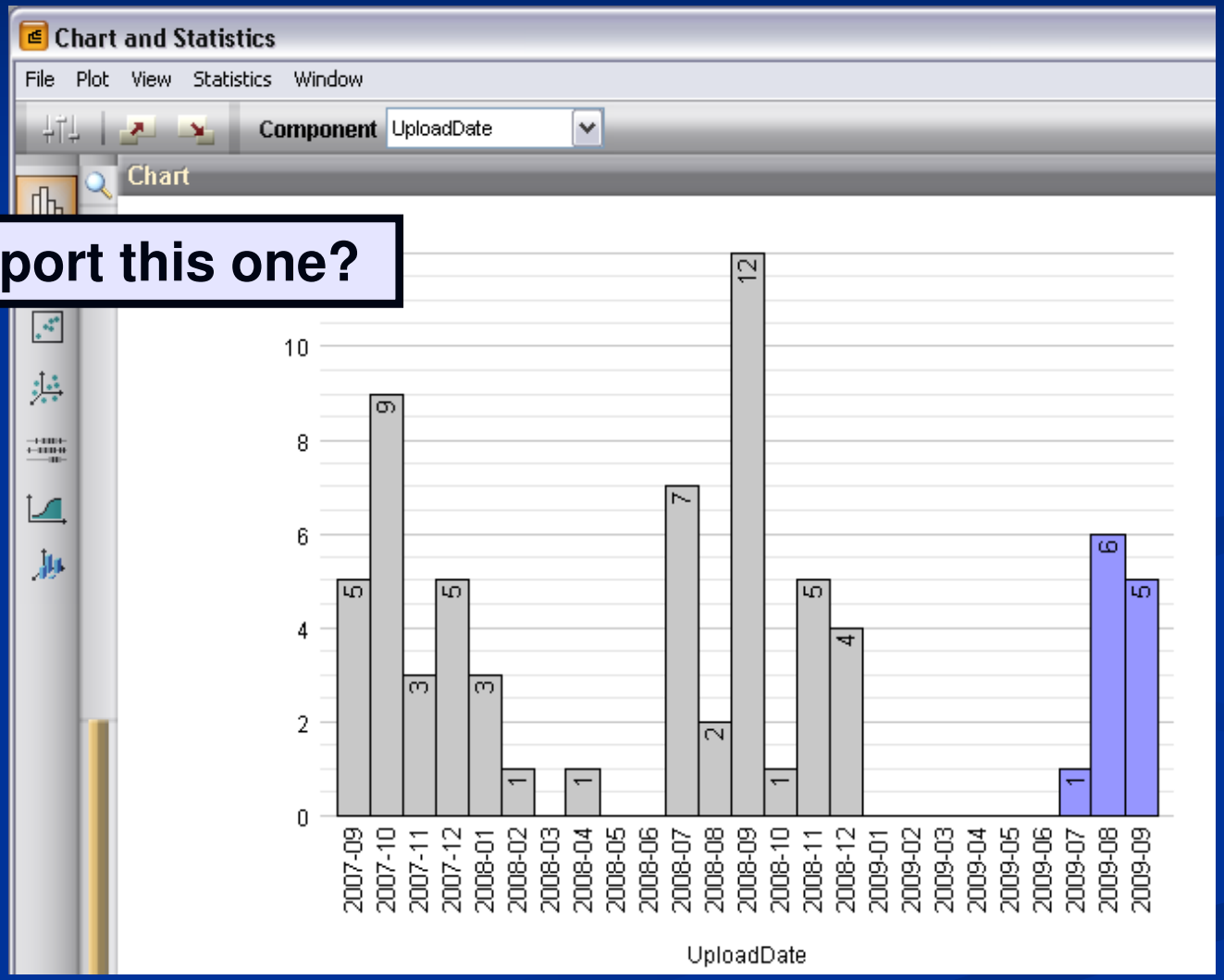
Utility of Graphs in Cluster Detection and Reporting: Example 1

The isolates uploaded in the past 60 days (to qualify in a cluster) are displayed as purple bars.



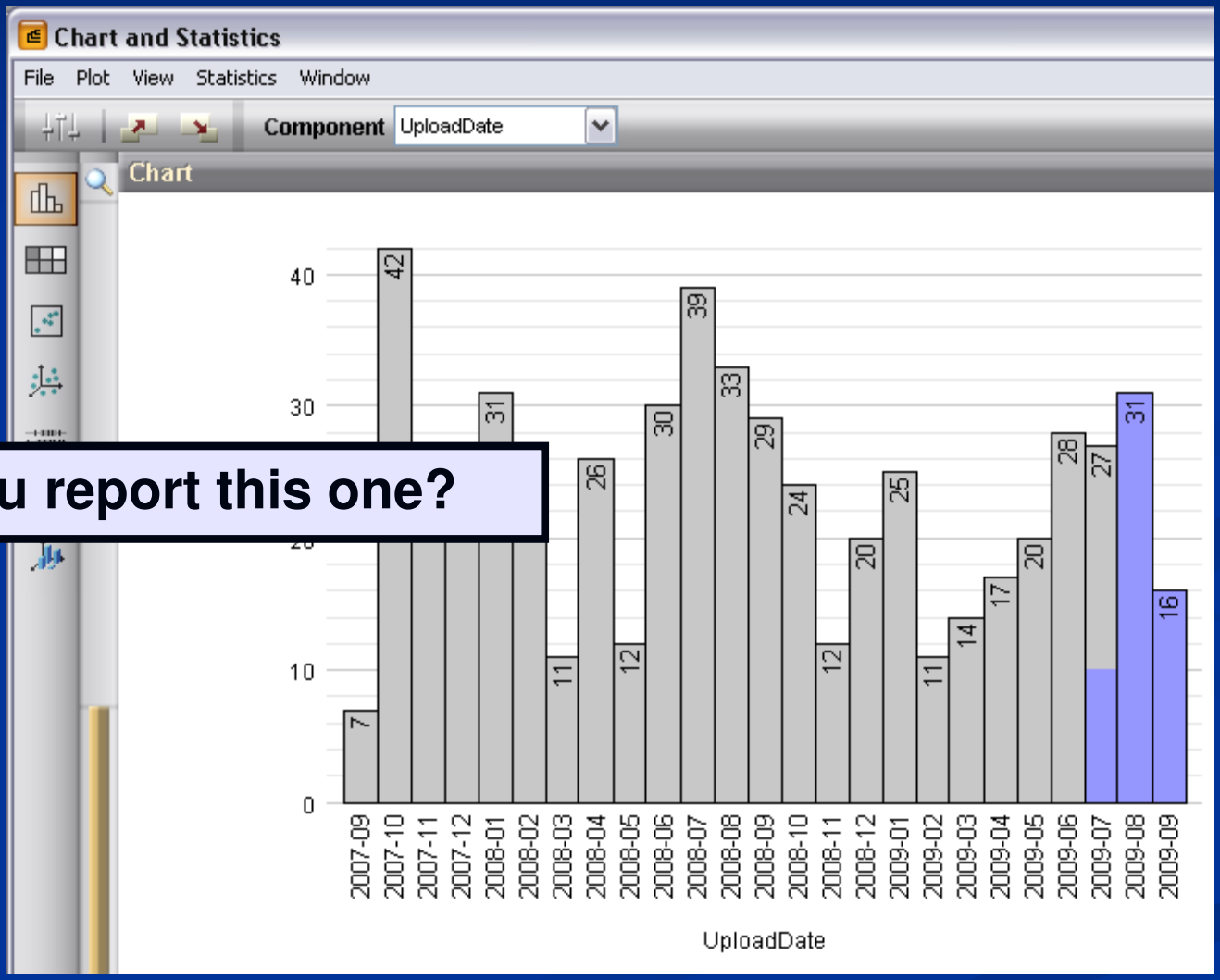
Should you report this one?

Utility of Graphs in Cluster Detection and Reporting: Example 2



Should you report this one?

Utility of Graphs in Cluster Detection and Reporting: Example 3

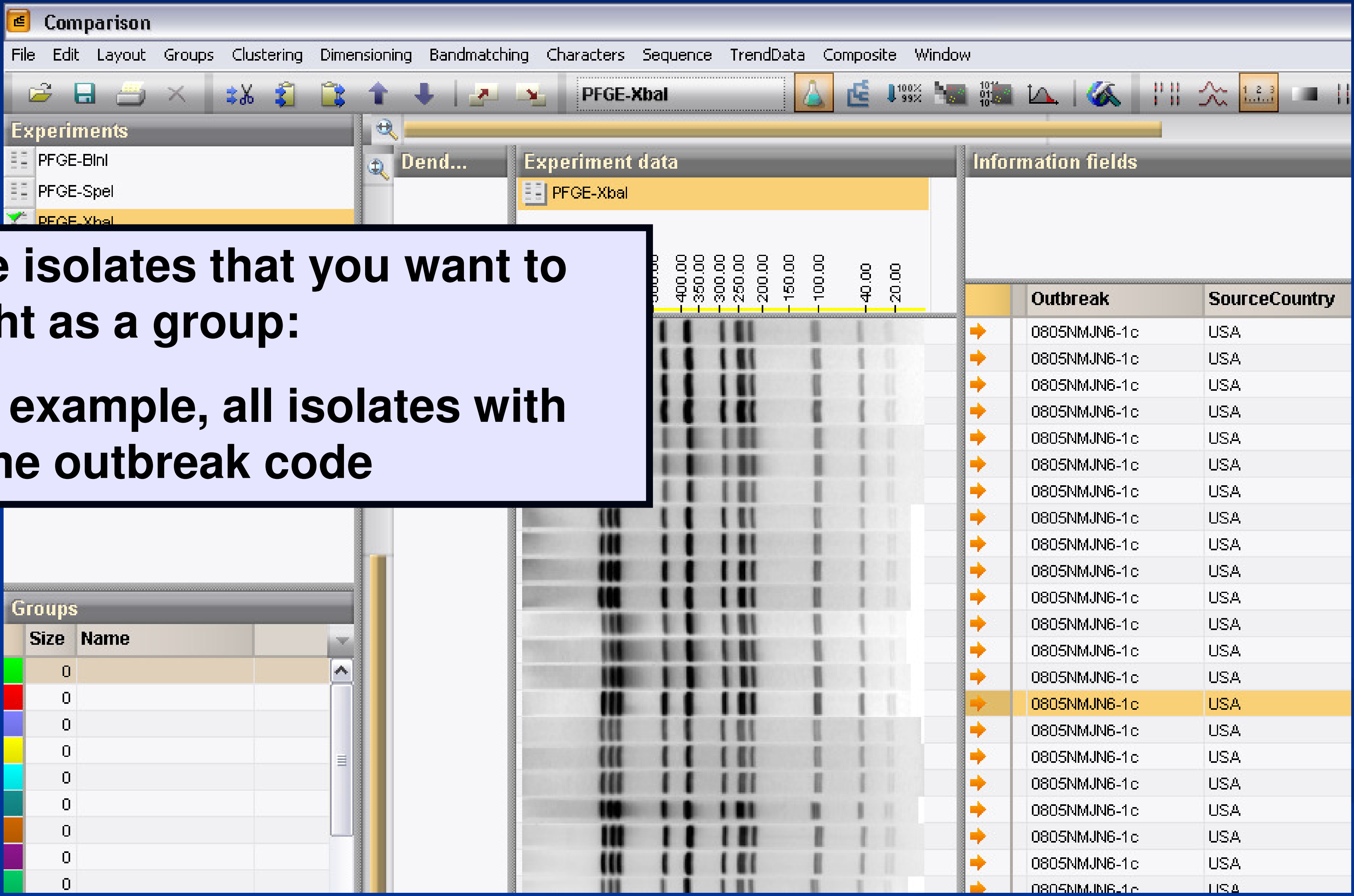


Should you report this one?

Utility of Groups

- The use of groups/colors can distinguish isolates to another level while working in a comparison
  - Example: Let's say you want to evaluate the number of 2009 isolates associated with each CDC outbreak code
    - In order for this to work in your local database, you need to regularly download pattern names and outbreak codes

Utility of Groups



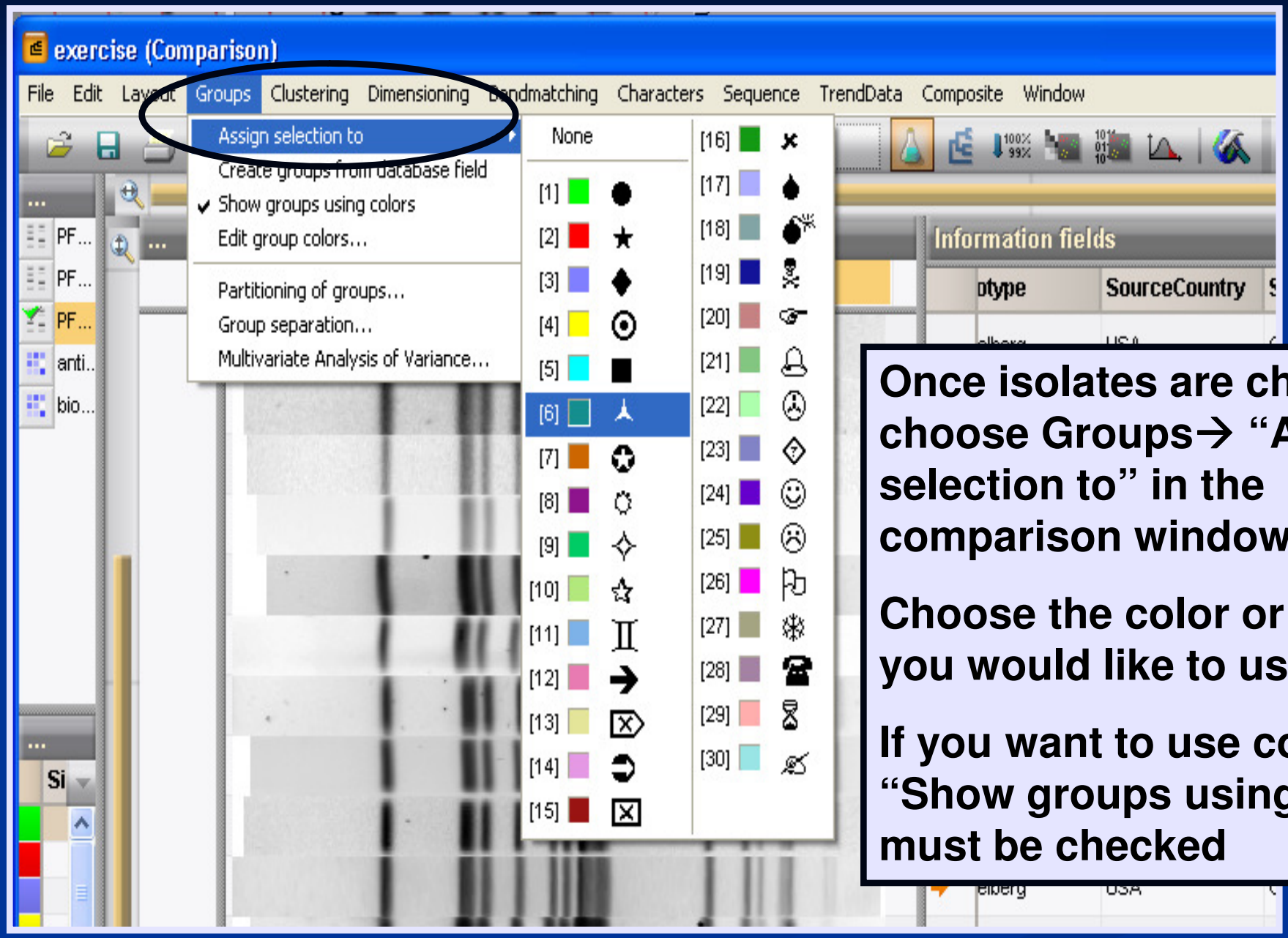
	Size	Name
🟩	0	
🟥	0	
🟦	0	
🟨	0	
🟦	0	
🟦	0	
🟧	0	
🟪	0	
🟩	0	

Outbreak	SourceCountry
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA
0805NMJN6-1c	USA

Choose isolates that you want to highlight as a group:

For the example, all isolates with the same outbreak code

Using Groups

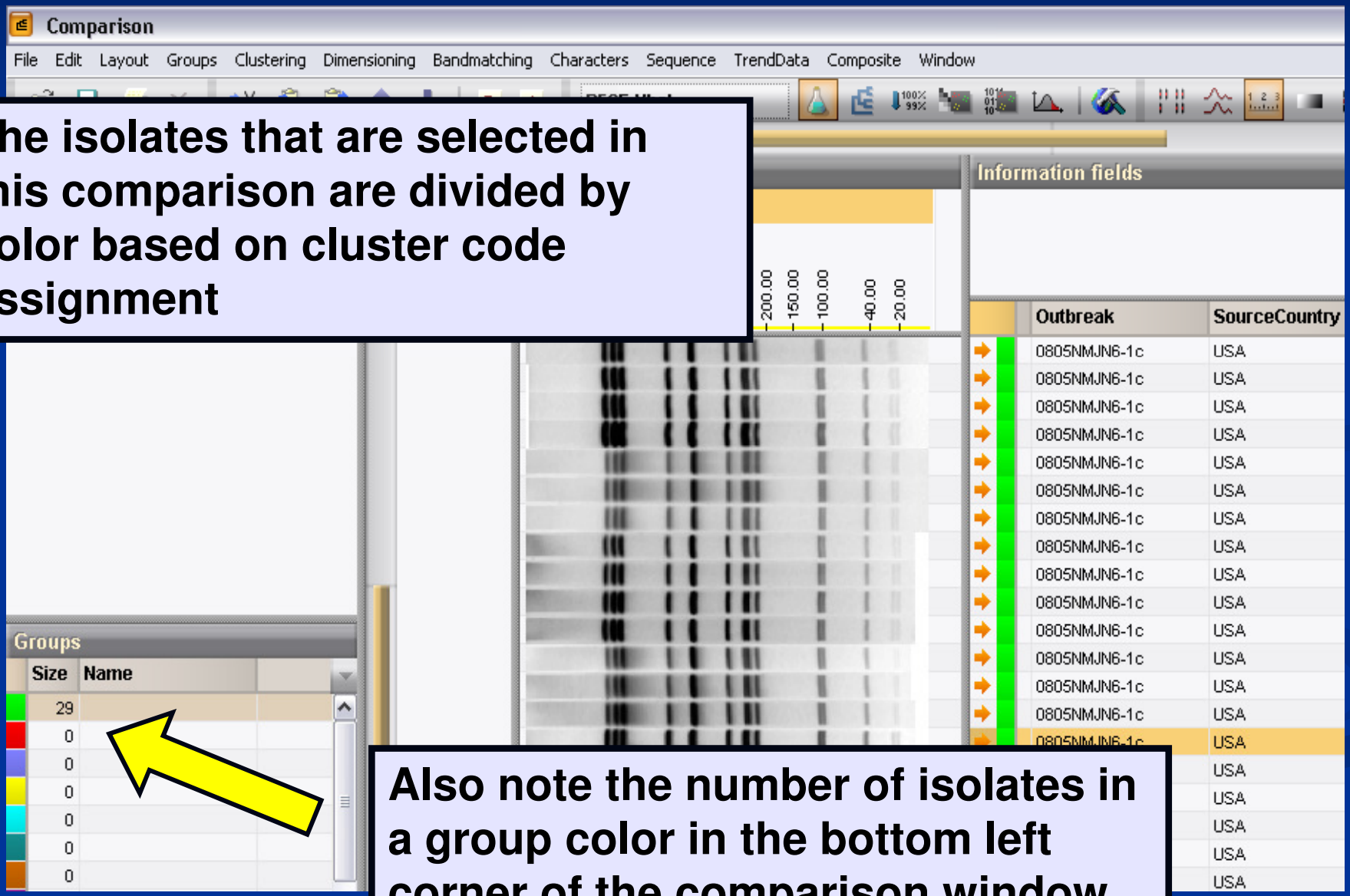


Once isolates are chosen, choose Groups→ "Assign selection to" in the comparison window

Choose the color or symbol you would like to use

If you want to use colors, "Show groups using colors" must be checked

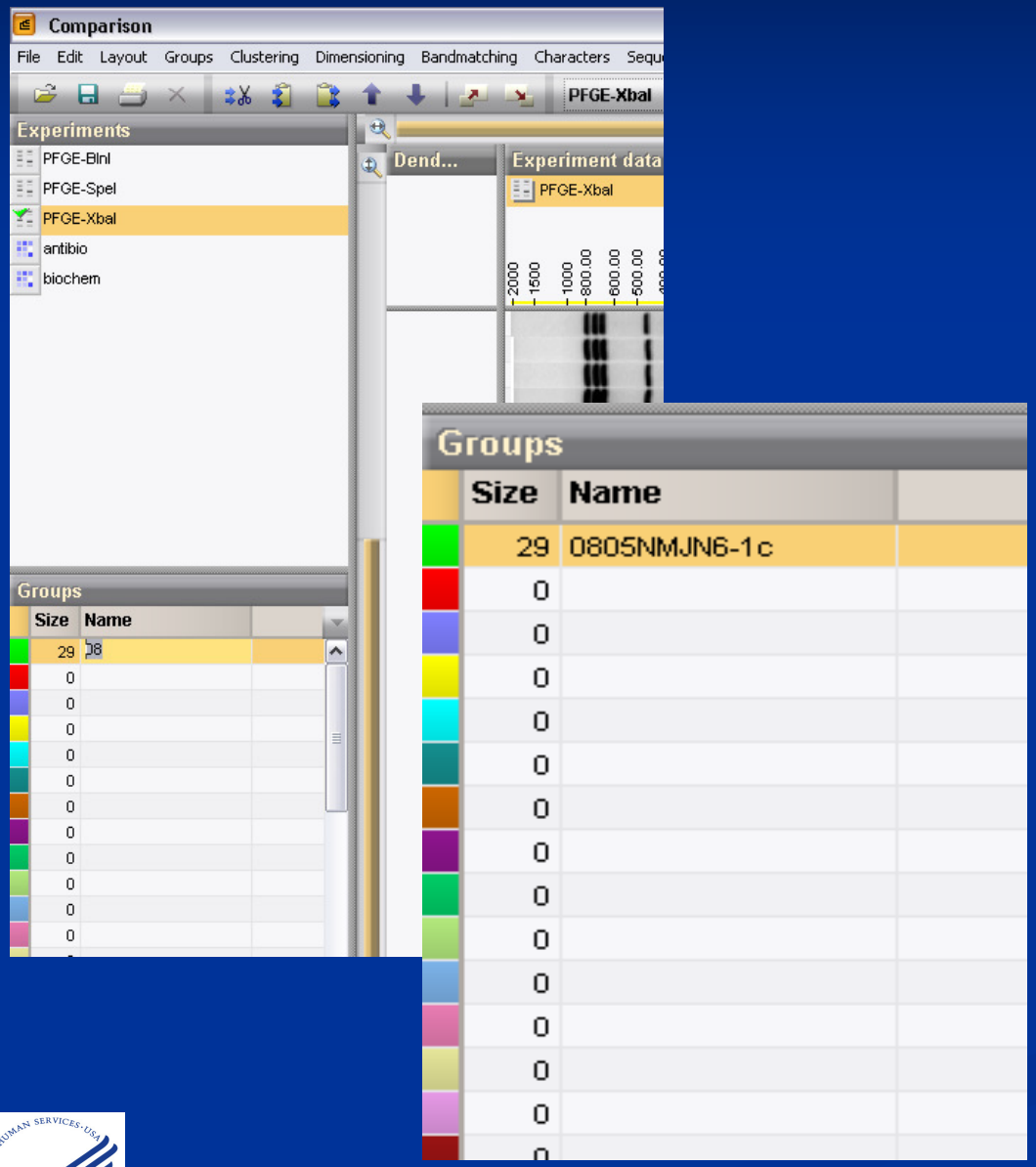
Utility of Groups



The isolates that are selected in this comparison are divided by color based on cluster code assignment

Also note the number of isolates in a group color in the bottom left corner of the comparison window

Utility of Groups



You can name each color group by highlighting the name field and click, then type in the appropriate name for each color designation

Utility of Groups

BioNumerics

File Edit Database Subsets Experiments Comparison Identification PulseNet Scripts Help Window

Complete view

Database entries

Comparison

File Edit Layout Groups Clustering Dimensioning Bandmatching Characters Sequence TrendData Composite

Assign selection to
Create groups from database field
Show groups using colors
Edit group colors...
Partitioning of groups...
Group separation...
Multivariate Analysis of Variance...

PFGE-XbaI

Experiments
PFGE-BlnI
PFGE-SpeI
PFGE-XbaI
antibio
biochem

Experiment data
PFGE-XbaI

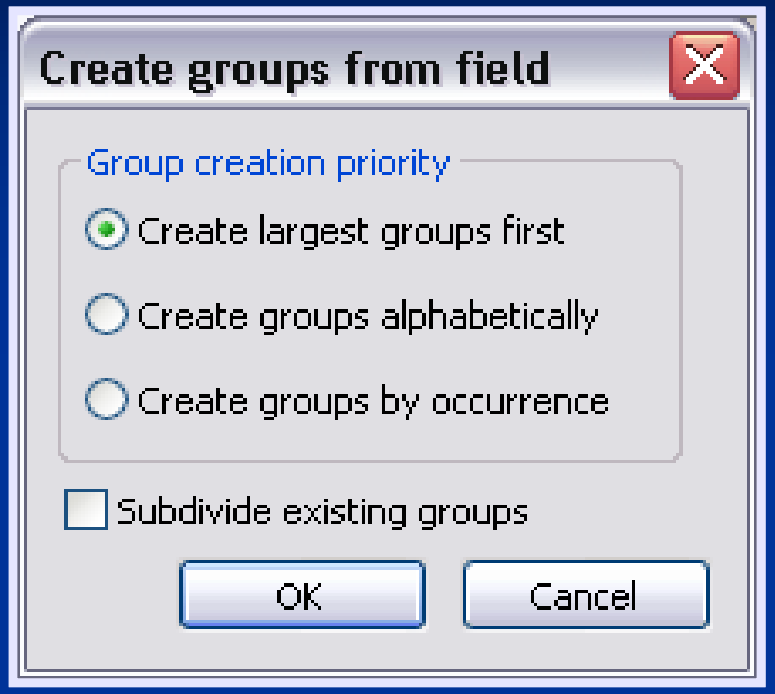
Information fields

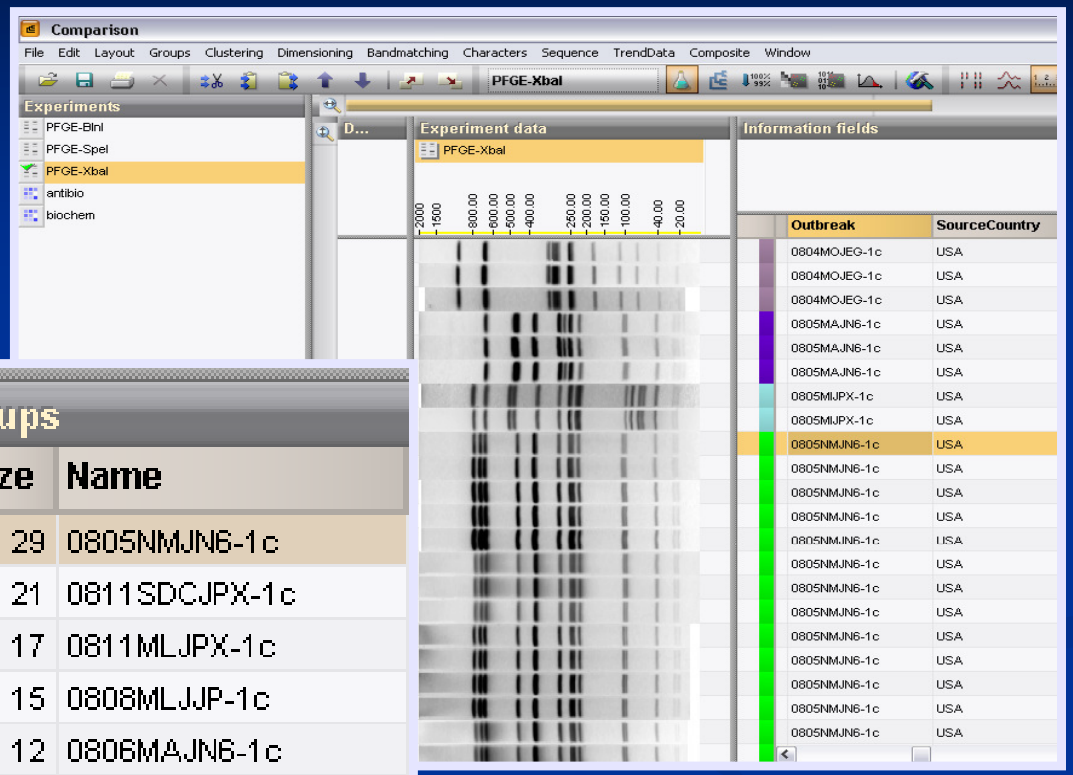
date	PFGE-SpeI-status	Outbreak
	Unconfirmed	0708NYCJM6-1c
	Unconfirmed	0801ORJPX-1c
	Unconfirmed	0802MLJPX-1c
	Unconfirmed	0803VAJPX-1c

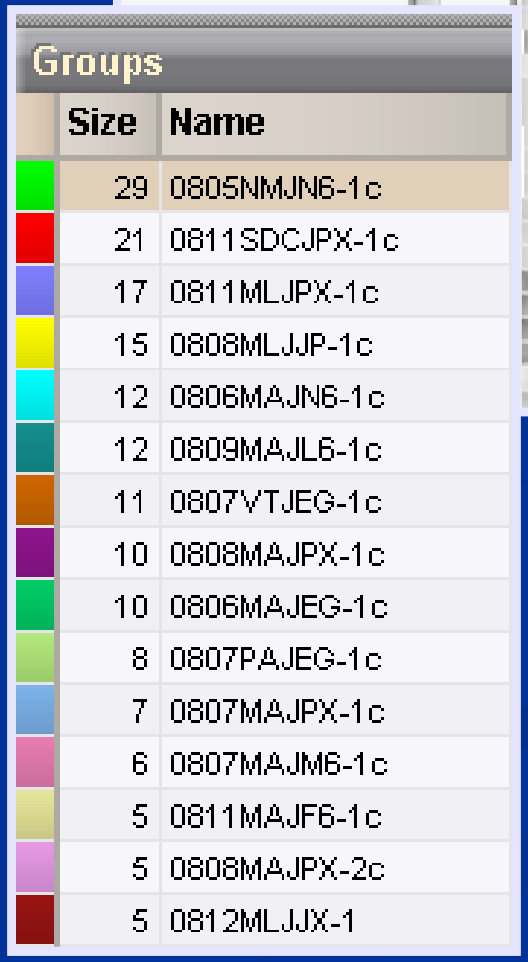
To automatically assign groups, select the database field of interest

Go to Groups and Select Create groups from database field

Utility of Groups







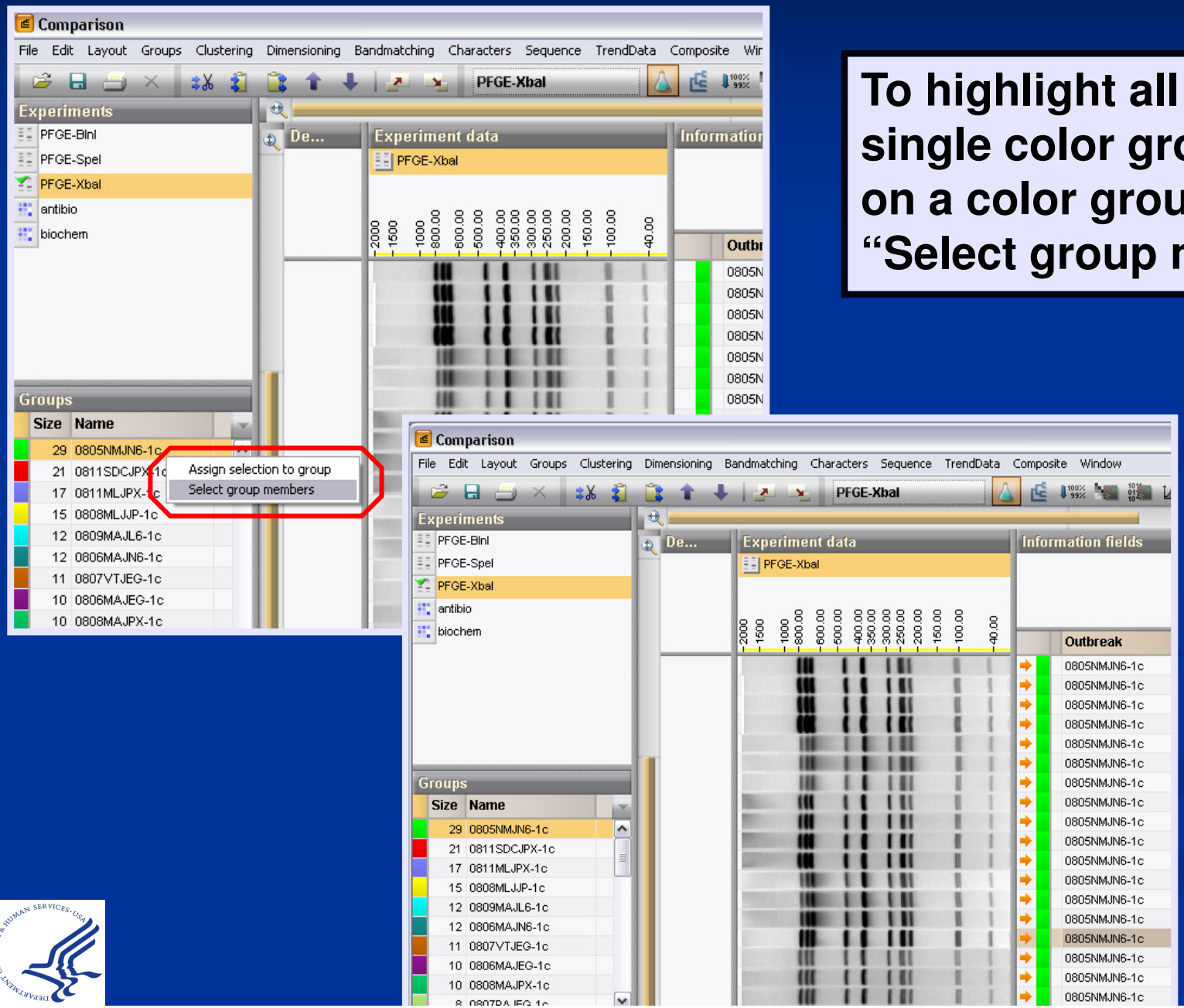
Groups

Size	Name
29	0805NMJN6-1c
21	0811SDCJPX-1c
17	0811MLJPX-1c
15	0808MLJJP-1c
12	0806MAJN6-1c
12	0809MAJL6-1c
11	0807VTJEG-1c
10	0808MAJPX-1c
10	0806MAJEG-1c
8	0807PAJEG-1c
7	0807MAJPX-1c
6	0807MAJM6-1c
5	0811MAJF6-1c
5	0808MAJPX-2c
5	0812MLJJX-1

Choose the method of creating the groups and click OK

Now you can easily determine the amount of isolates per outbreak code

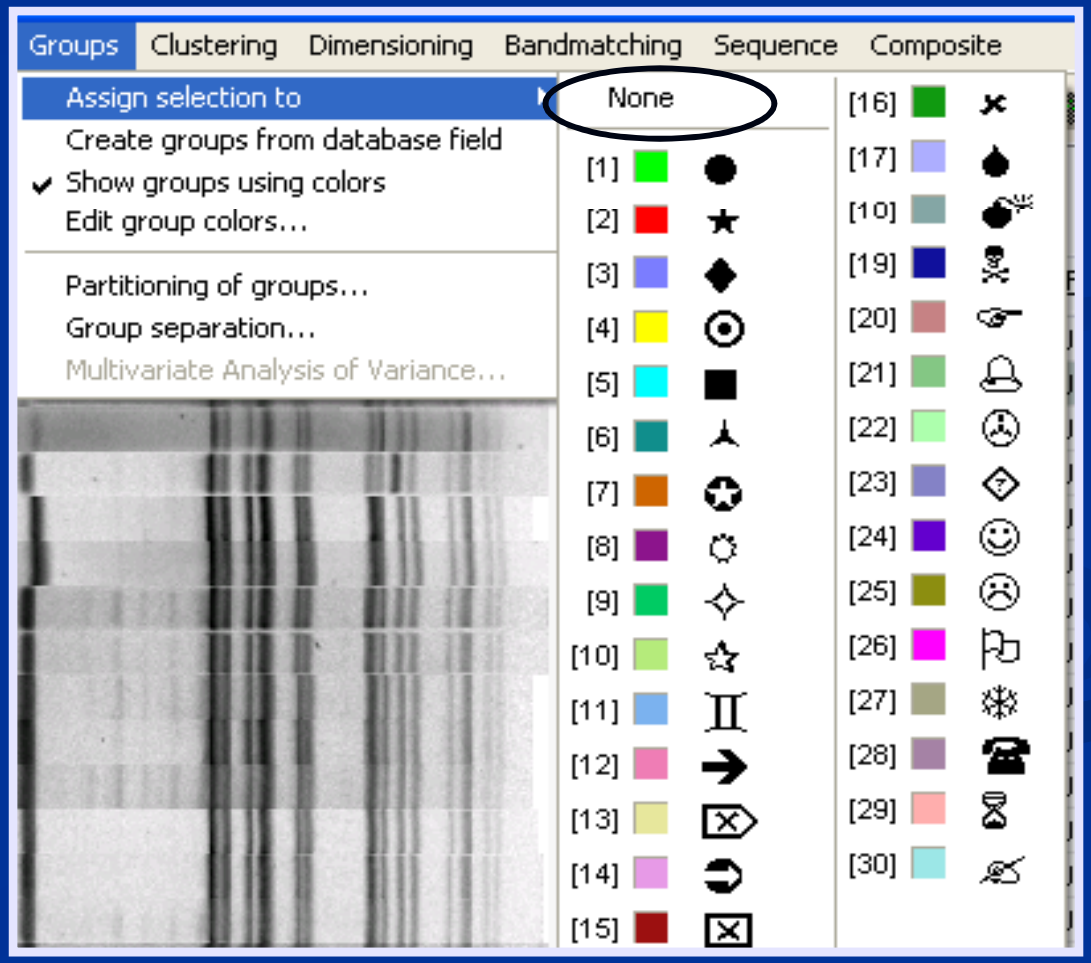
Utility of Groups



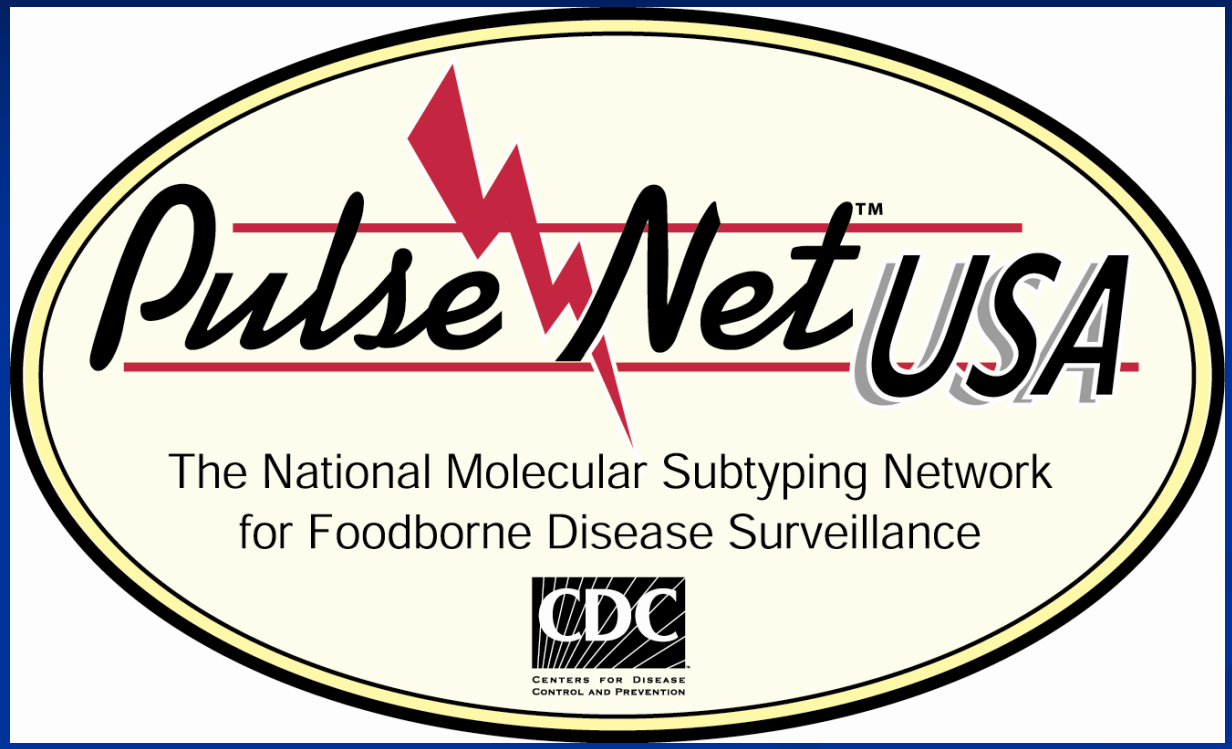
To highlight all members of a single color group, right click on a color group and choose "Select group members"

Utility of Groups

- To remove a group selection, select all the entries within the group and within the group window 'assign selection to' none.



Questions?



Thank you for your attention

The findings and conclusions in this presentation are those of the author and do not necessarily represent the views of the Centers for Disease Control and Prevention