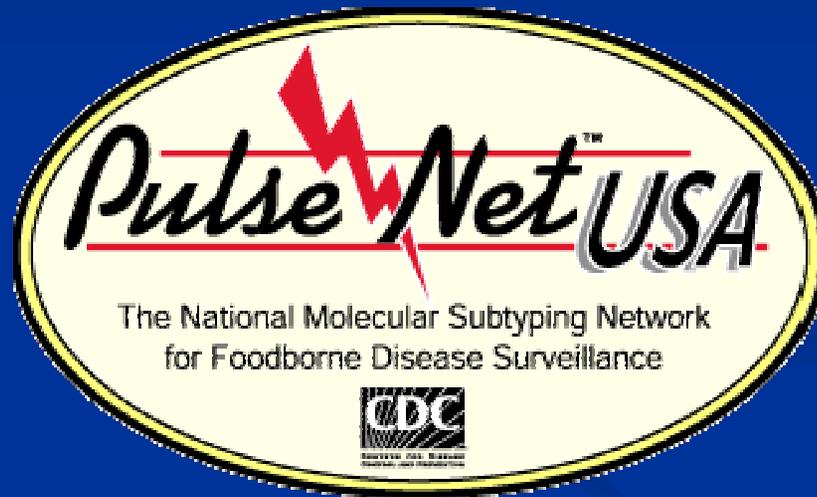


Using the Chart & Statistics Tool and Groups



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

Entry search

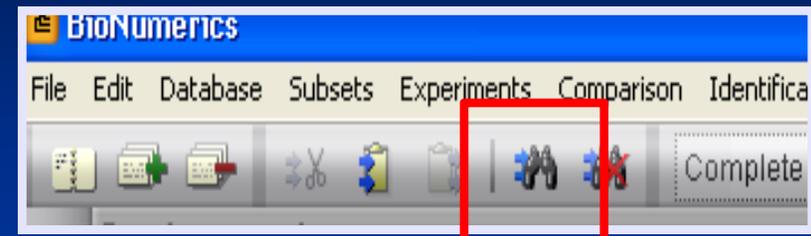
Simple query | **Advanced query tool**

ReceivedDate	
UploadDate	*2008*
AntigenForm	
OtherStateIsolate	
UploadModifiedDate	
cdc_id	
ListMember	
TEMP	
Status	
NARMS-EB	
FoodNet	
Phagetype	
Toxin	
ToxinAdditInfo	
Traveled_To	
Exposure	
Subspecies	

PFGE-BlnI
 PFGE-Spel
 PFGE-Xbal
 antibio
 biochem

Search in list
 Negative search
 Case sensitive

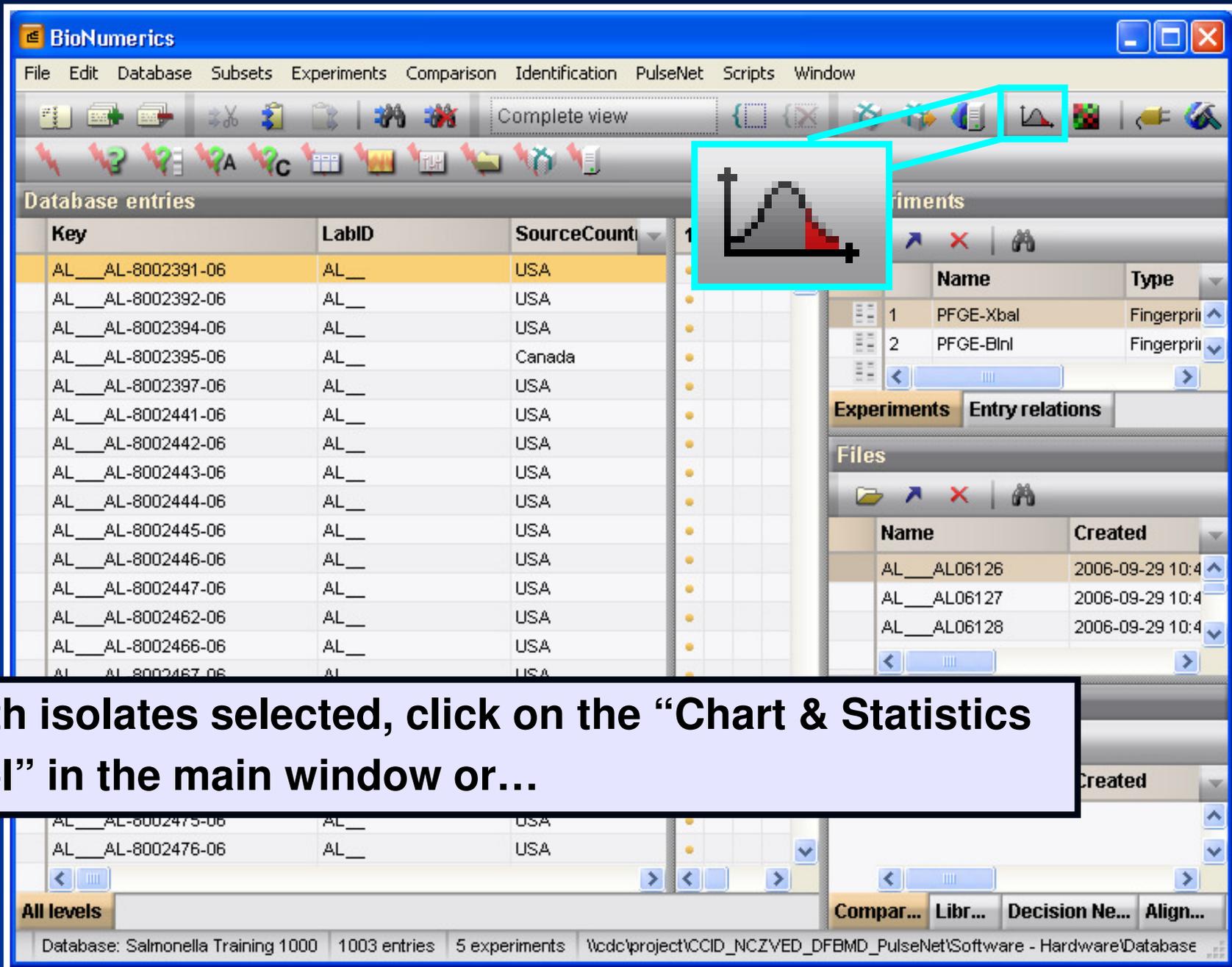
Clear Search
Cancel



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



The screenshot displays the BioNumerics software interface. The main window shows a table of database entries with columns for Key, LabID, and SourceCount. The 'SourceCount' column is expanded to show the source of each isolate. A red box highlights the 'Chart & Statistics' icon in the toolbar. A callout window shows a histogram of the data, with the x-axis representing the source and the y-axis representing the count. The histogram shows a distribution of isolates across different sources, with a peak for 'USA' and a smaller peak for 'Canada'.

Key	LabID	SourceCount
AL__AL-8002391-06	AL__	USA
AL__AL-8002392-06	AL__	USA
AL__AL-8002394-06	AL__	USA
AL__AL-8002395-06	AL__	Canada
AL__AL-8002397-06	AL__	USA
AL__AL-8002441-06	AL__	USA
AL__AL-8002442-06	AL__	USA
AL__AL-8002443-06	AL__	USA
AL__AL-8002444-06	AL__	USA
AL__AL-8002445-06	AL__	USA
AL__AL-8002446-06	AL__	USA
AL__AL-8002447-06	AL__	USA
AL__AL-8002462-06	AL__	USA
AL__AL-8002466-06	AL__	USA
AL__AL-8002467-06	AL__	USA

With isolates selected, click on the “Chart & Statistics tool” in the main window or...

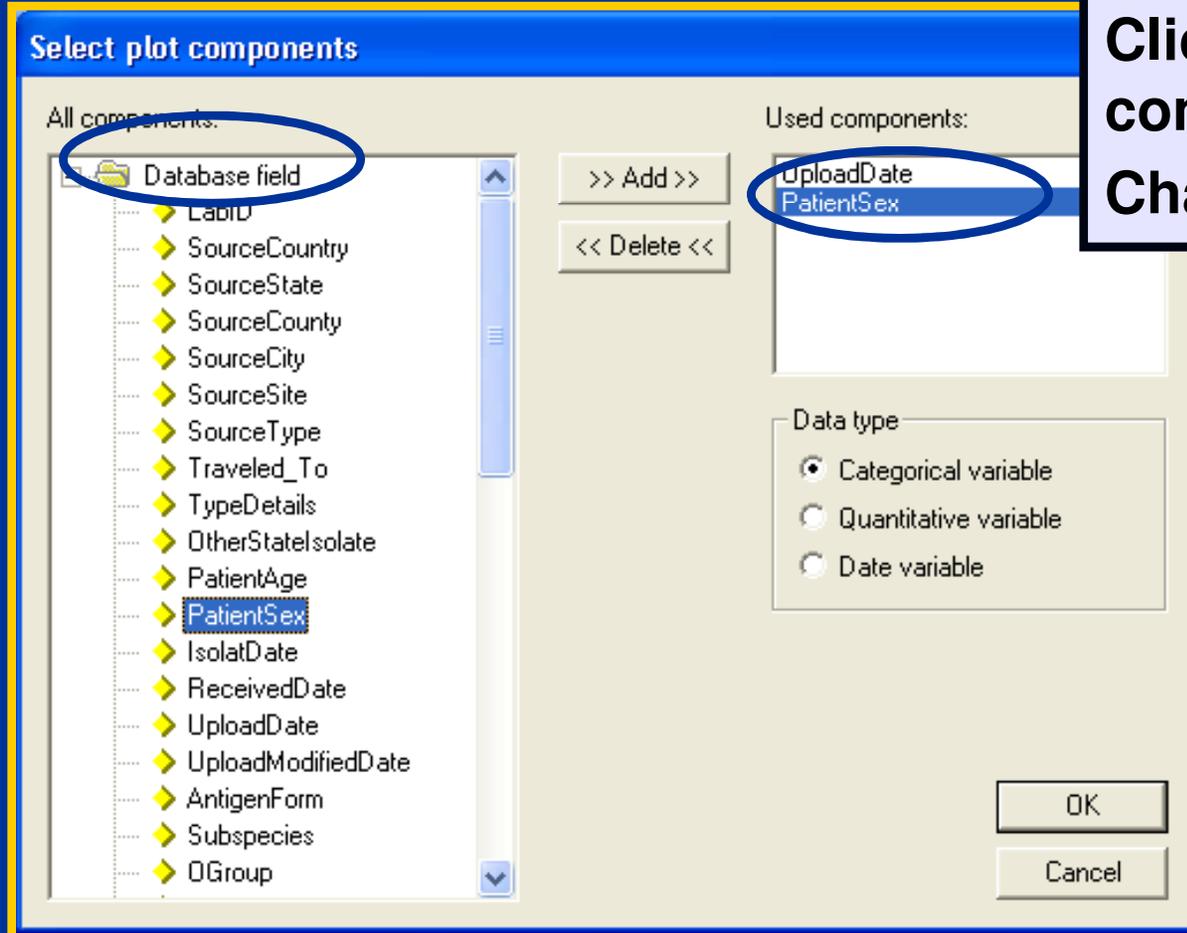
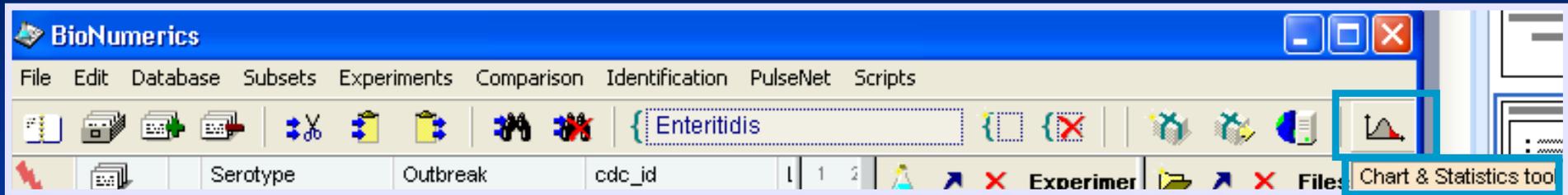
Chart & Statistics

The screenshot shows a software window titled "Comparison" with a menu bar (File, Edit, Layout, Groups, Clustering, Dimensioning, Bandmatching, Characters, Sequence, TrendData, Composite, Window) and a toolbar. The toolbar includes a "Chart & Statistics" icon, which is highlighted with a red box. A tooltip for this icon shows a small chart with a bell curve and a red shaded area under the curve. The main window area is divided into three panes: "Dendrogr...", "Experiment data", and "Information field". The "Information field" pane contains a table with a "Key" column and several entries, each preceded by an orange arrow. The status bar at the bottom left indicates "18 entries".

Key
AL__AL-8002391-06
AL__AL-8002392-06
AL__AL-8002394-06
AL__AL-8002395-06
AL__AL-8002397-06
AL__AL-8002441-06
AL__AL-8002442-06

...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

Select plot components

All components:

- Database field
 - LabID
 - SourceCountry
 - SourceState
 - SourceCounty
 - SourceCity
 - SourceSite
 - SourceType
 - OtherStateIsolate
 - PatientAge
 - PatientSex
 - IsolatDate
 - ReceivedDate
 - UploadDate
 - UploadModifiedDate
 - AntigenForm
 - Subspecies
 - OGroup
 - Serotype
 - Outbreak

Used components:

- LabID
- UploadDate

>> Add >>

<< Delete <<

Data type

- Categorical variable
- Quantitative variable
- Date variable

Convert to interval data

Group by day

- Group by day
- Group by week
- Group by month
- Group by quarter
- Group by year

Choose data type

If date, choose whether to make it interval data

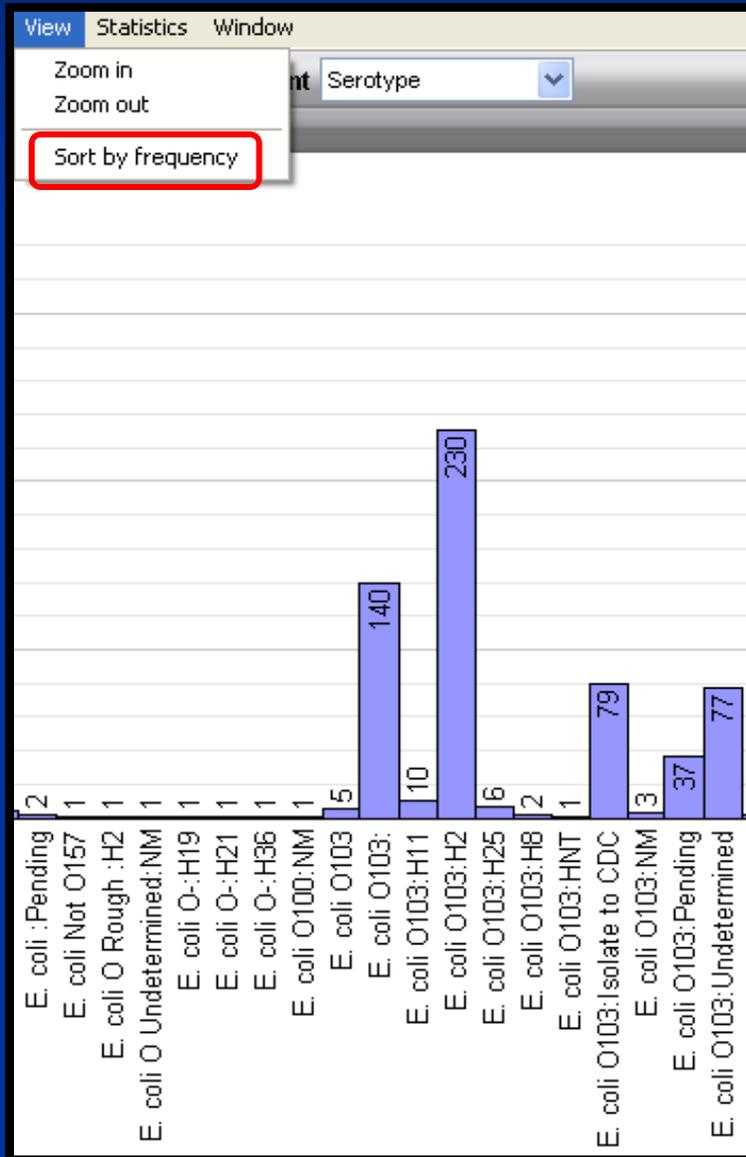
Chart & Statistics Plot Components

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

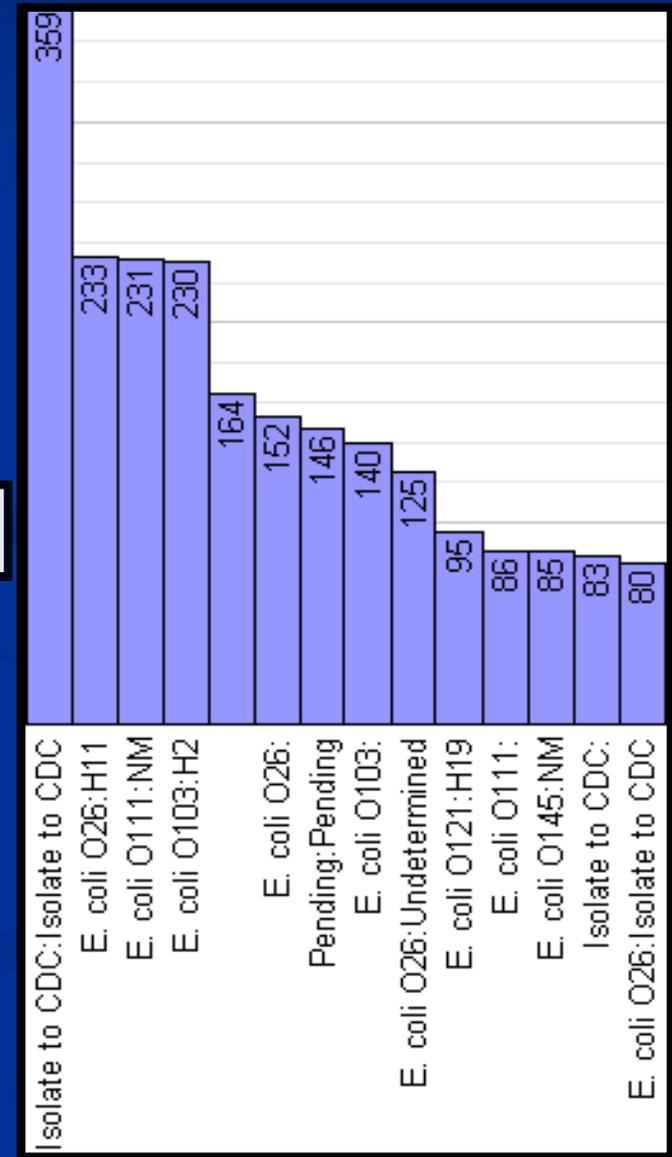
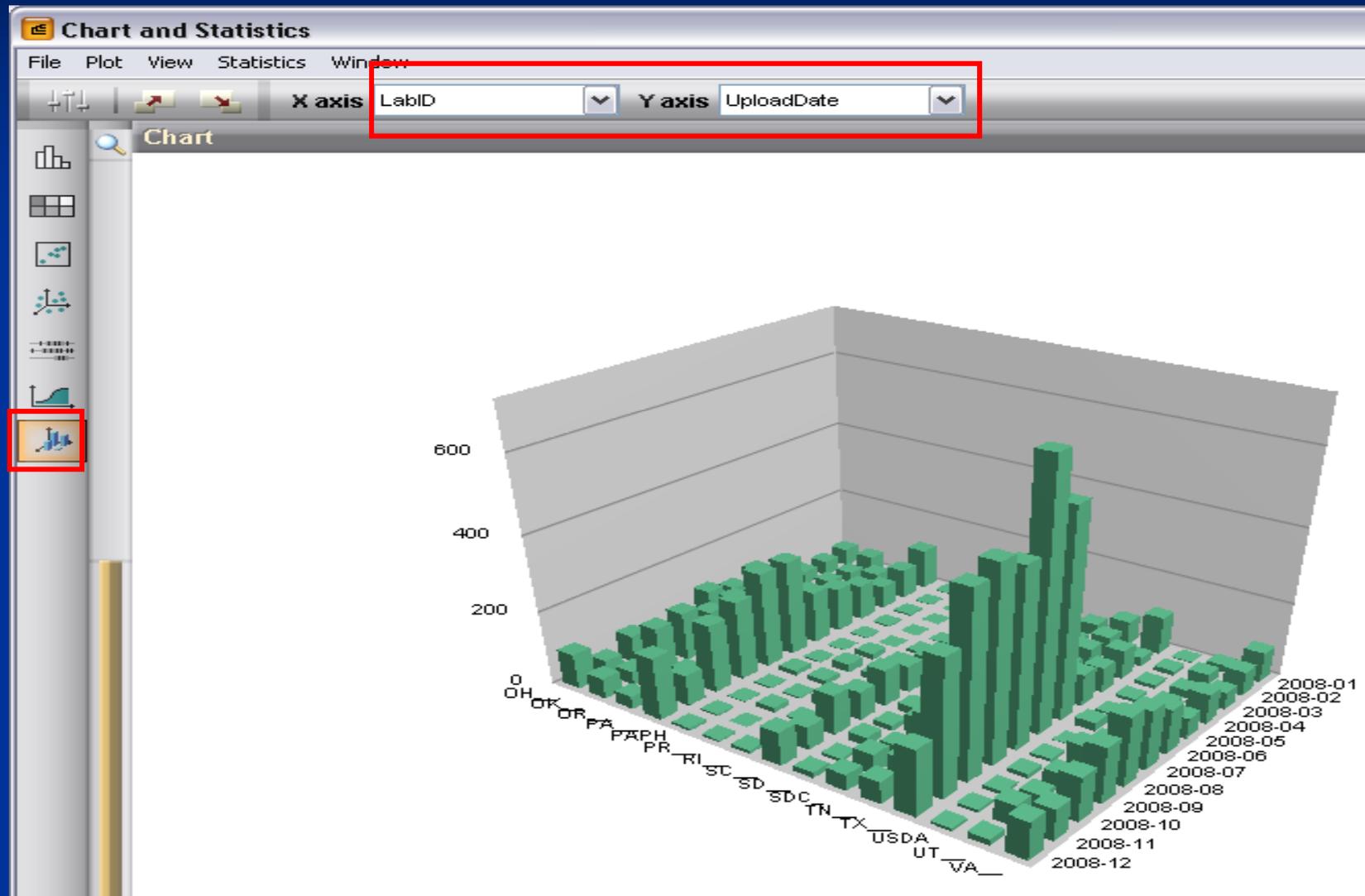


Chart & Statistics: Types of Graphs



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
 - OtherStatelolate
 - UploadDate
 - UploadModifiedDate

Used components:

- SourceState

Data type

- Categorical variable
- Quantitative variable
- Date variable

>> Add >>

<< Delete <<

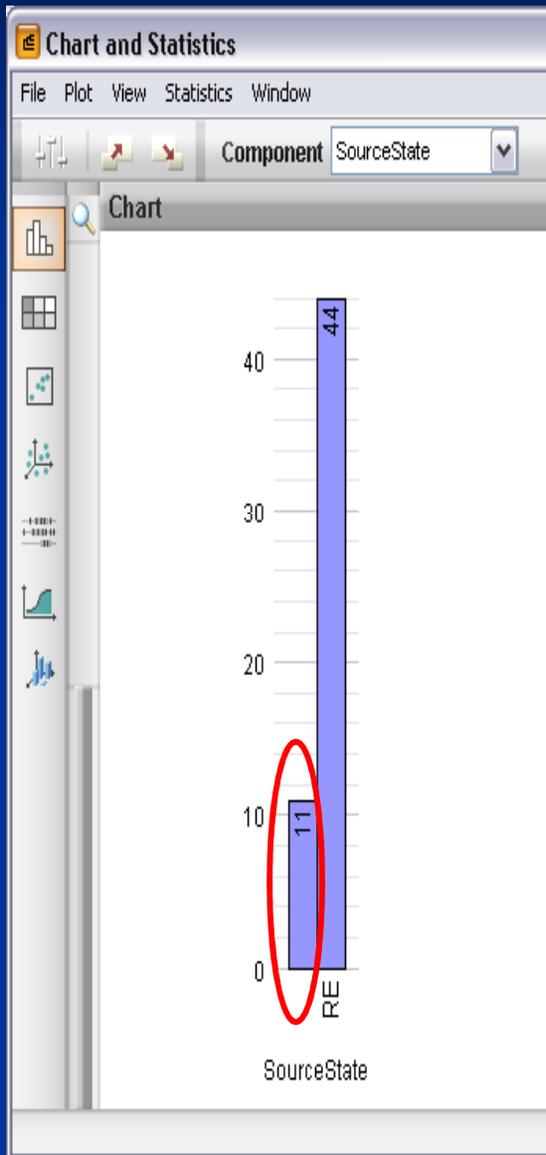
OK

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)

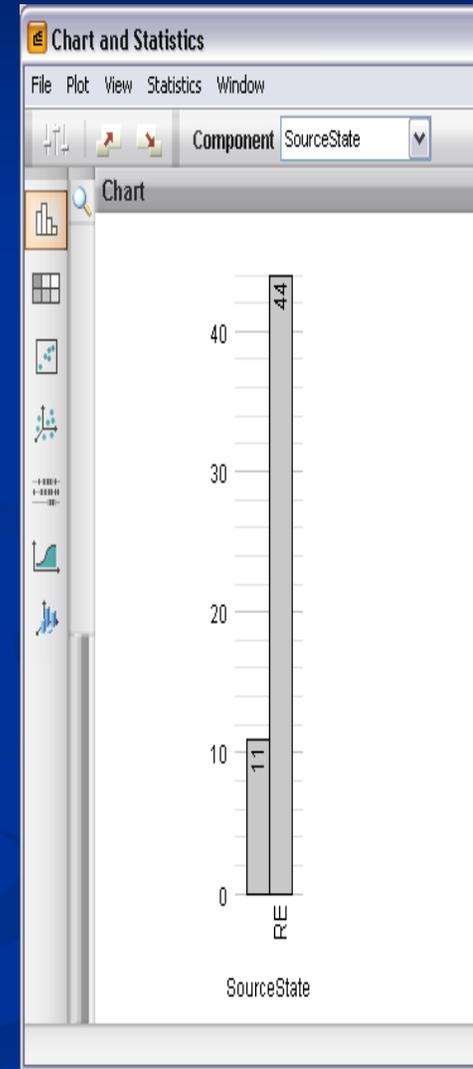
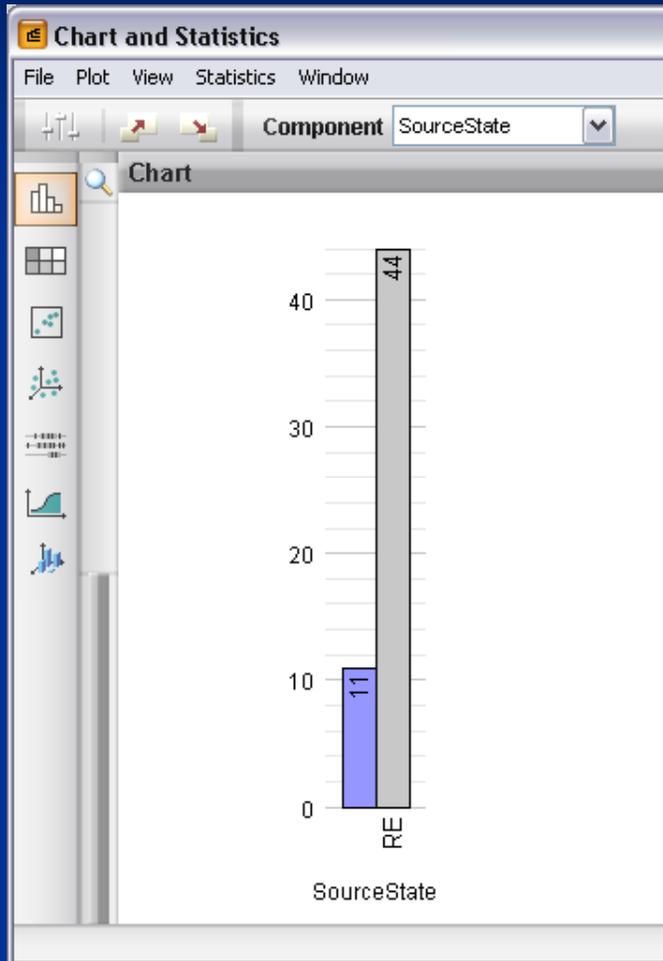


Chart & Statistics Tool: Example



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

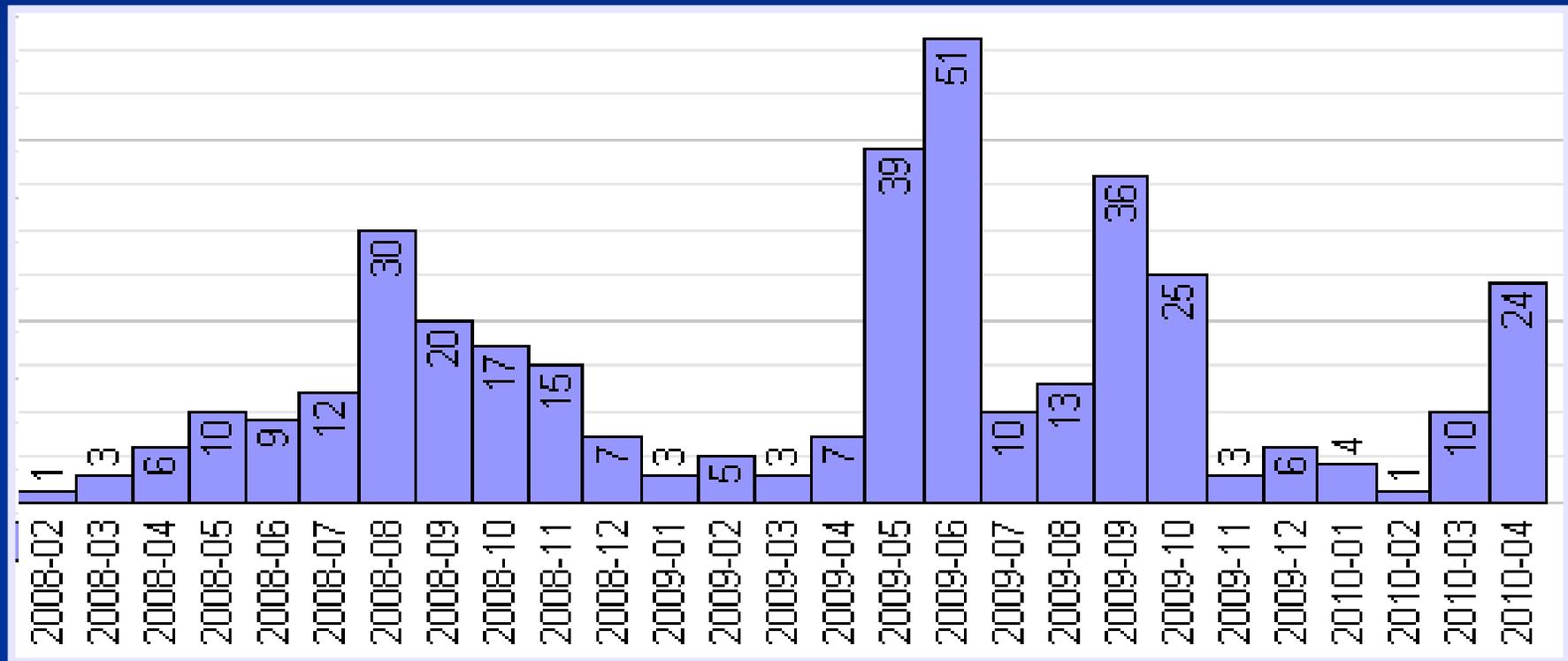
Utility of Graphs in Cluster Detection and Reporting

- 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

The screenshot shows the BioNumerics software interface. The main window displays a table of database entries with columns for Index, abID, Serotype, and PFGE-Xbal-patt... The table contains several rows of data, including entries with Index 7, 8, 9, and 10. A toolbar at the top includes icons for file operations, editing, and analysis. A tooltip 'Unselect all entries (F4)' is visible over the 'Unselect all entries' icon. A smaller window titled 'Database entries' is open over the table, showing a list of entries with columns for Index and Key. A tooltip 'Query recent uploads' is visible over the 'Query recent uploads' icon. A dialog box is open in the bottom right corner, prompting the user to 'Enter number of days before now' with the value '60' entered in the input field. The dialog has 'OK' and 'Cancel' buttons.

Index	abID	Serotype	PFGE-Xbal-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**

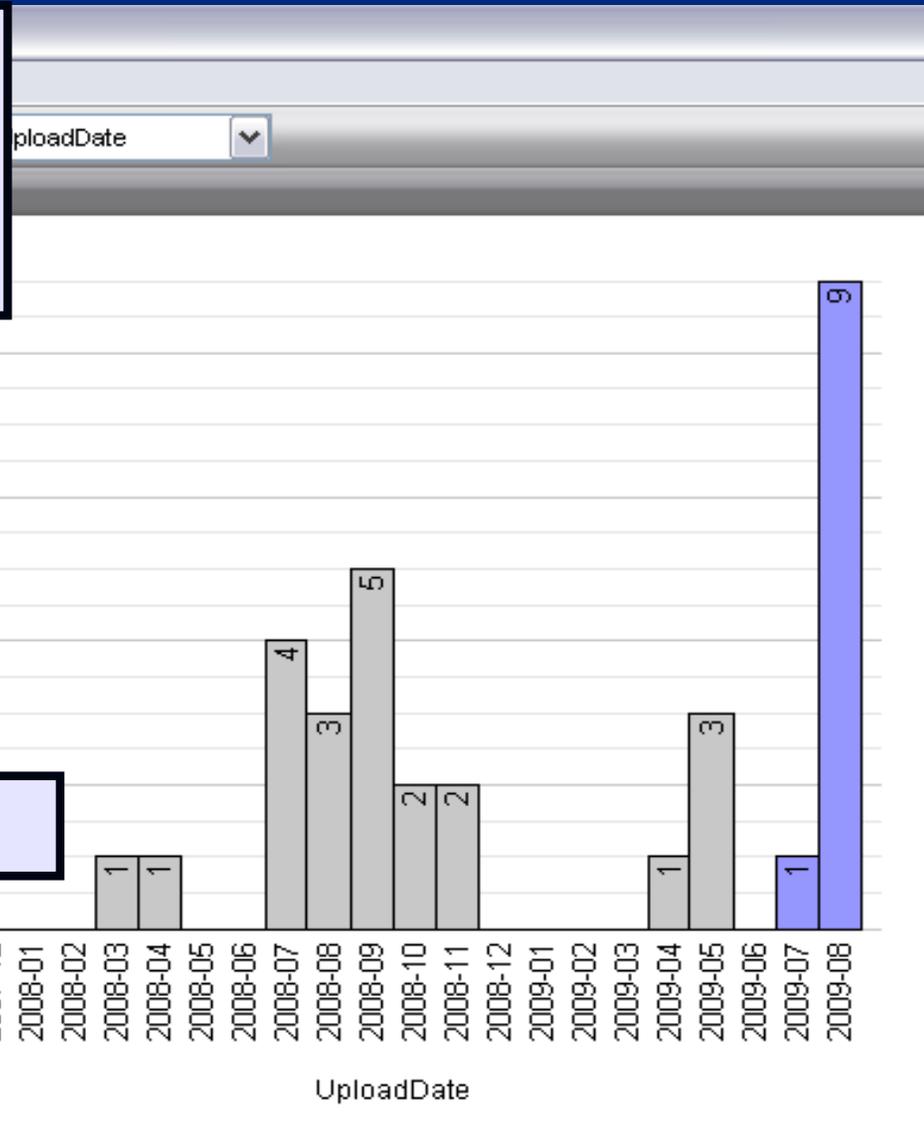
**Perform a Hot
List search of the
past 60 days**

Enter number of days before now

OK Cancel

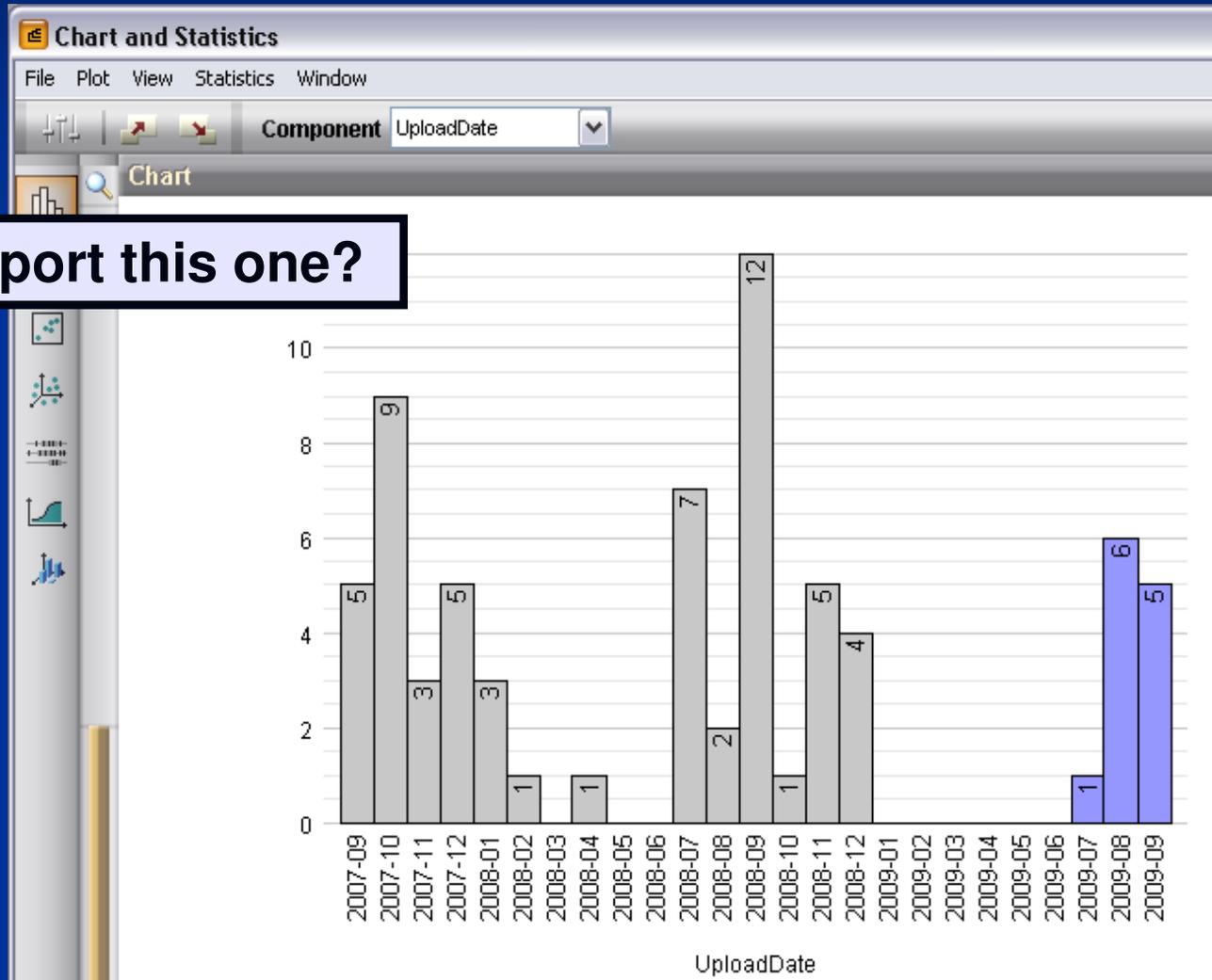
Utility of Graphs in Cluster Detection and Reporting: Example 1

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

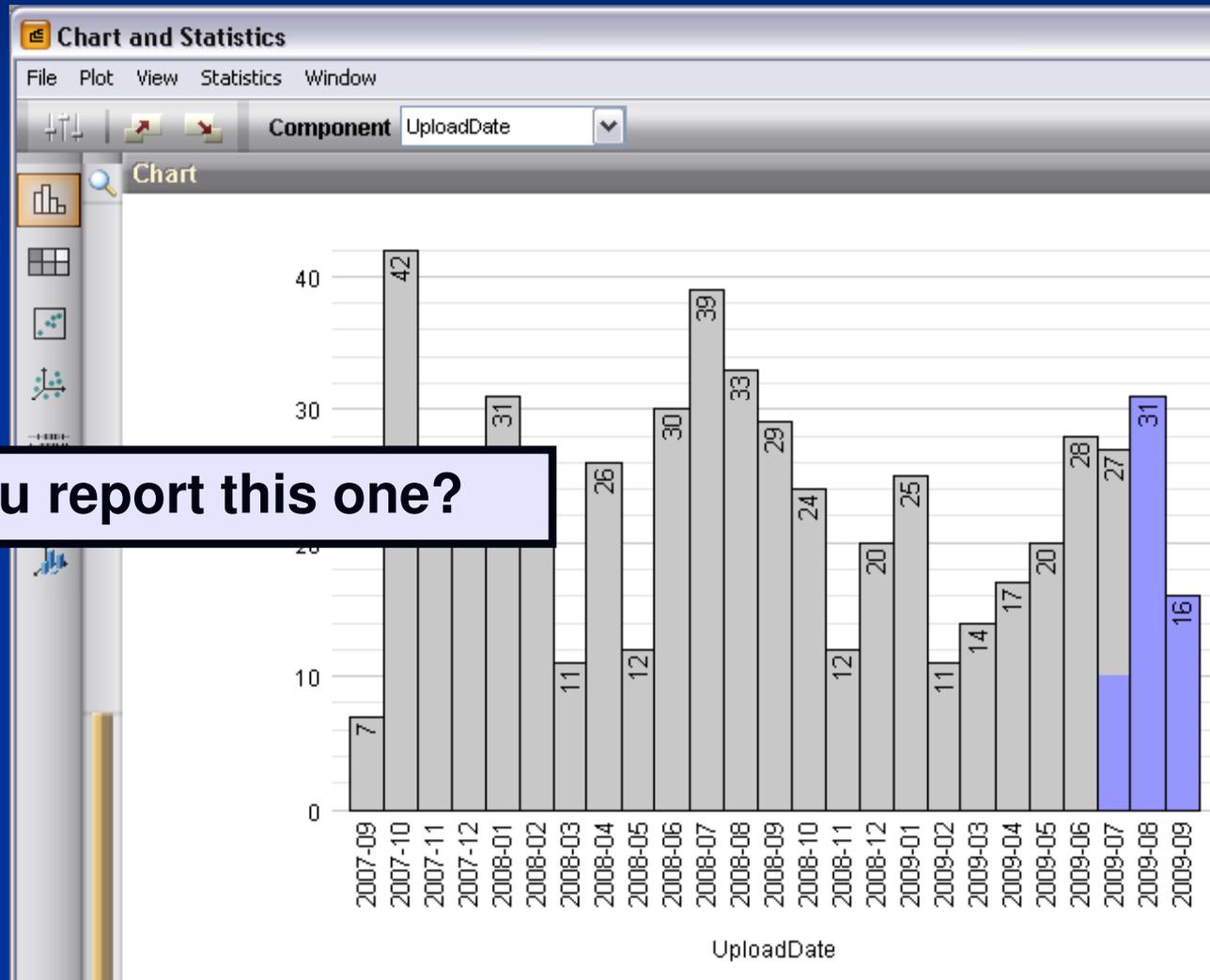


Should you report this one?

Utility of Graphs in Cluster Detection and Reporting: Example 2



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

Using Groups

The screenshot shows a software window titled "exercise (Comparison)". The menu bar includes "File", "Edit", "Layout", "Groups", "Clustering", "Dimensioning", "Bandmatching", "Characters", "Sequence", "TrendData", "Composite", and "Window". The "Groups" menu is open, with "Assign selection to" circled in black. Other menu items include "Create groups from database field", "Show groups using colors" (checked), "Edit group colors...", "Partitioning of groups...", "Group separation...", and "Multivariate Analysis of Variance...". Below the menu is a list of 30 group options, each with a color and a symbol. Option [6] is highlighted in blue. The background shows a data table with columns "otype" and "SourceCountry".

Group	Color	Symbol
[1]	Green	Black circle
[2]	Red	Black star
[3]	Purple	Black diamond
[4]	Yellow	Black target
[5]	Cyan	Black square
[6]	Teal	Black triangle
[7]	Brown	Black cross
[8]	Purple	Black gear
[9]	Green	Black diamond
[10]	Light green	Black star
[11]	Blue	Black Roman numeral II
[12]	Pink	Black arrow
[13]	Yellow	Black box with X
[14]	Pink	Black arrow
[15]	Red	Black box with X
[16]	Green	Black X
[17]	Purple	Black drop
[18]	Teal	Black circle with X
[19]	Dark blue	Black X
[20]	Brown	Black hand
[21]	Light green	Black hand
[22]	Light green	Black hand
[23]	Purple	Black diamond
[24]	Purple	Black smiley
[25]	Olive	Black frowny
[26]	Pink	Black hand
[27]	Olive	Black snowflake
[28]	Purple	Black hand
[29]	Red	Black hourglass
[30]	Cyan	Black hand

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 screenshot shows the BioNumerics software interface. The 'Groups' menu is open, highlighting the option 'Create groups from database field'. A yellow arrow points from this menu option to the 'Outbreak' column in the 'Information fields' table below.

late	PFGE-Spel-status	Outbreak
→	Unconfirmed	0708NYCJM6-1c
→	Unconfirmed	0801ORJPX-1c
→	Unconfirmed	0802MLJPX-1c
→	Unconfirmed	0803VAJPX-1c

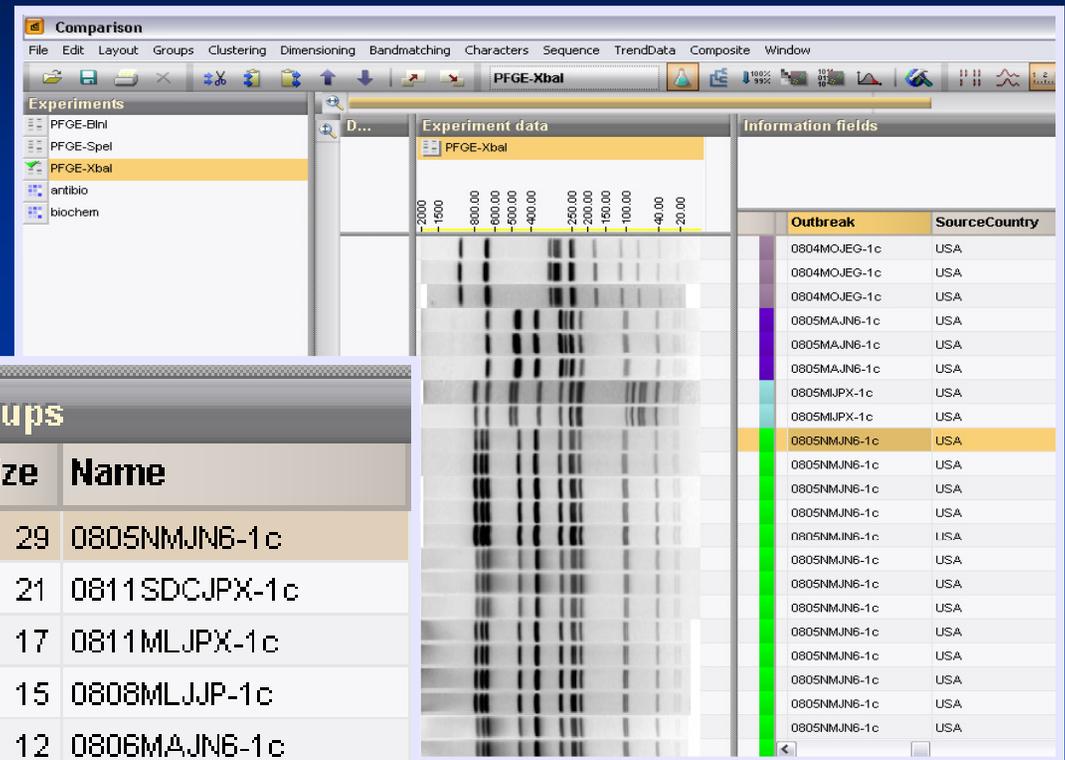
To automatically assign groups, select the database field of interest

Go to Groups and Select Create groups from database field

Utility of Groups



Choose the method of creating the groups and click OK

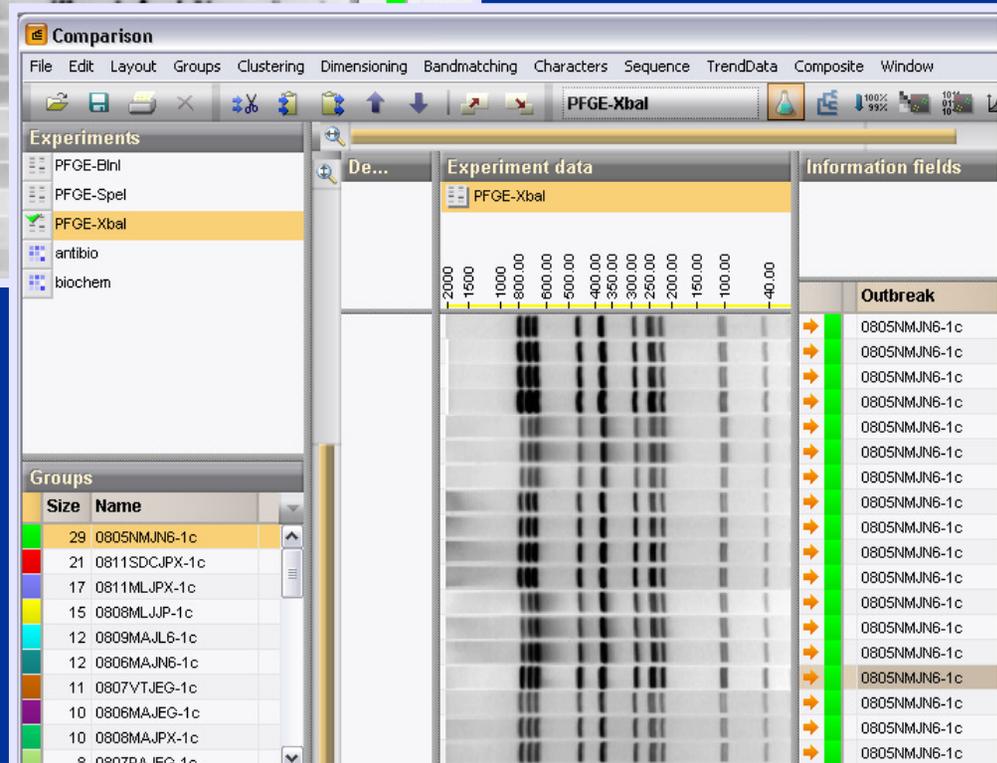
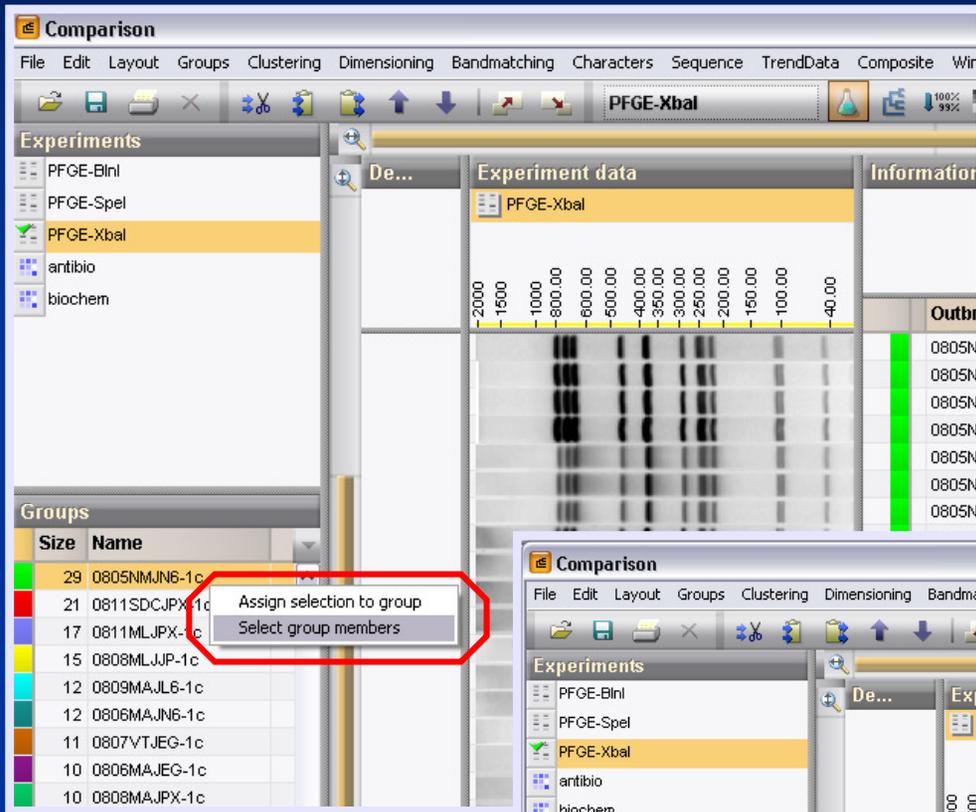


Groups		
	Size	Name
■	29	0805NMJN6-1c
■	21	0811SDCJPX-1c
■	17	0811MLJPX-1c
■	15	0808MLJJP-1c
■	12	0806MAJN6-1c
■	12	0809MAJL6-1c
■	11	0807VTJEG-1c
■	10	0808MAJPX-1c
■	10	0806MAJEG-1c
■	8	0807PAJEG-1c
■	7	0807MAJPX-1c
■	6	0807MAJM6-1c
■	5	0811MAJF6-1c
■	5	0808MAJPX-2c
■	5	0812MLJJX-1

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

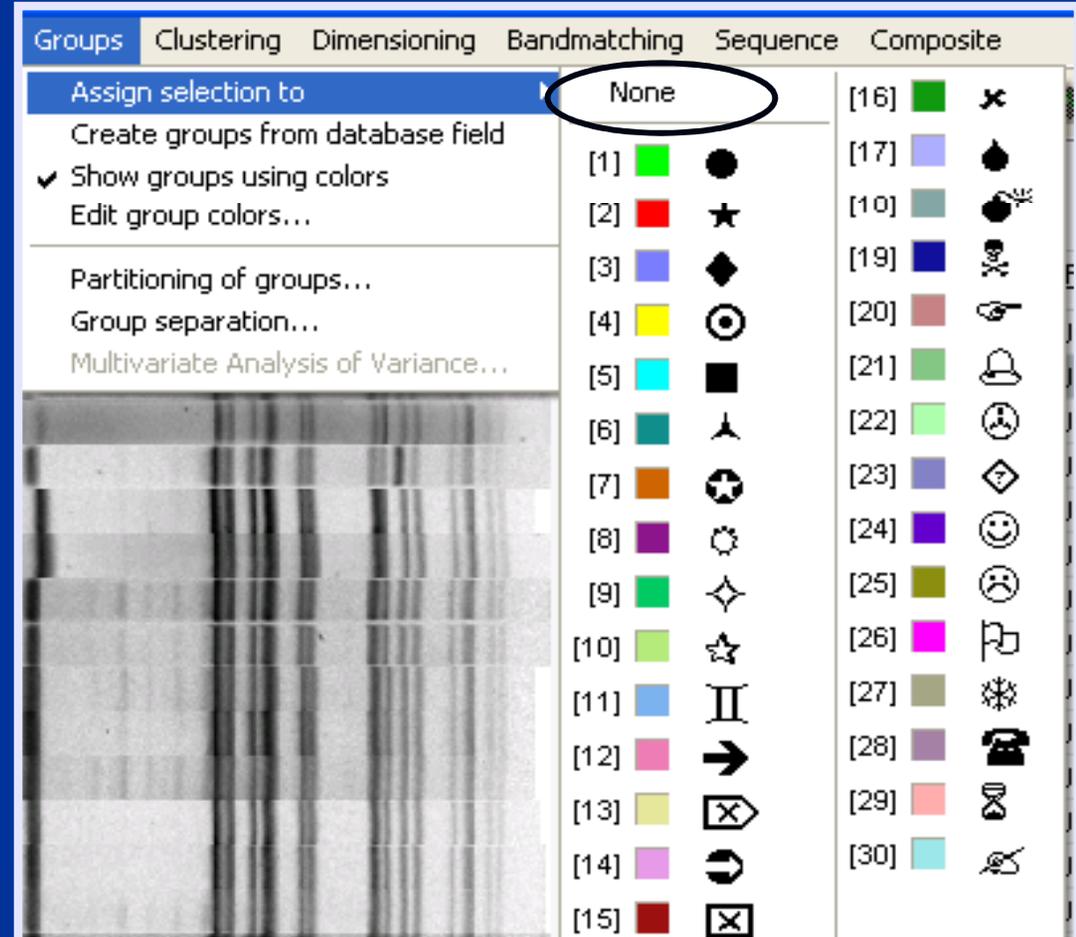
Utility of Groups

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

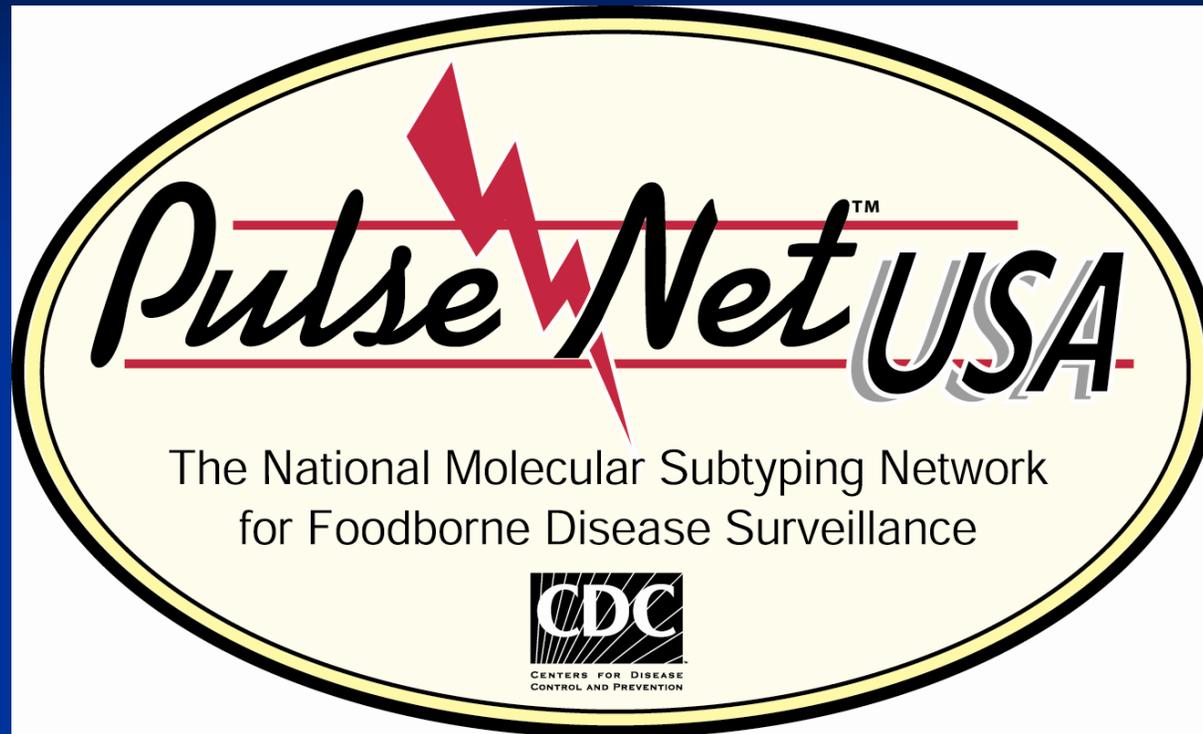


Utility of Groups

- 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