

# Using the Groups and Chart and Statistics Tools



Shannon Brown

May 2009



# Overview

- Utility of Groups
- Printing Groups
- Customizing Colors
- Using the Chart and Statistics Tool
- Types of Graphs
- Utility of Graphs in Cluster Detection and Reporting

# Utility of Groups

- In using groups/colors you can distinguish isolates to another level while working with a comparison
  - Example: naming patterns
    - If you make a comparison of all the similar isolates at first glance, you can go back through with a fine tooth comb and designate alternative patterns

# Utility of Groups

Experiment data		Information fields			
PFGE-Xbal		otype	SourceCountry	SourceState	SourceCounty
		elberg	USA	CA	Orange County
		elberg	USA	WI	
		elberg	USA	NV	Washoe
		→ elberg	USA	WY	Campbell
		elberg		IL	Peoria
		→ elberg	USA	PA	Philadelphia
		→ elberg		WA	King
		elberg	USA	CA	Los Angeles
		→ elberg	USA	OH	Clermont
		→ elberg	USA	OK	Custer
		→ elberg	USA	IL	
		→ elberg	USA	AR	Pulaski
		→ elberg	USA	NY	Rockland

**Choose isolates that you want to highlight as a group**

# Using Groups

The screenshot shows a software interface with a menu bar at the top containing 'File', 'Edit', 'Layout', 'Groups', 'Clustering', 'Dimensioning', 'Bandmatching', 'Characters', 'Sequence', 'TrendData', and 'Composite'. The 'Groups' menu is open, displaying options: 'Assign selection to', 'Create groups from database field', 'Show groups using colors', 'Edit group colors...', 'Partitioning of groups...', 'Group separation...', and 'Multivariate Analysis of Variance...'. The 'Show groups using colors' option is checked and circled in green. Below the menu is a table with 30 rows, each containing a number in brackets, a color swatch, and a symbol. The 'Groups' menu title is circled in blue. In the background, a table with columns 'otype' and 'SourceCountry' is visible, showing rows for 'elberg' and 'USA'.

otype	SourceCountry
elberg	USA

None		
[16]	Green	✕
[17]	Light Blue	🍷
[18]	Grey	🌿
[19]	Dark Blue	🍷
[20]	Brown	👉
[21]	Light Green	👤
[22]	Light Green	👤
[23]	Light Purple	👤
[24]	Purple	😊
[25]	Olive	😞
[26]	Magenta	👤
[27]	Olive	❄️
[28]	Purple	👤
[29]	Red	🕒
[30]	Cyan	👤

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

Experiment data		Information fields			
PFGE-XbaI		Serotype	SourceCountry	SourceState	SourceC
		Heidelberg	USA	CA	Orange Coun
		Heidelberg	USA	WI	
		Heidelberg	USA	NV	Washoe
		Heidelberg	USA	WY	Campbell
		Heidelberg		IL	Peoria
		Heidelberg	USA	PA	Philadelphia
		Heidelberg		WA	King
		Heidelberg	USA	CA	Los Angeles
		Heidelberg	USA	OH	Clermont
		Heidelberg	USA	OK	Custer
		Heidelberg	USA	AR	Pulaski
		Heidelberg	USA	NY	Rockland

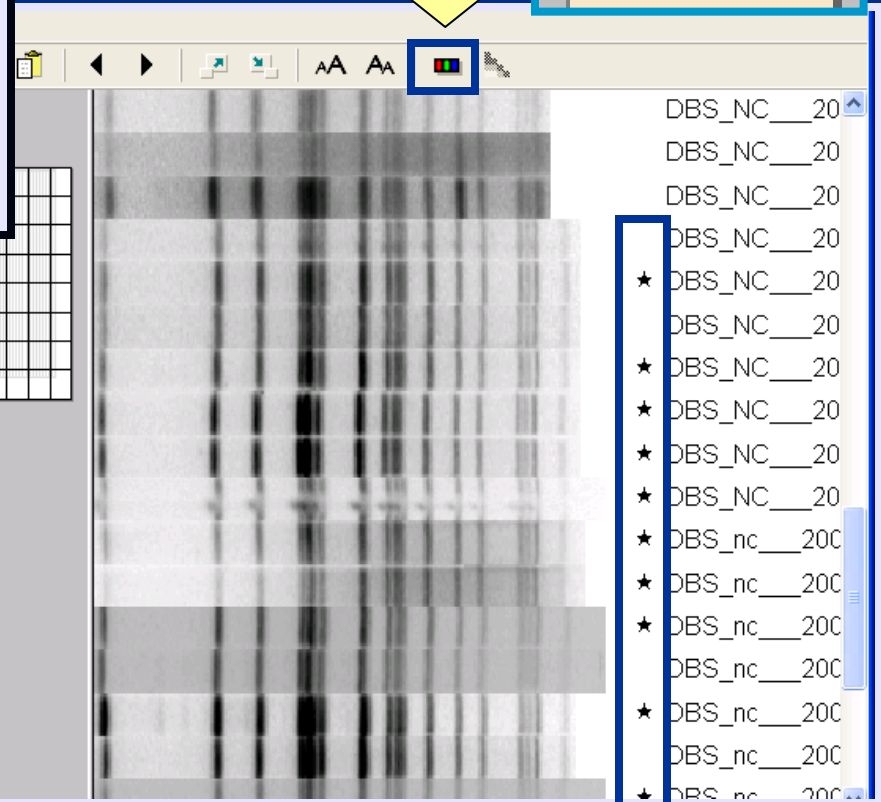
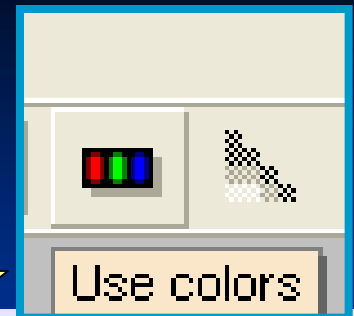
The isolates that are selected in this comparison are further divided by color after more detailed analysis

# Overview

- Utility of Groups
- Printing Groups
- Customizing Colors
- Using the Chart and Statistics Tool
- Types of Graphs
- Utility of Graphs in Cluster Detection and Reporting

# Printing Groups

When printing grouped isolates in a comparison, the colors will automatically change to shapes unless you choose to print in color



CT__02032150	Typhi	JPPX01.049
MI__08ST000498	Typhi	JPPX01.053
VA__08-0631	Typhi	JPPX01.002
WV__M08001411	Typhi	JPPX01.048
LAC_Z20894	Typhi	JPPX01.056
NY__BAC08000024...	Typhi	JPPX01.002
NY__BAC08000024...	Typhi	JPPX01.002
CASC_08SCPH06708	Typhi	JPPX01.045
PA__08E00636	Typhi	JPPX01.034
GA__08C0365113	Typhi	JPPX01.034
HI__N08-148	Typhi	JPPX01.002
NJ__800895	Typhi	JPPX01.022
NYC_nyc08-100601...	Typhi	JPPX01.074
NYC_nyc08-100601...	Typhi	JPPX01.004
NYC_nyc08-100608...	Typhi	JPPX01.014
NYC_nyc08-100610...	Typhi	JPPX01.002
VA__08-0691	Typhi	JPPX01.0026
CA__M08X01512	Typhi	JPPX01.0459
NY__BAC08000028...	Typhi	JPPX01.0480
PA__08E00706	Typhi	JPPX01.0026
MO__MOENT0773-08	Typhi	JPPX01.0704



# Overview

- Utility of Groups
- Printing Groups
- Customizing Colors
- Using the Chart and Statistics Tool
- Types of Graphs
- Utility of Graphs in Cluster Detection and Reporting

# Customizing Colors

The 'Group colors' dialog box is shown with a grid of 30 color swatches. The swatches are arranged in 10 rows and 3 columns, numbered 1 through 30. The first row contains swatches 1 (blue), 11 (light blue), and 21 (green). The second row contains 2 (red), 12 (pink), and 22 (light green). The third row contains 3 (purple), 13 (yellow-green), and 23 (dark purple). The fourth row contains 4 (yellow), 14 (light purple), and 24 (dark purple). The fifth row contains 5 (cyan), 15 (dark red), and 25 (olive green). The sixth row contains 6 (teal), 16 (green), and 26 (magenta). The seventh row contains 7 (orange), 17 (light purple), and 27 (olive green). The eighth row contains 8 (purple), 18 (grey-blue), and 28 (grey-purple). The ninth row contains 9 (green), 19 (dark blue), and 29 (pink). The tenth row contains 10 (light green), 20 (brown), and 30 (light blue). To the right of the grid is a color gradient control with a large color swatch and three sliders labeled R (0%), G (100%), and B (0%). Below the grid are three buttons: 'Default', 'Pastels', and 'Range...'. At the bottom right are 'Ok' and 'Cancel' buttons. A 'Saved color schemes' section contains a dropdown menu and 'Save as...' and 'Delete selected...' buttons.

Here, under edit group colors, you can change the tone of each color from default or change entirely to “pastels.” You can also choose to do a color gradient with “range.”

# Customizing Colors

The image displays two screenshots of the 'Group colors' dialog box. The left screenshot shows a grid of 30 color swatches (numbered 1-30) and three buttons at the bottom: 'Default', 'Pastels', and 'Range...'. The 'Pastels' button is circled in green. The right screenshot shows the same dialog with a different color scheme selected. It includes three sliders for Red (R: 60%), Green (G: 89%), and Blue (B: 89%) and a 'Saved color schemes' section with a dropdown menu, 'Save as...', and 'Delete selected...' buttons. A yellow arrow points to the 'Save as...' button, and a text box below it says 'Save your color choices that you create'.

# Overview

- Utility of Groups
- Printing Groups
- Customizing Colors
- Using the Chart and Statistics Tool
- Types of Graphs
- Utility of Graphs in Cluster Detection and Reporting

# Chart and Statistics Tool: Functionality

- The utility of the Chart and Statistics Tool:
  - Database statistics
    - Pattern frequencies
    - Date-oriented calculations
    - Multi-variable options

# Chart and Statistics Tool Components

**Entry search**

**Simple query** **Advanced query tool**

Subspecies

OGroup

Serotype Enteritidis\*

Outbreak

cdc\_id

ListMember

PFGE-Xbal

PFGE-BlnI

PFGE-Spel

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 and Statistics Tool Components

The screenshot displays the BioNumerics software interface. The main window is titled "BioNumerics" and features a menu bar (File, Edit, Database, Subsets, Experiments, Comparison, Identification, PulseNet, Scripts, Window) and a toolbar. A red box highlights the "Chart & Statistics" icon in the toolbar. A callout box shows a magnified view of this icon, which depicts a histogram with a bell curve. Below the toolbar is a "Database entries" table with columns for Key, LabID, and SourceCount. The table lists various entries with keys like AL\_\_AL-8002391-06 and source counts like USA and Canada. To the right, there are panels for "Experiments" (listing PFGE-Xbal and PFGE-BlnI) and "Files" (listing AL\_\_AL06126, AL\_\_AL06127, and AL\_\_AL06128). The status bar at the bottom shows "Database: Salmonella Training 1000" and "1003 entries 5 experiments".

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

Name	Created
AL__AL06126	2006-09-29 10:4
AL__AL06127	2006-09-29 10:4
AL__AL06128	2006-09-29 10:4

Database: Salmonella Training 1000 1003 entries 5 experiments \\cdd\project\CCID\_NCZVED\_DFBMD\_PulseNet\Software - Hardware\Database

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

# Chart and Statistics Tool Components

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. A red box highlights a chart icon in the toolbar. Below the toolbar, there are three main panels: "Dendrogr...", "Experiment data", and "Information field". The "Information field" panel contains a table with a "Key" column and several rows of alphanumeric strings. A red box highlights a small chart icon within the "Information field" panel. 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

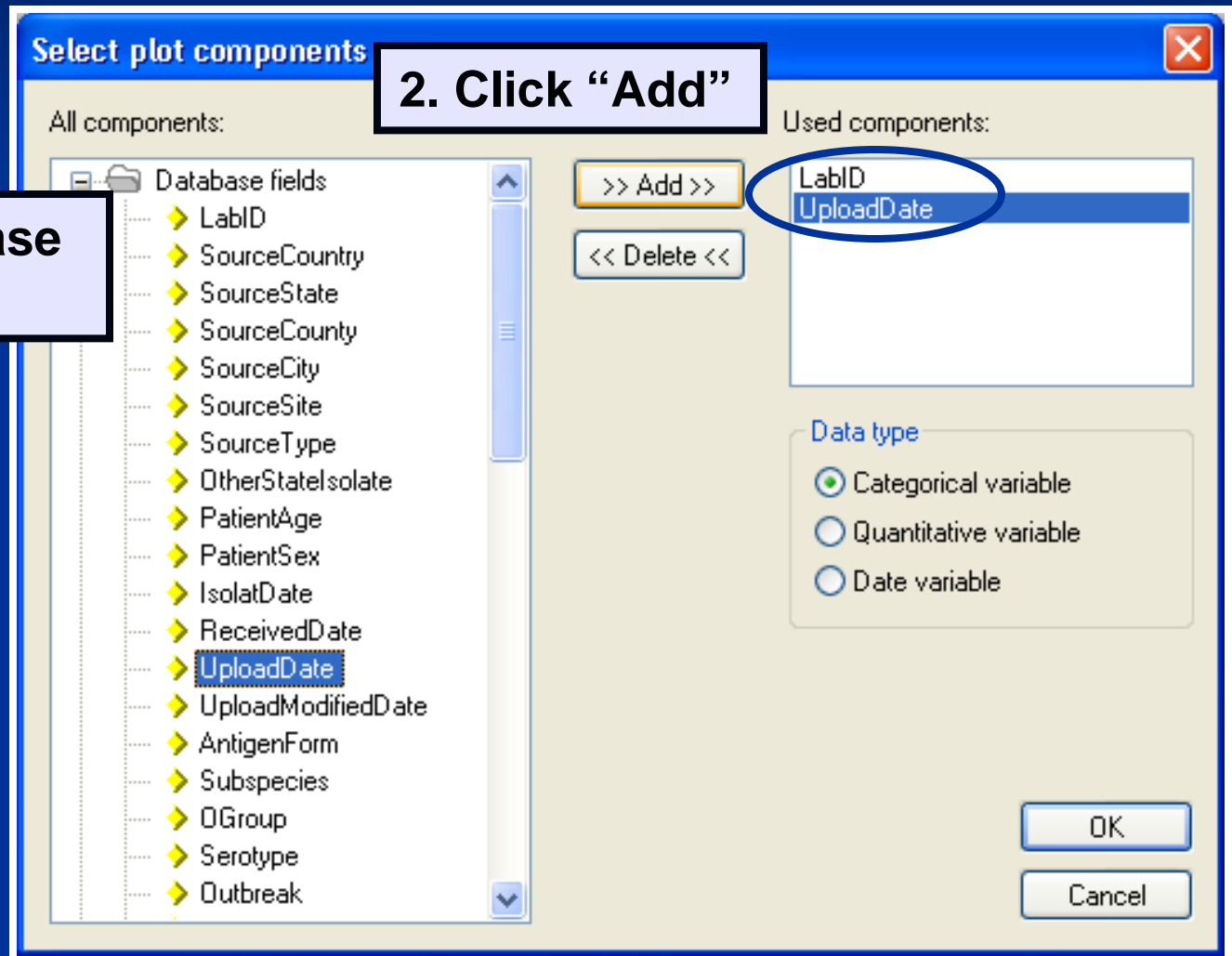
...from a comparison window



# Chart and Statistics Tool Components

1. Choose database components

2. Click "Add"



# Chart and Statistics Tool Components

**Select plot components**

All components:

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

Used components:

- LabID
- UploadDate

**3. Choose data type**

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

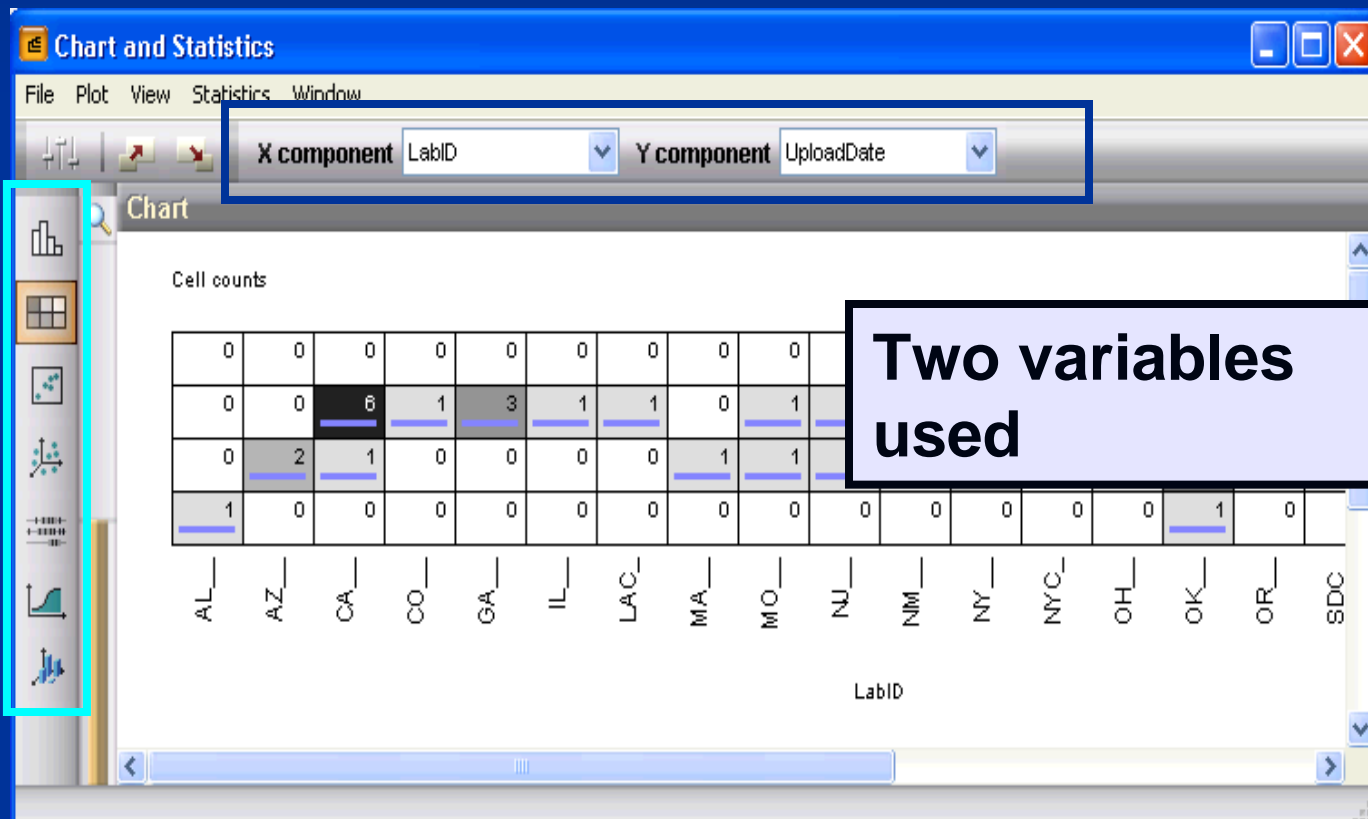
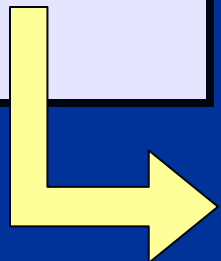
**4. If date, choose whether to make it interval data**

# Overview

- Utility of Groups
- Printing Groups
- Customizing Colors
- Using the Chart and Statistics Tool
- Types of Graphs
- Utility of Graphs in Cluster Detection and Reporting

# Chart & Statistics: Changing graph type

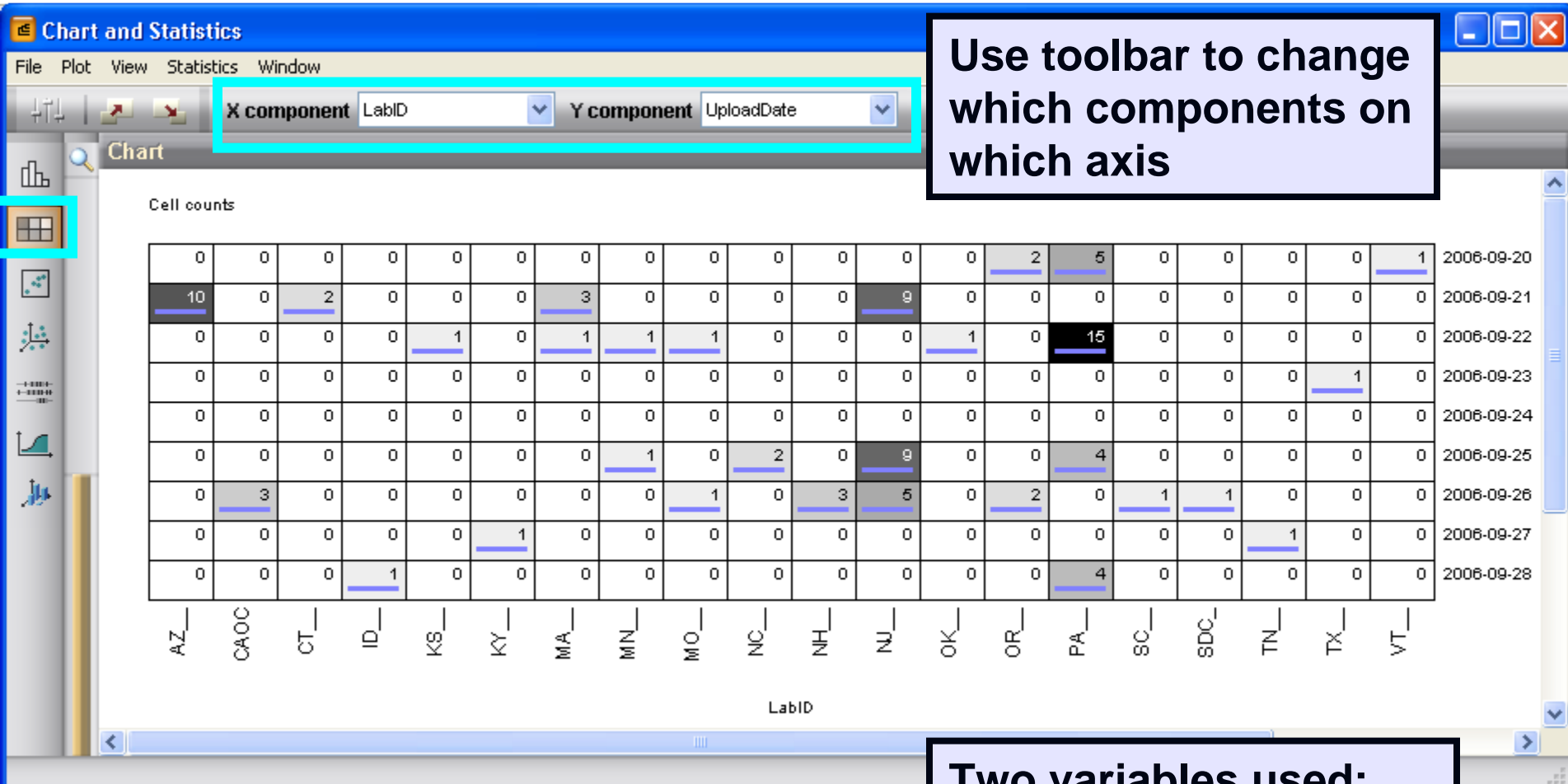
Use  
toolbar to  
change  
chart and  
type



Two variables  
used

# Types of Graphs

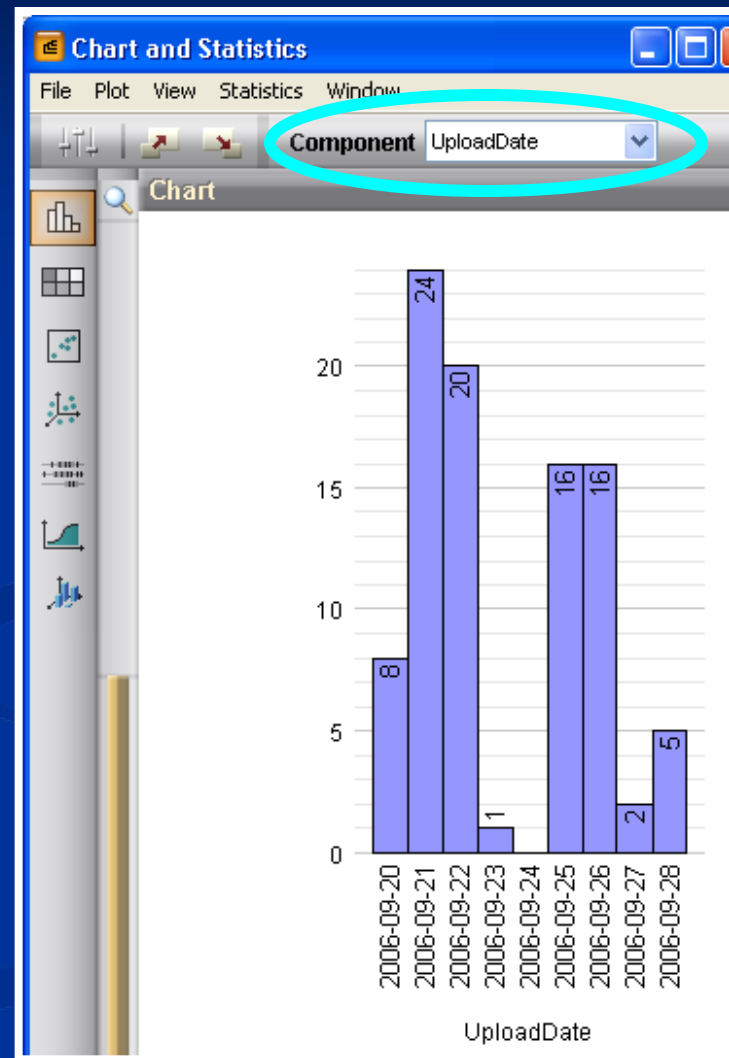
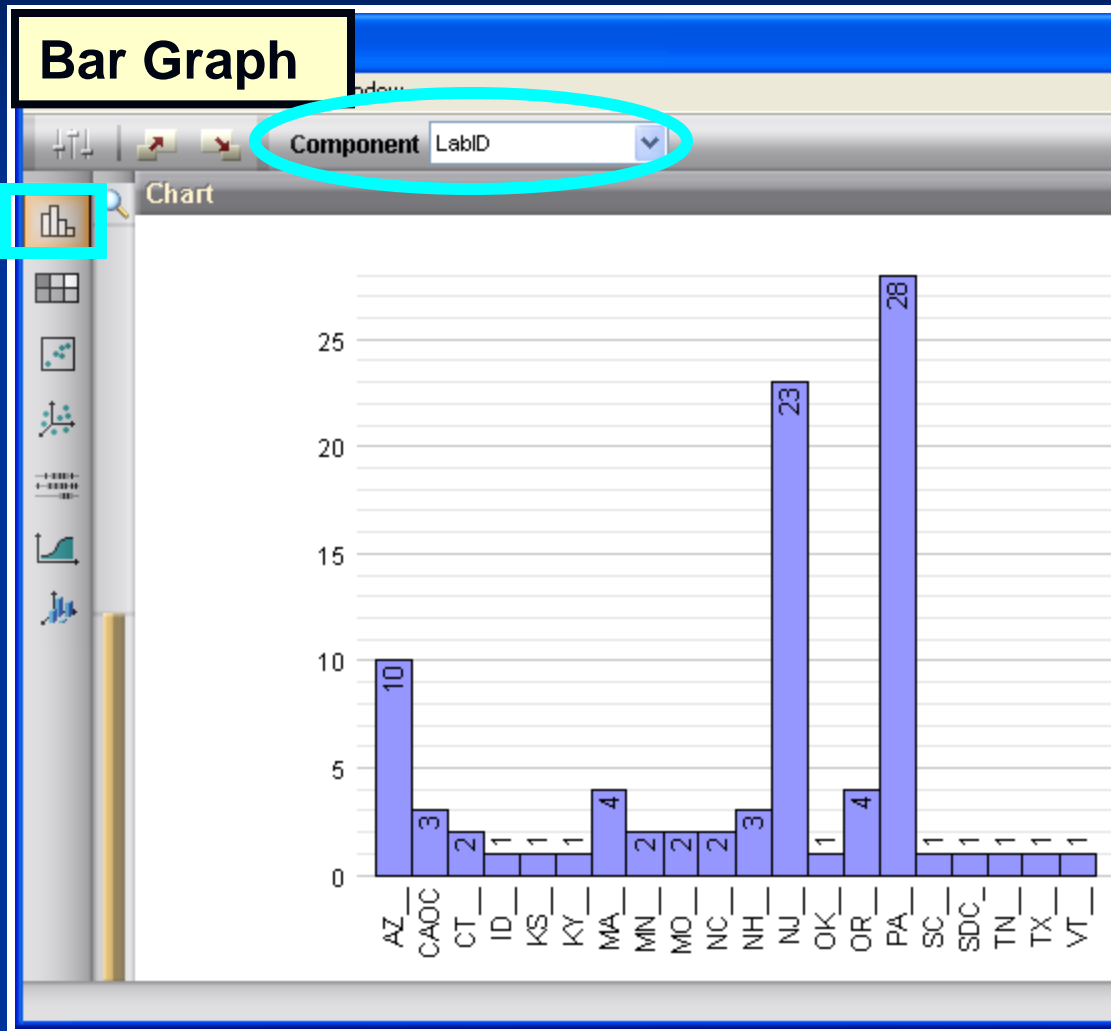
**2D Contingency Table: shows the association between 2 categorical variables**



**Use toolbar to change which components on which axis**

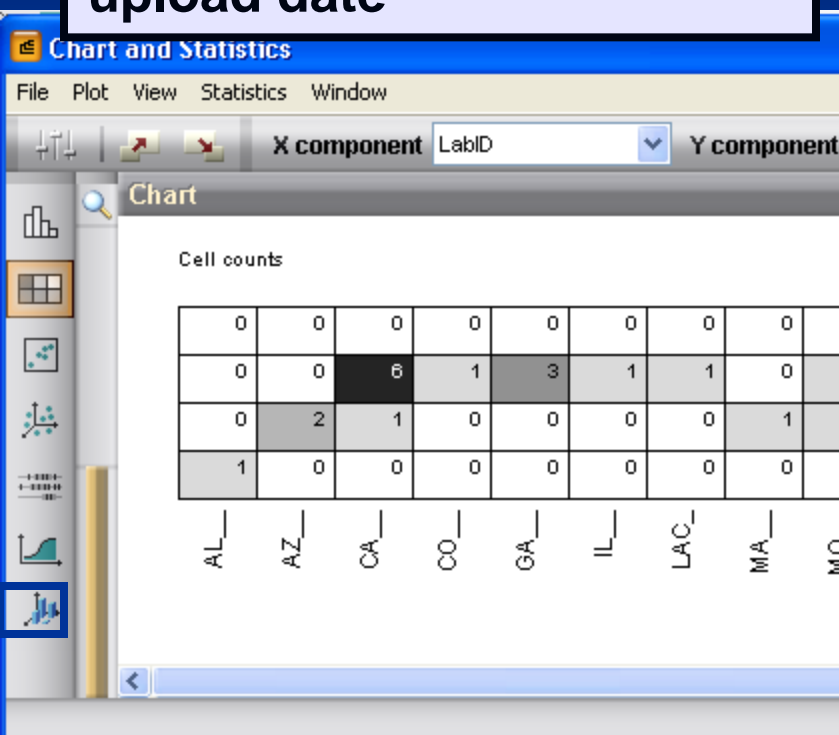
**Two variables used: LabID and UploadDate**

# Types of Graphs

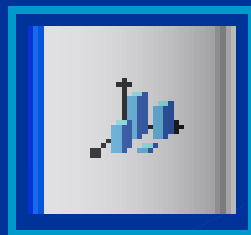
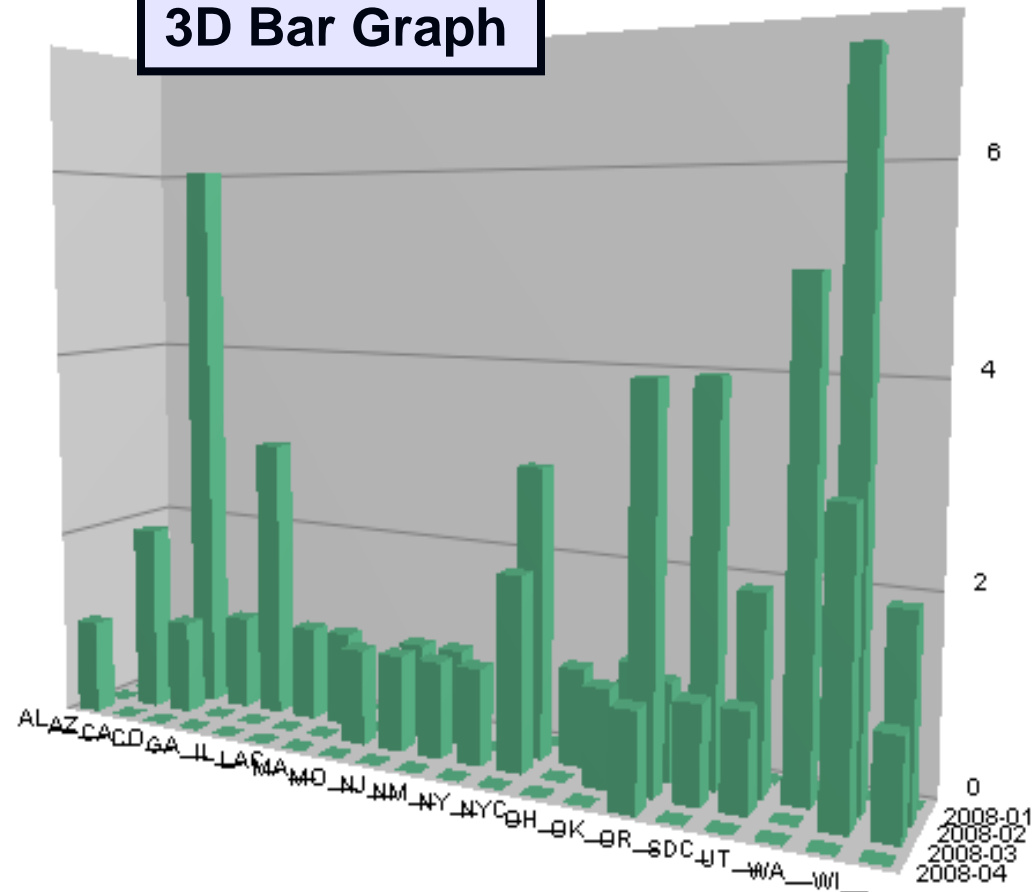


# Chart & Statistics: Changing graph types

2 components Lab ID and upload date



3D Bar Graph

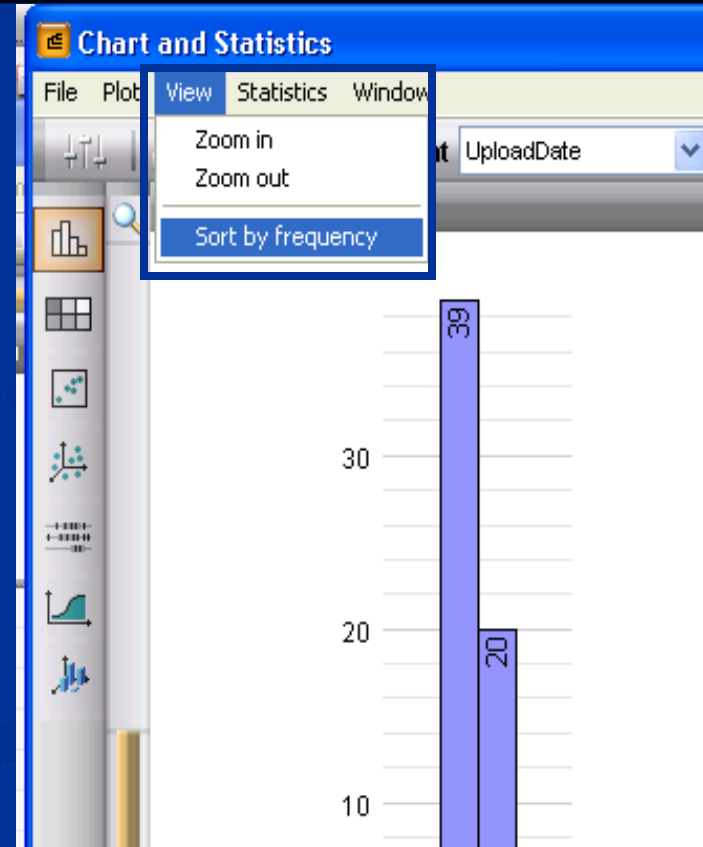
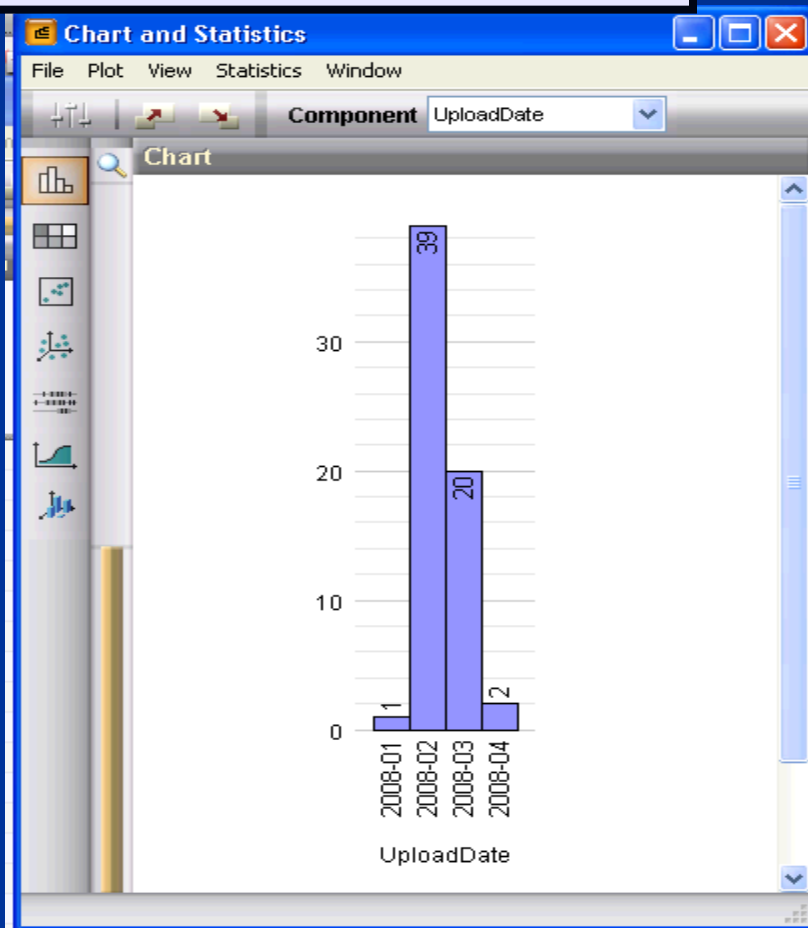


# Chart & Statistics:

## Changing graph type (bar graph)

Used upload date as variable (component)

You can also choose to sort by frequency using “view” in the tool bar





# Overview

- Utility of Groups
- Printing Groups
- Customizing Colors
- Using the Chart and Statistics Tool
- Types of Graphs
- Utility of Graphs in Cluster Detection and Reporting

# Utility of Graphs in Cluster Detection and Reporting

- You recently saw a noticeable amount of *Salmonella* Berta come into the lab. You want to determine if Berta is seasonal
  - Step 1: decide on the duration for which you want to do your seasonal tracking (2 years)
  - Step 2: select those entries in your database
  - Step 3: use the chart and statistics tool and choose a database component (isolate, received or upload date), check date variable box and then choose monthly intervals

# Utility of Graphs in Cluster Detection and Reporting

Select plot components

All components:

- SourceCity
- SourceSite
- SourceType
- TypeDetails
- PatientAge
- PatientSex
- IsolatDate
- ReceivedDate
- UploadDate
- AntigenForm
- OtherStatesIsolate
- cdc\_id
- Traveled\_To
- NARMS-EB
- Phagetype
- Exposure
- Comment
- AdditionalComment
- UploadModifiedDate
- Subspecies

Used components:

UploadDate

Data type

Categorical variable

Quantitative variable

Date variable

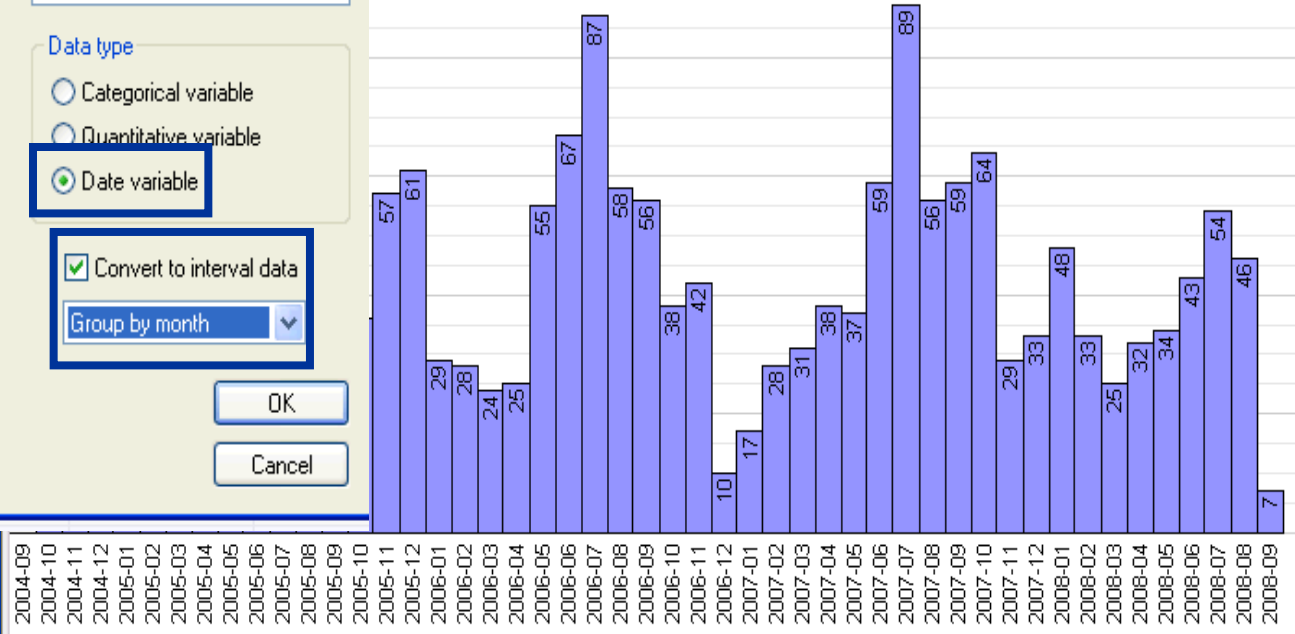
Convert to interval data

Group by month

OK

Cancel

**Seasonal Peaks**

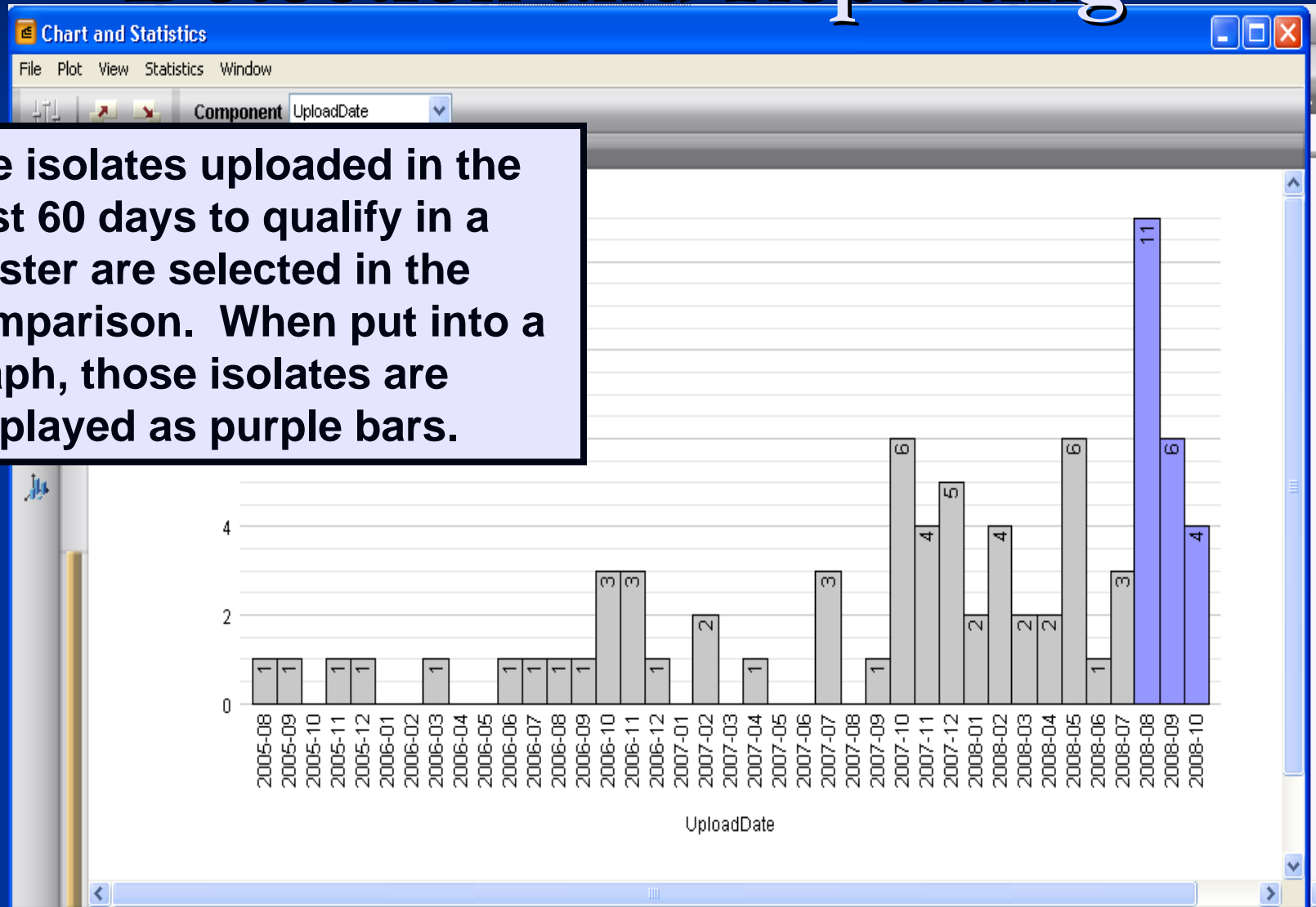


# Utility of Graphs in Cluster Detection and Reporting

- You are trying to decide if there is an increase in a certain pattern of *Salmonella* to post a cluster
  - Step 1: go to your main screen and search for all isolates assigned the pattern of interest and create a comparison
  - Step 2: select all current isolates in your comparison
  - Step 3: select charts and statistics, create a bar graph by upload data and select “date variable” by month

# Utility of Graphs in Cluster Detection and Reporting

The isolates uploaded in the past 60 days to qualify in a cluster are selected in the comparison. When put into a graph, those isolates are displayed as purple bars.



# Utility of Graphs in Cluster Detection and Reporting

- Your lab has seen a fair amount of *Salmonella* Enteritidis in the past few weeks and you want to see if this is an increase for a specific pattern
- You discern that your pattern is JEGX01.0005 based on your local database, and from downloading the CDC pattern name

**Search your database for pattern JEGX01.0005**

Entry search

Simple query | Advanced query tool

Key:

LabID:

Serotype:

PFGE-Xbal-pattern: \*JEGX01.0005\*

PFGE-Xbal-file:

PFGE-Xbal-rundate:

PFGE-Xbal-status:

PFGE-Blnl-pattern:

PFGE-Blnl

PFGE-Spel

PFGE-Xbal

antibio

biochem

Search in list

Negative search

Case sensitive

Clear Search Cancel

# Utility of Graphs in Cluster Detection and Reporting

**BioNumerics**

File Edit Database Subsets Experiments Comparison Identification PulseNet Scripts Window

Complete view

Chart & Statistics tool

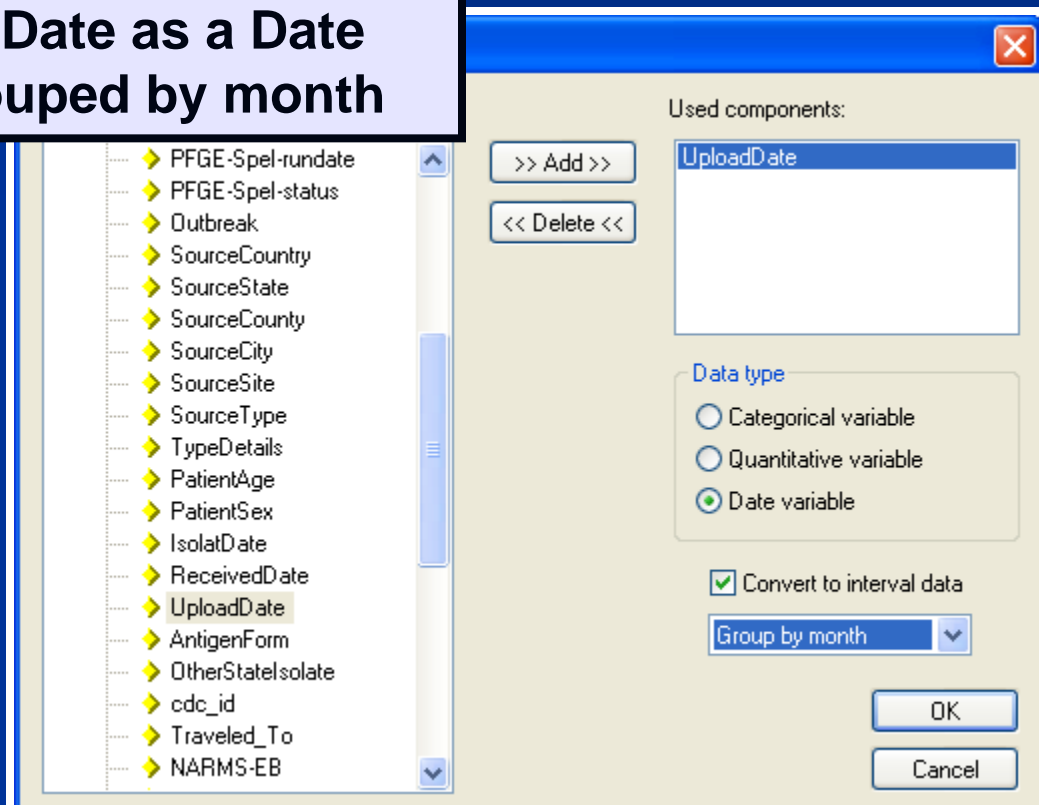
Database entries

Index	Key	LabID	Serotype	PFGE-Xbal-patt	PFGE-Xbal-file
7	T			JEGX01	TN05040,8
8	T			JEGX01.0004	TN05051,9
9	T			JEGX01.0004	TN05064,8
10	T			JEGX01.0004	TN05072,4
11	T			JEGX01.0004	TN05072,11
12	T			JEGX01.0004	TN05081,8

Click on the Chart & Statistics tool

# Utility of Graphs in Cluster Detection and Reporting

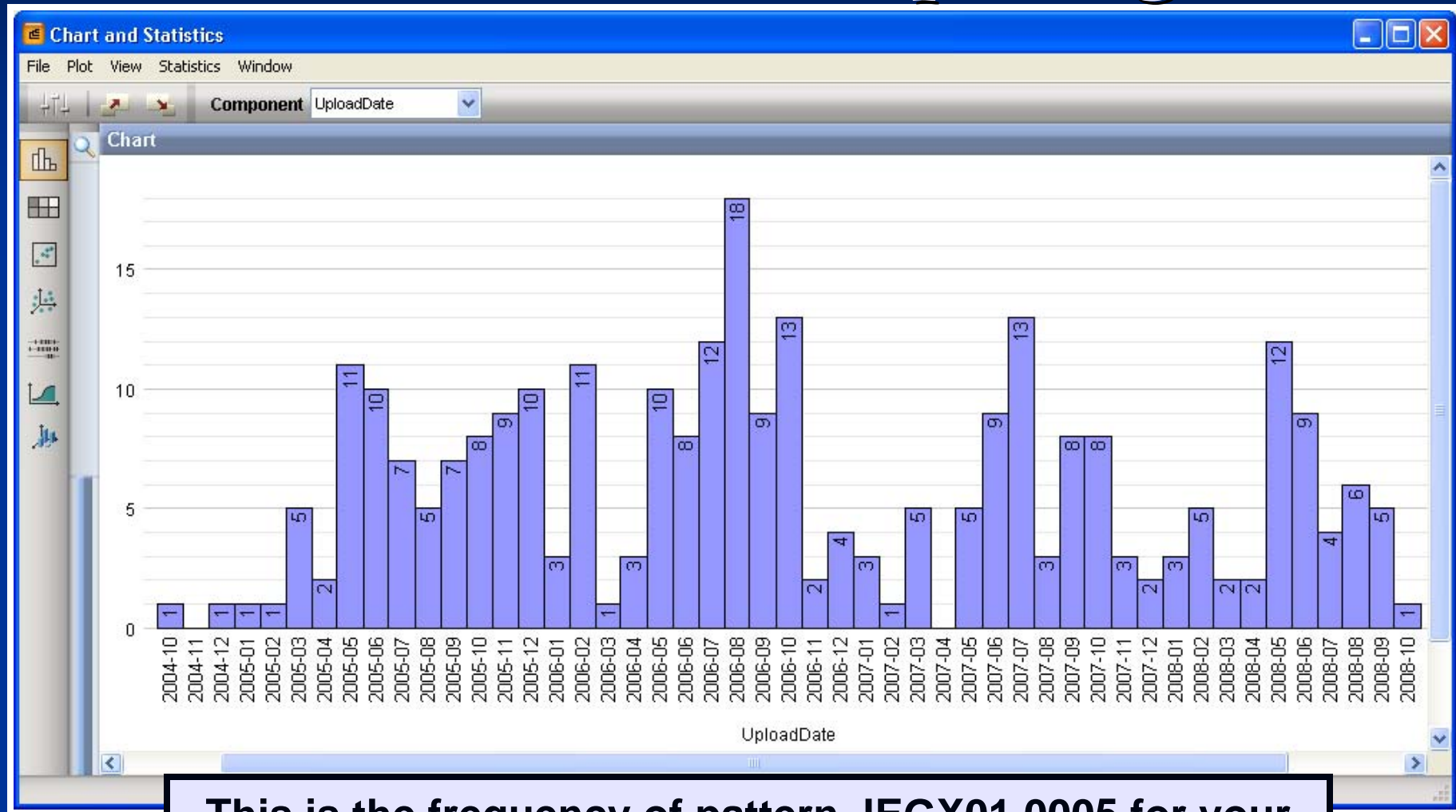
**Add UploadDate as a Date variable grouped by month**



**Grouping by month gives a good visual for frequencies over longer periods of time. You can use day or week for shorter periods.**



# Utility of Graphs in Cluster Detection and Reporting



**This is the frequency of pattern JEGX01.0005 for your database...is this an increase in the past 60 days?**

# 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, Key, abID, Serotype, and PFGE-XbaI-patt... The table contains 20 rows of data. A yellow callout box points to the 'Unselect all entries (F4)' button in the toolbar. A second yellow callout box points to the 'Query recent uploads' button in the bottom toolbar.

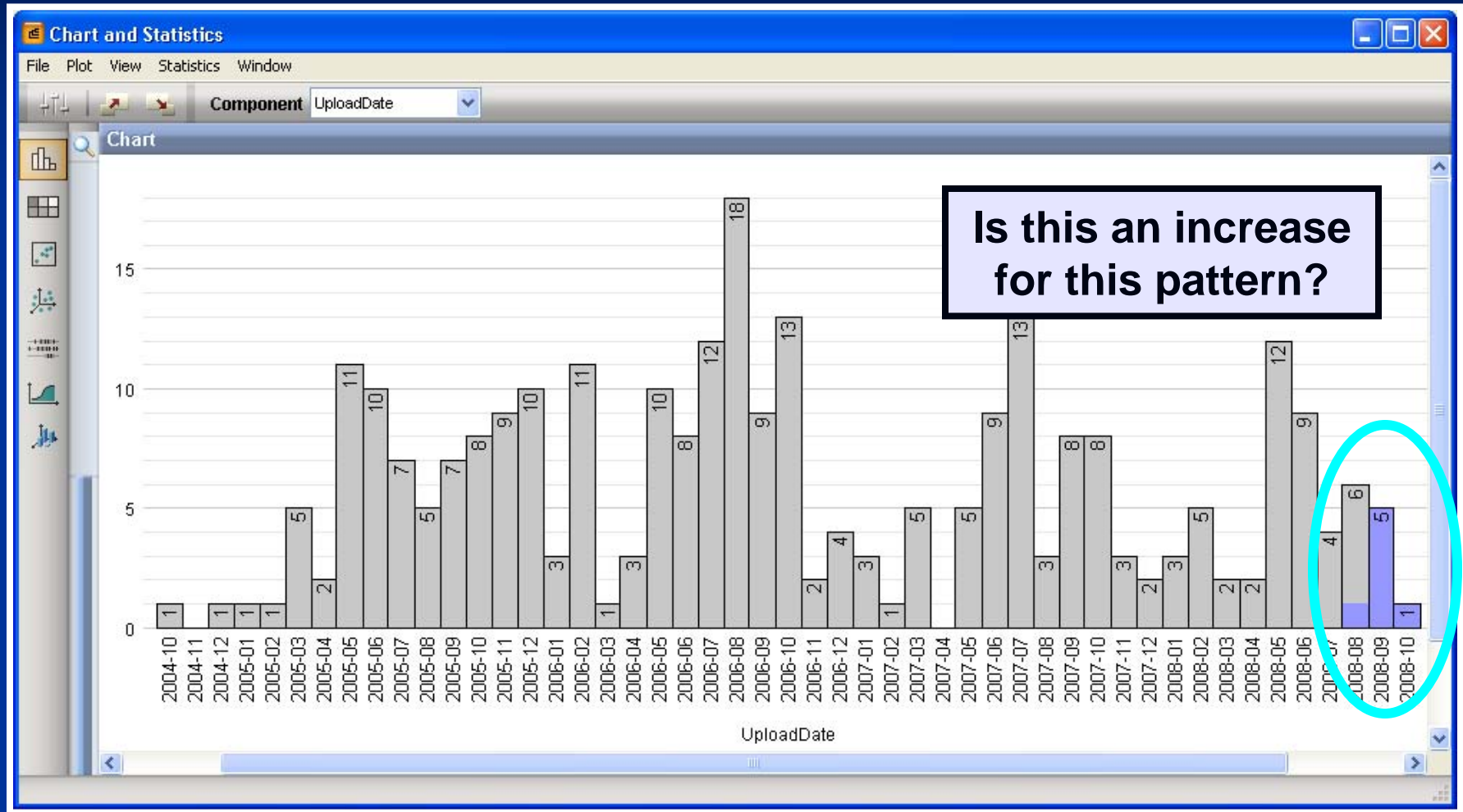
Index	Key	abID	Serotype	PFGE-XbaI-patt...	P
7	TN_043793		Enteritidis	JEGX01.0004	T
8			Enteritidis	JEGX01.0004	T
9			Enteritidis	JEGX01.0004	T
10			Enteritidis	JEGX01.0004	T
11			Enteritidis	JEGX01.0004	T
12			Enteritidis	JEGX01.0004	T
13			Enteritidis	JEGX01.0004	T
14	TN_045394		Enteritidis	JEGX01.0004	T
15	TN_050256		Enteritidis	JEGX01.0004	T
16	TN_050354		Enteritidis	JEGX01.0004	T
17	TN_050422		Enteritidis	JEGX01.0004	T
18	TN_050938		Enteritidis	JEGX01.0004	T
19	TN_050982		Enteritidis	JEGX01.0004	T
20	TN_051256		Enteritidis	JEGX01.0004	T

1. Deselect all entries *without* closing your graph

2. Perform a Hot List search of the past 60 days

A dialog box titled "Enter number of days before now" with a text input field containing the value "60". The dialog has "OK" and "Cancel" buttons.

# Utility of Graphs in Cluster Detection and Reporting



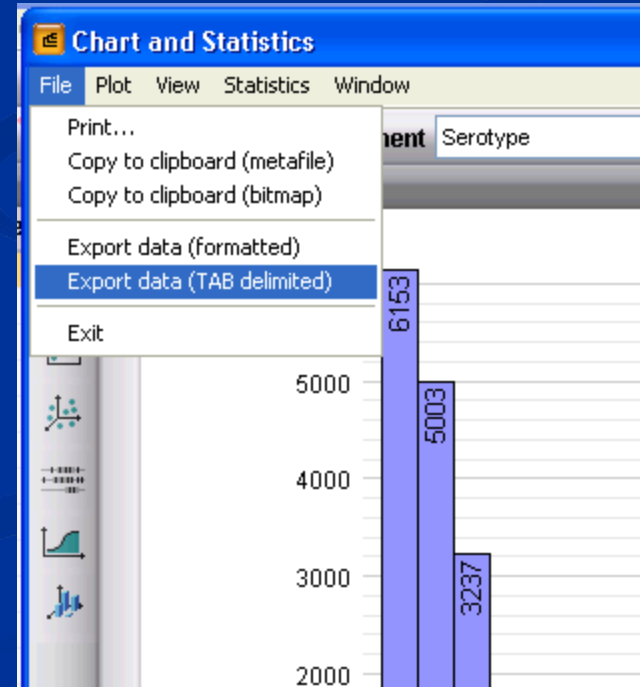
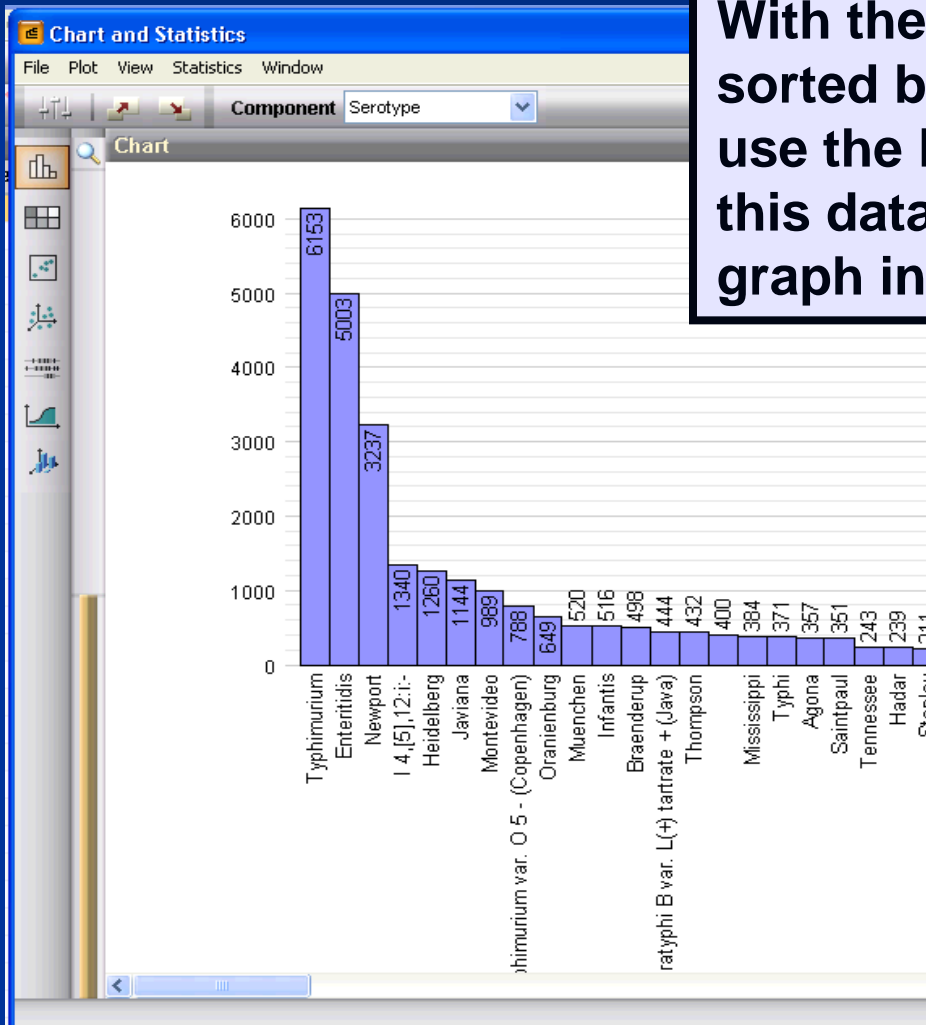
Open your graph to see the past 60 days selected (in blue, all others grayed out)

# Utility of Graphs in Cluster Detection and Reporting

- Your supervisor has asked how much of each *Salmonella* Serotype was analyzed in 2006
  - Step 1: go to your main screen and search for 2006 isolates and select the chart and statistics tool
  - Step 2: select “serotype” as the database component
  - Step 3: select “categorical variable”

# Utility of Graphs in Cluster Detection and Reporting

With the serotype information sorted by frequency, you can use the File menu to export this data into excel or copy graph into power point...



# Summary

- Utility of Groups
- Printing Groups
- Customizing Colors
- Using the Chart and Statistics Tool
- Types of Graphs
- Utility of Graphs in Cluster Detection and Reporting

# 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