## Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

<table>
<thead>
<tr>
<th>Name of Data Matrix</th>
<th>Name of Variable Matrix</th>
<th>Name of Filename Matrix</th>
<th>Name of Totalcounts Matrix</th>
<th>Name of Samplenames Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndatat</td>
<td>exactmass</td>
<td>filenames</td>
<td>Select Totalcounts...</td>
<td>samplegroups</td>
</tr>
</tbody>
</table>

This tutorial contains navigation buttons that enable you to move throughout the tutorial.

Please use the navigation buttons and not the page up/page down or arrow keys to navigate through the tutorials.

- This is the 'Next' button. It takes you to the next frame or stop point.
- This is the 'Previous' button. It takes you to the previous frame or stop point.
- This is the 'Go to frame' button. It takes you to a specified frame.
- This is the 'Go to URL' button. It takes you to a website link.

Press the 'Next' button below to start this tutorial.
This tutorial covers how to use a function called the PC Data browser. The function allows the user to look at preprocessed data after subtracting previous PCs from a data set.

This function was created in order to look at the data that PCA is using when it calculates PCs after PC1. As illustrated in the figure on the right, each subsequent PC is calculated from the data matrix not captured by the previous PC. This means it is using the data matrix minus the previous PC.

It is always recommended that one looks back at the "original" data when interpreting PCA results. This can be easily done for PC1, where even looking back at the non-preprocessed data will show trends similar to those seen in PCA. However, for subsequent PCs one cannot look at the original data matrix and expect to see the trends shown in the PCA results.

In fact, one should really look at the preprocessed data used for a given PC when looking at trends in the PCA results.

The PC Data browser enables this functionality and provides a way for the user to look at peak area images of the preprocessed data for a given PC (preprocessed matrix - previous PCs).

In this tutorial I will provide examples of how this works and why it can be useful.
## Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

<table>
<thead>
<tr>
<th>Name of Data Matrix</th>
<th>Name of Variable Matrix</th>
<th>Name of Filename Matrix</th>
<th>Name of Totalcounts Matrix</th>
<th>Name of Samplenames Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndatassp</td>
<td>exactmass</td>
<td>filenames</td>
<td>Select Totalcounts</td>
<td>samplenames</td>
</tr>
</tbody>
</table>

---

This tutorial assumes that you have already loaded some data (see tutorial 02/03/04), ran PCA and saved the PCA results (see tutorial 07).
Select 'PC Data Browser' from the 'Data Display' menu.
Select the desired data from both the 'Data Selection Panel', and the 'MVA Data Selection Panel' above.
### Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

<table>
<thead>
<tr>
<th>Name of Data Matrix</th>
<th>Name of Variable Matrix</th>
<th>Name of Filename Matrix</th>
<th>Name of Totalcounts Matrix</th>
<th>Name of Samplenames Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndataset</td>
<td>exactmass</td>
<td>filenames</td>
<td>Select Totalcounts</td>
<td>sampleinames</td>
</tr>
</tbody>
</table>

### MVA Data Selection Panel

<table>
<thead>
<tr>
<th>Name of Scores Matrix</th>
<th>Name of Loadings Matrix</th>
<th>Name of % Variance Matrix</th>
<th>Name of Model Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>scores</td>
<td>loads</td>
<td>var</td>
<td>model</td>
</tr>
</tbody>
</table>

Choose score/loads/var above

- After selecting the desired data. Press the 'Continue' button.
Data Selection Panel
These are the main input data that will be used in further analysis unless you specify otherwise.
Use the drop down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ndatassp
Name of Variable Matrix: exactmass
Name of Filename Matrix: filenames
Name of Totalcounts Matrix: Select Totalcounts
Name of Samplnames Matrix: samplnames

Then press the 'Load Selected Data' button.

Data Preprocessing
Choose an option below
PC #
Choose one:
Plot Scores and Loads

Peak List
Select Plot Type
Average + Stdev Colored
Plot Peak Area Image
This plot will show the scaled for the selected peak after subtracting the PCs selected above.

Save Figure
Make Ext
Close Panel
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

- **Name of Data Matrix**: ndatassp
- **Name of Variable Matrix**: exactmass
- **Name of Filename Matrix**: filenames
- **Name of Totalcounts Matrix**: Select Totalcounts
- **Name of Samplenames Matrix**: samplenames

**Load Selected Data**
- **Image**: ndatassp
- **Variable**: exactmass
- **Scores**: scores
- **Loads**: loads

**Data Preprocessing**
Choose an option below:
- None
- Mean Center
- Autoscale
- Square root & mean center

**Plot Scores and Loads**

**Peak List**
- 285.2647 –[M-H]C18
- 292.885 –AuS3
- 333.2482 –C18H37SC
- 371.1106 –AuMC10
- 393.9333 –Au2
- 403.0822 –AuS2C10l
- 425.9052 –Au2S
- 457.8771 –Au2S2
- 483.2357 –AuMC18
- 489.8492 –Au2S3
- 515.2086 –Au[M-H]20
- 543.2399 –Au[M-H]2

**Select Plot Type**
- Average + Stdev Colored

**Plot Peak Area Image**

This plot will show the scaled for the selected peak after subtracting the PCs selected above.

Select the pre-processing method used on the data when you carried out PCA.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: 
Name of Variable Matrix: exactmass
Name of Filename Matrix: filenames
Name of Totalcounts Matrix: Select Totalcounts
Name of Samplenames Matrix: sample

Load Selected Data

Image: ndatassp
Variable: exactmass
Scores: scores
Loads: loads

Data Preprocessing

Squareroot & mean center

PC #
Choose one:
Choose one:
1
2

Peak List

285.2647–[M-H]C18
292.885–AuS3
333.2482–C18H37S
371.1106–AuMC10
393.9333–Au2
403.0822–AuS2C10t
425.9052–Au2S
457.8771–Au2S2
483.2357–AuMC18
489.8492–Au2S3
515.2086–Au[M-H]20
543.2399–Au[M-H]2

Select Plot Type

Average + Stdev Colored

Plot Peak Area Image

This plot will show the scaled for the selected peak after subtracting the PCs selected above.

Choose which PC you want to display the data for.

We will start with PC1.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ndatassp
Name of Variable Matrix: exactmass
Name of Filename Matrix: filenames
Name of Totalcounts Matrix: Select Totalcounts
Name of Samplenames Matrix: samplenames

Load Selected Data

Image: ndatassp
Variable: exactmass
Scores: scores
Loads: loads

Data Preprocessing

Squareroot & mean center
PC #
1

Plot Scores and Loads

Peak List
285.2647—[M-H]C18
292.885—AuS3
333.2482—C18H37SC
371.1106—AuMC10
393.9333—Au2
403.0822—AuS2C10H
425.9052—Au2S
457.8771—Au2S2
483.2357—AuMC18
489.8492—Au2S3
515.2086—Au[M-H]20
543.2399—Au[M-H]2

Select Plot Type
Average + Stdev Colored

Plot Peak Area Image

This plot will show the scaled for the selected peak after subtracting the PCs selected above.

Once you have chosen the options you want, press the 'Plot Scores and Loads' button.
The scores and loadings for the chosen PC are displayed.

Here we can see that the peak at m/z 543 has a high positive loading on PC1.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

- **Data Matrix**: ndatassp
- **Variable Matrix**: exactmass
- **Filename Matrix**: filenames
- **Totalcounts Matrix**: Select Totalcounts
- **Sample Names Matrix**: samenames

![Data Selection Panel](image)

**Load Selected Data**
- **Image**: ndatassp
- **Variable**: exactmass
- **Scores**: scores
- **Loads**: loads

**Data Preprocessing**
- Square root & mean center
- **PC #**: 1

**Plot Scores and Loads**

**Peak List**
- 333.2482-C18H37S
- 371.1106-AuMC10
- 393.9333-Au2
- 403.0822-AuS2C10t
- 425.9052-Au2S
- 457.8771-Au2S2
- 483.2357-AuMC18
- 489.8492-Au2S3
- 515.2086-Au[M-H]2
- **543.2399-Au[M-H]2**
- 567.0683-Au2[M-H]2
- 590.8999-Au3

**Select Plot Type**
- Average + Stdev Colored

**Plot Peak Area Image**
- This plot will show the scaled for the selected peak after subtracting the PCs selected above.

Let's select the m/z 543 peak.
Choose the type of peak area plot you want to use. Here we choose 'Average + Stdev Colored'.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ndatassp
Name of Variable Matrix: exactmass
Name of Filename Matrix: filenames
Name of Totalcounts Matrix: Select Totalcounts
Name of Samplenames Matrix: samplenames

Load Selected Data
Image: ndatassp
Variable: exactmass
Scores: scores
Loads: loads

Data Preprocessing
Squareroot & mean center
PC #

Plot Scores and Loads

Peak List
333.2482--C18H37S(△)
371.1106--AuMC10
393.9333--Au2
403.0822--AuS2C10△
425.9052--Au2S
457.8771--Au2S2
483.2357--AuMC18
489.8492--Au2S3
515.2086--Au[M-H]2△
543.2399--Au[M-H]2△
567.0683--Au2[M-H]1△
590.8999--Au3

Select Plot Type
Average + Stdev Colored

Plot Peak Area Image
This plot will show the scaled for the selected peak after subtracting the PCs selected above.

Then press the 'Plot Peak Area Data' button.

Save Figure
Make Ext
Close Panel
Here we see the preprocessed peak area data for the m/z 543 peak. It looks very similar to the PC1 scores image as we would expect since this peak has a high positive loading on this PC and therefore we would expect it to have a higher relative intensity in areas with positive scores on PC1.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

<table>
<thead>
<tr>
<th>Name of Data Matrix</th>
<th>Name of Variable Matrix</th>
<th>Name of Filename Matrix</th>
<th>Name of Totalcounts Matrix</th>
<th>Name of Samplenames Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndatassp</td>
<td>exactmass</td>
<td>filenames</td>
<td>Select Totalcounts</td>
<td>samplenames</td>
</tr>
</tbody>
</table>

**Load Selected Data**

**Image:** ndatassp  
**Variable:** exactmass  
**Scores:** scores  
**Loads:** loads

**Data Preprocessing**

Squareroot & mean center

**PC #**

Choose one:  
1  
2

**Peak List**

333.2482—C18H37S  
371.1106—AuMC10  
393.9333—Au2  
403.0822—AuS2C10I  
425.9052—Au2S  
457.8771—Au2S2  
483.2357—AuMC18  
489.8492—Au2S3  
515.2086—Au[M-H]2+  
543.2399—Au[M-H]2+  
567.0683—Au2[M-H]2+  
590.8999—Au3

**Select Plot Type**

Average + Stdev Colored

**Plot Peak Area Image**

This plot will show the scaled for the selected peak after subtracting the PCs selected above.

Now let's look at PC2 by selecting it from the drop down menu.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

Name of Data Matrix: ndatassp
Name of Variable Matrix: exactmass
Name of Filename Matrix: filenames
Name of Totalcounts Matrix: Select Totalcounts
Name of Samplenames Matrix: sample names

Load Selected Data
Image: ndatassp
Variable: exactmass
Scores: scores
Loads: loads

Data Preprocessing
Squareroot & mean center
PC #

Plot Scores and Loads
Peak List

Select Plot Type
Average + Stdev Colored

Plot Peak Area Image
This plot will show the scaled for the selected peak after subtracting the PCs selected above.

And pressing the 'Plot Scores and Loads' button.

Save Figure
Make Ext
Close Panel
The PC2 scores show a different pattern than PC1. The m/z 543 peak shows a positive loading on PC2, but it no longer has the highest loading.
Let's select the m/z 543 peak and press the 'Plot Peak Area Image' button to look at the original pre-processed peak area data for PC2 for that peak.
Here we see the preprocessed peak area data for the m/z 543 peak after subtracting PC1 from the data set. The general pattern is similar to that seen in the PC2 scores, and it is very different from that seen before subtracting PC1 from the data set.

This plot makes more sense than the plot from the preprocessed data before subtracting PC1 for understanding what is happening in PC2. Before subtracting PC1 the m/z 543 peak area data showed an increase with increasing sample number. For PC2 the data shows a decrease and then an increase exactly like is seen in the scores plot.
Here I have gone back and plotted another peak for PC1. In this case I chose the m/z 679 peak which shows a high negative loading.

As expected the data for the m/z 679 peak shows an inverse trend as seen in the PC1 scores (since it has a negative loading).
Here I plotted the m/z 679 pre-processed peak area data for PC2. On PC2 this peak also shows a negative loading.

The m/z 679 peak shows an inverse trend as seen in the PC2 scores (since it has a negative loading). The trend makes sense based off the PC2 scores.
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

- **Image**: ndataspp
- **Variable**: exactmass
- **Scores**: scores
- **Loads**: loads

Data Preprocessing

Squareroot & mean center

PC #

Plot Scores and Loads

Peak List

- 489.8492--Au2S3
- 515.2086--Au[M-H]2C
- 543.2399--Au[M-H]2C
- 567.0683--Au2[M-H]C
- 590.8999--Au3
- 622.869--Au3S
- 654.8471--Au3S2O2C
- 655.3643--Au[M-H]2C
- 679.196--Au2[M-H]C
- 686.8168--unknown
- 718.789--unknown
- 767.498--Au[M-H]2C

Select Plot Type

- Average + Stdev Colored

Plot Peak Area Image

This plot will show the scaled for the selected peak after subtracting the PCs selected above.

- m/z = 679.196

Any of the generated peak area images can be saved by using the ‘Save Figure’ button.

- Relative Intensity

- Sample number

Save Figure

Make Ext

Close Panel
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

- **Name of Data Matrix**: ndatassp
- **Name of Variable Matrix**: exactmass
- **Name of Filename Matrix**: filenames
- **Name of Totalcounts Matrix**: Select Totalcounts
- **Name of Samplenames Matrix**: samenames

**Load Selected Data**

- **Image**: ndatassp
- **Variable**: exactmass
- **Scores**: scores
- **Loads**: loads

**Data Preprocessing**

- **Squareroot & mean center**
- **PC #**: 2

**Plot Scores and Loads**

**Peak List**

- 489.8492–Au2S3
- 515.2086–Au[M-H]2C
- 543.2399–Au[M-H]2C
- 567.0683–Au2[M-H]C
- 590.8999–Au3
- 622.869–Au3S
- 654.8471–Au3S2O2l
- 655.3643–Au[M-H]2C
- 679.196–Au2[M-H]C
- 686.8168–unknown
- 718.789–unknown
- 767.498–Au[M-H]2C

**Select Plot Type**

- **Average + Stdev Colored**

**Plot Peak Area Image**

This plot will show the scaled for the selected peak after subtracting the PCs selected above.

**m/z = 679.196**

**Relative Intensity**

- 0.03
- 0.02
- 0.01
- 0.00
- 0.0

**Sample Number**

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

You can also make an external Matlab figure of any generated peak area image by pressing the 'Make Ext' button.

**Save Figure**

**Make Ext**

**Close Panel**
Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop-down menus to select the data and information you want to use in your analysis.

<table>
<thead>
<tr>
<th>Name of Data Matrix</th>
<th>Name of Variable Matrix</th>
<th>Name of Filename Matrix</th>
<th>Name of Totalcounts Matrix</th>
<th>Name of Samplenames Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndatassp</td>
<td>exactmass</td>
<td>filenames</td>
<td>Select Totalcounts</td>
<td>samplenames</td>
</tr>
</tbody>
</table>

Load Selected Data

- **Image:** ndatassp
- **Variable:** exactmass
- **Scores:** scores
- **Loads:** loads

Data Preprocessing

- Squareroot & mean center
- PC #

Plot Scores and Loads

Peak List

- 489.8492 -- Au2S3
- 515.2086 -- Au[M-H]2C
- 543.2399 -- Au[M-H]2C
- 567.0683 -- Au2[M-H]C
- 590.8999 -- Au3
- 622.8699 -- Au3S
- 654.8471 -- Au3S2O2l
- 655.3643 -- Au[M-H]2C
- 679.196 -- Au2[M-H]C
- 686.8168 -- unknown
- 718.789 -- unknown
- 767.498 -- Au[M-H]2C

Select Plot Type

- Average + Stdev Colored

Plot Peak Area Image

This plot will show the scaled for the selected peak after subtracting the PCs selected above.

You can close the panel by pressing the ‘Close Panel’ button.
## Data Selection Panel

These are the main input data that will be used in further analysis unless you specify otherwise. Use the drop down menus to select the data and information you want to use in your analysis.

<table>
<thead>
<tr>
<th>Name of Data Matrix</th>
<th>Name of Variable Matrix</th>
<th>Name of Filename Matrix</th>
<th>Name of Totalcounts Matrix</th>
<th>Name of Samplenames Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndatassp</td>
<td>exactmass</td>
<td>filenames</td>
<td>Select Totalcounts</td>
<td>samplenames</td>
</tr>
</tbody>
</table>

That ends this tutorial. Press the button on the left to go back to the previous step. Press the button on the right to start the tutorial over.