

LC-MS/MS data handling with **xcms**

Johannes Rainer

Eurac Research, Bolzano, Italy

johannes.rainer@eurac.edu - github: jorainer - twitter: jo_rainer

SWEMSA 2019 | <https://doi.org/10.5281/zenodo.3499650>

LC-MS data analysis with `xcms`

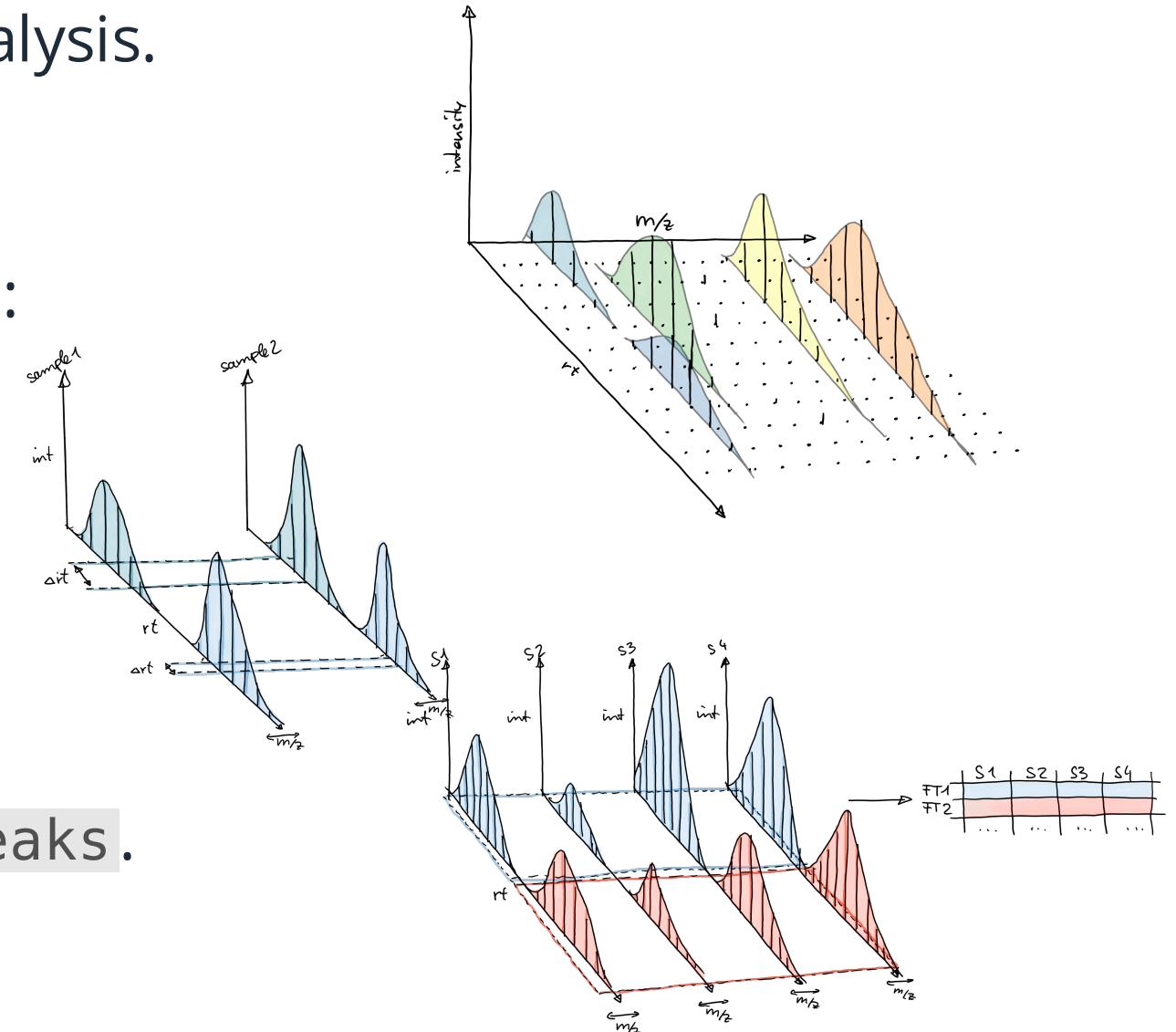
- `xcms`: toolbox for LC-MS data analysis.

- Chromatographic peak detection:
`findChromPeaks`.

- Alignment: `adjustRtime`.

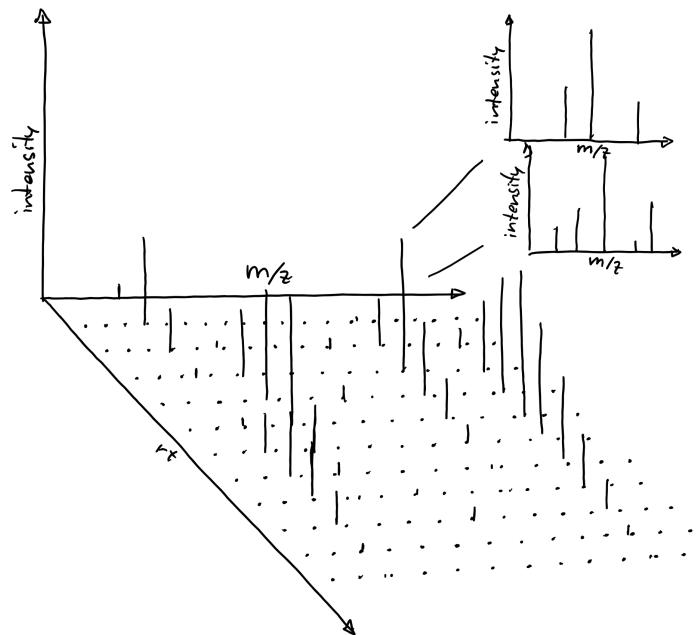
- Correspondence: `groupChromPeaks`.

- **Don't use defaults!**
see [Metabolomics data pre-processing using xcms workshop](#).



- Result: `matrix` with feature abundances in samples.
- Annotation of features from LC-MS experiments challenging.
- LC-MS/MS data: MS2 spectra assist in annotation.
- Added support for LC-MS/MS data analysis in `xcms`.
- Possible because we re-use functionality and classes from the `MSnbase` package.

Analyzing DDA data with **xcms**



```
dda_file <- system.file("TripleTOF-SWATH/PestMix1_DDA.mzML",
                        package = "msdata")
dda_data <- readMSData(dda_file, mode = "onDisk")
table(msLevel(dda_data))
```

```
##
##    1    2
## 4627 2975
```

Analyzing DDA data with **xcms**

- MS1 chromatographic peak detection:

```
cwp <- CentWaveParam(snthresh = 5, noise = 100, ppm = 10,  
                      peakwidth = c(3, 30))  
dda_data <- findChromPeaks(dda_data, param = cwp)
```

```
dda_spectra <- chromPeakSpectra(dda_data)  
dda_spectra
```

```
## Spectra with 158 spectra and 1 metadata column(s):  
##          msLevel      rtime peaksCount |    peak_id  
##          <integer> <numeric> <integer> | <character>  
## CP01.F1.S1000        2   128.237       16 |    CP01  
## CP01.F1.S1008        2   128.737       40 |    CP01  
## ...                 ...     ...       ... |    ...  
## CP98.F1.S5266        2   596.054       88 |    CP98  
## CP99.F1.S7344        2   873.714       20 |    CP99
```

Analyzing DDA data with **xcms**

- Example: annotate chrom peak with an m/z of 304.1131.

```
chromPeaks(dda_data, mz = 304.1131, ppm = 20)
```

```
##          mz      mzmin      mzmax       rt     rtmin     rtmax     into     intb
## CP53 304.1133 304.1126 304.1143 424.614 417.985 430.784 13709.7 13658.01
##          maxo sn sample
## CP53 3978.987 74      1
```

- Get MS2 spectra associated with that peak

```
ex_spectra <- dda_spectra[mcols(dda_spectra)$peak_id == "CP53"]
```

Analyzing DDA data with **xcms**

```
## Spectra with 5 spectra and 1 metadata column(s):
##          msLevel      rtime peaksCount |    peak_id
##          <integer> <numeric> <integer> | <character>
## CP53.F1.S3505      2   418.926      10 |    CP53
## CP53.F1.S3510      2   419.306      30 |    CP53
## CP53.F1.S3582      2   423.036     694 |    CP53
## CP53.F1.S3603      2   423.966     783 |    CP53
## CP53.F1.S3609      2   424.296     753 |    CP53
```

- Build *consensus* spectrum.

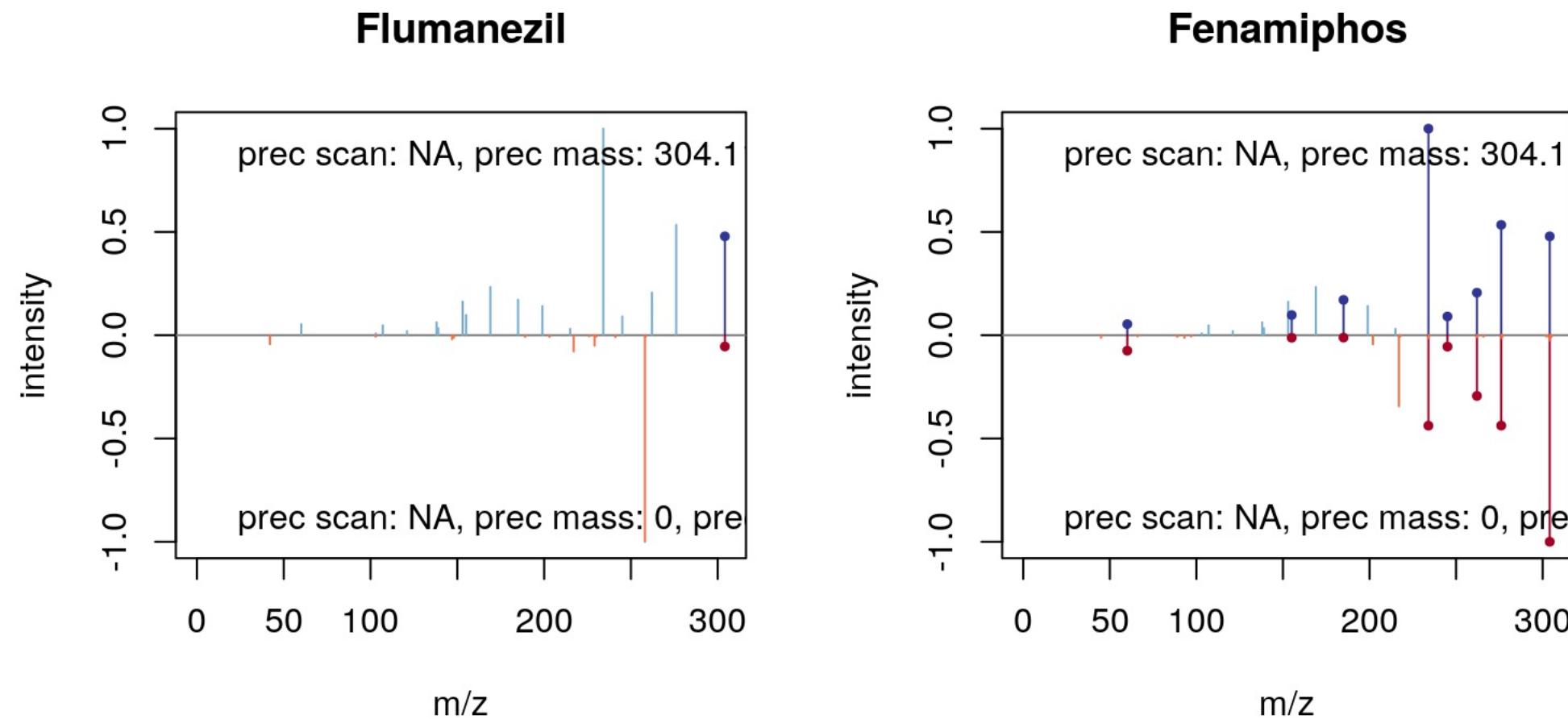
```
ex_spectrum <- combineSpectra(ex_spectra, method = consensusSpectrum,
                                ppm = 10, minProp = 0.8)
ex_spectrum
```

```
## Spectra with 1 spectra and 1 metadata column(s):
##          msLevel      rtime peaksCount |    peak_id
##          <integer> <numeric> <integer> | <character>
## CP53.F1.S3505      2   418.926      17 |    CP53
```

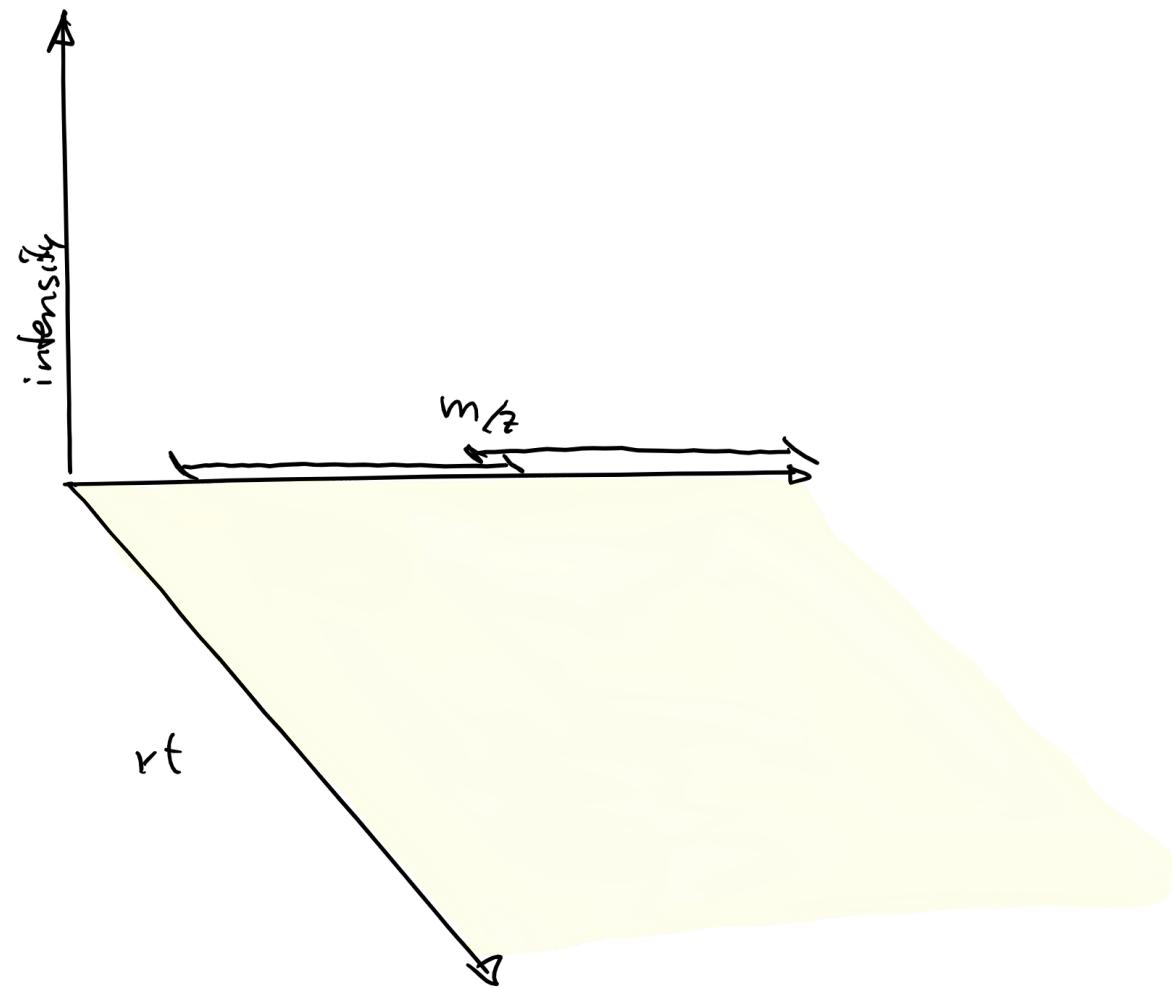
Analyzing DDA data with **xcms**

- Compare the consensus spectrum against 2 candidates with same m/z.

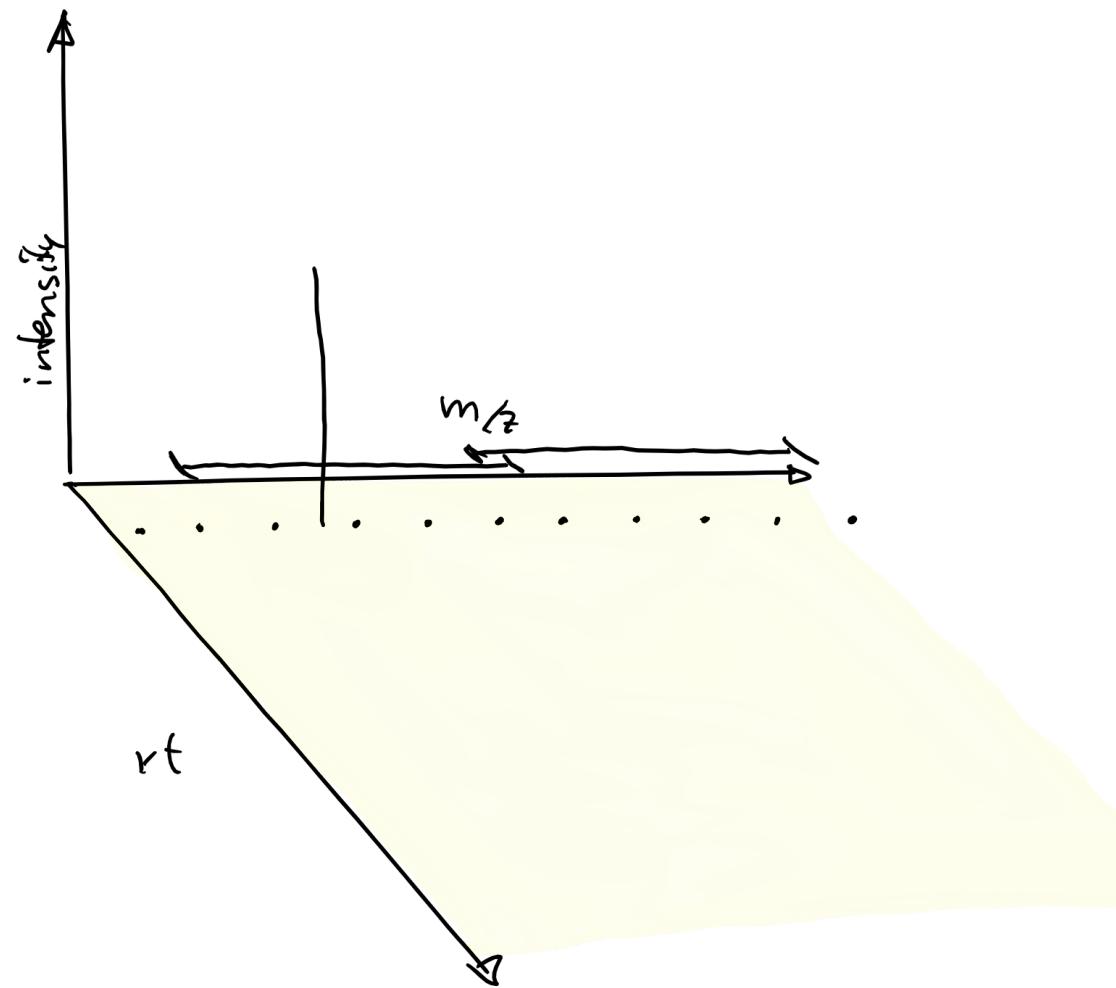
```
par(mfrow = c(1, 2))
plot(ex_spectrum, flumanezil, main = "Flumanezil", tolerance = 40e-6)
plot(ex_spectrum, fenamiphos, main = "Fenamiphos", tolerance = 40e-6)
```



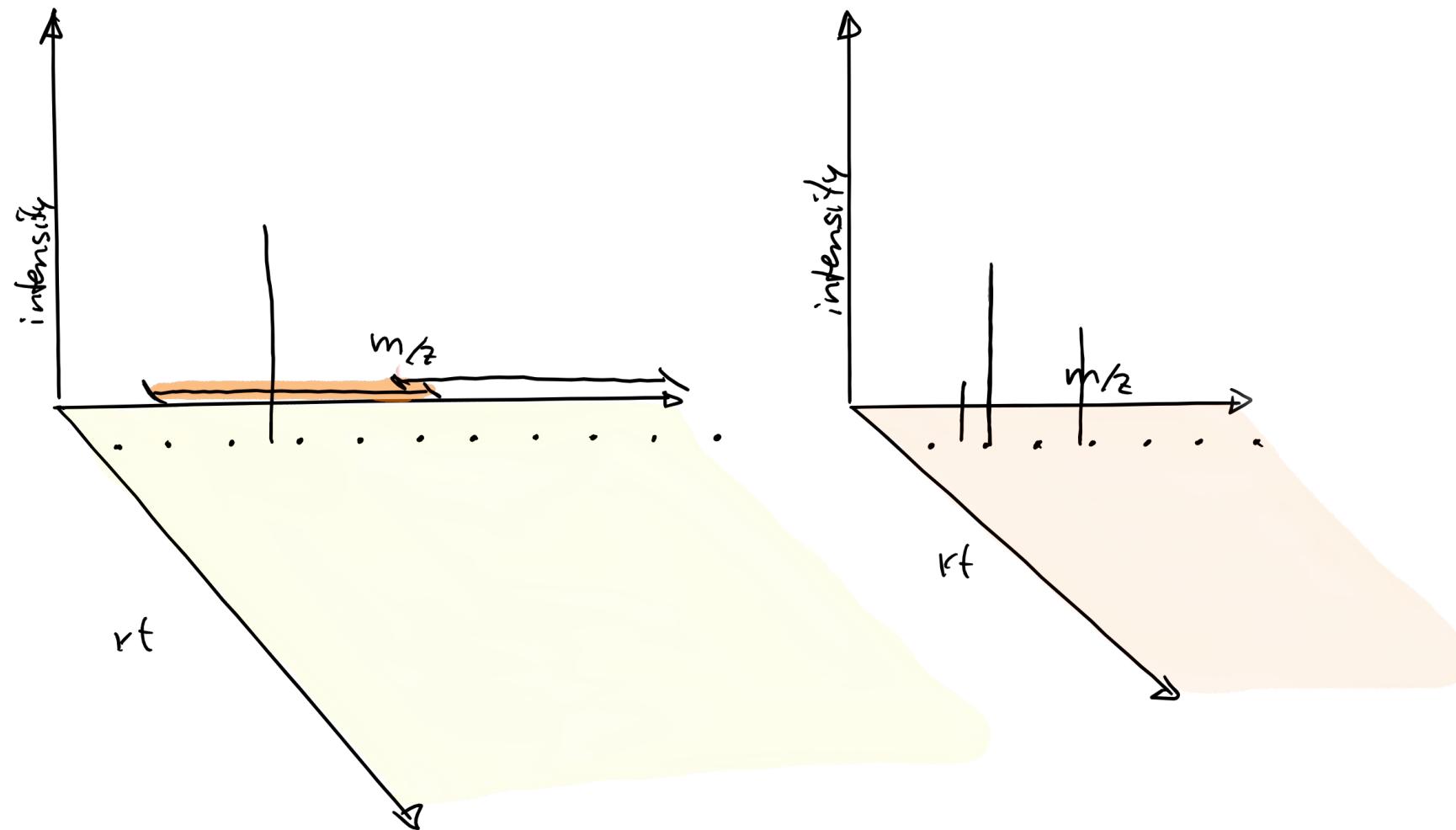
SWATH data



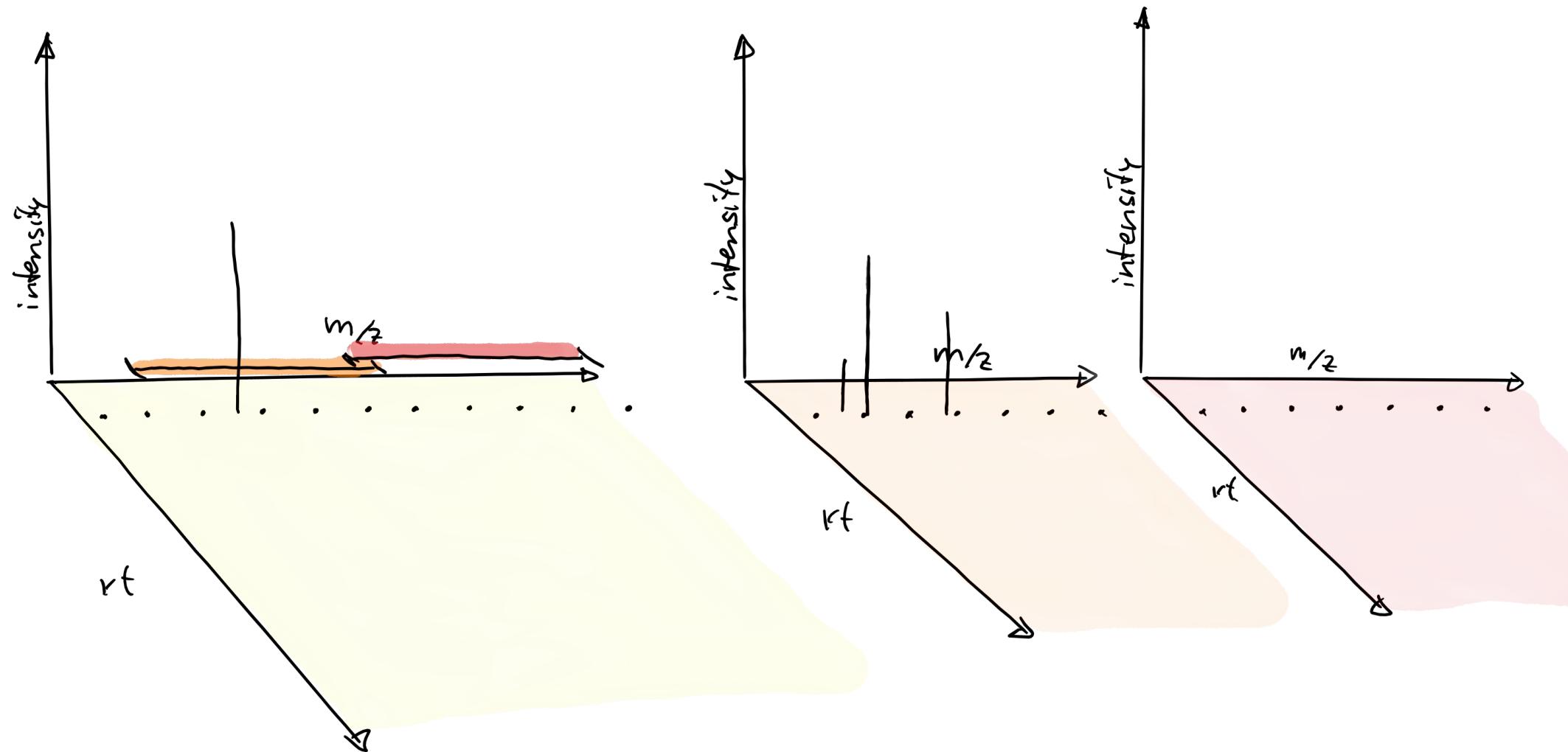
SWATH data



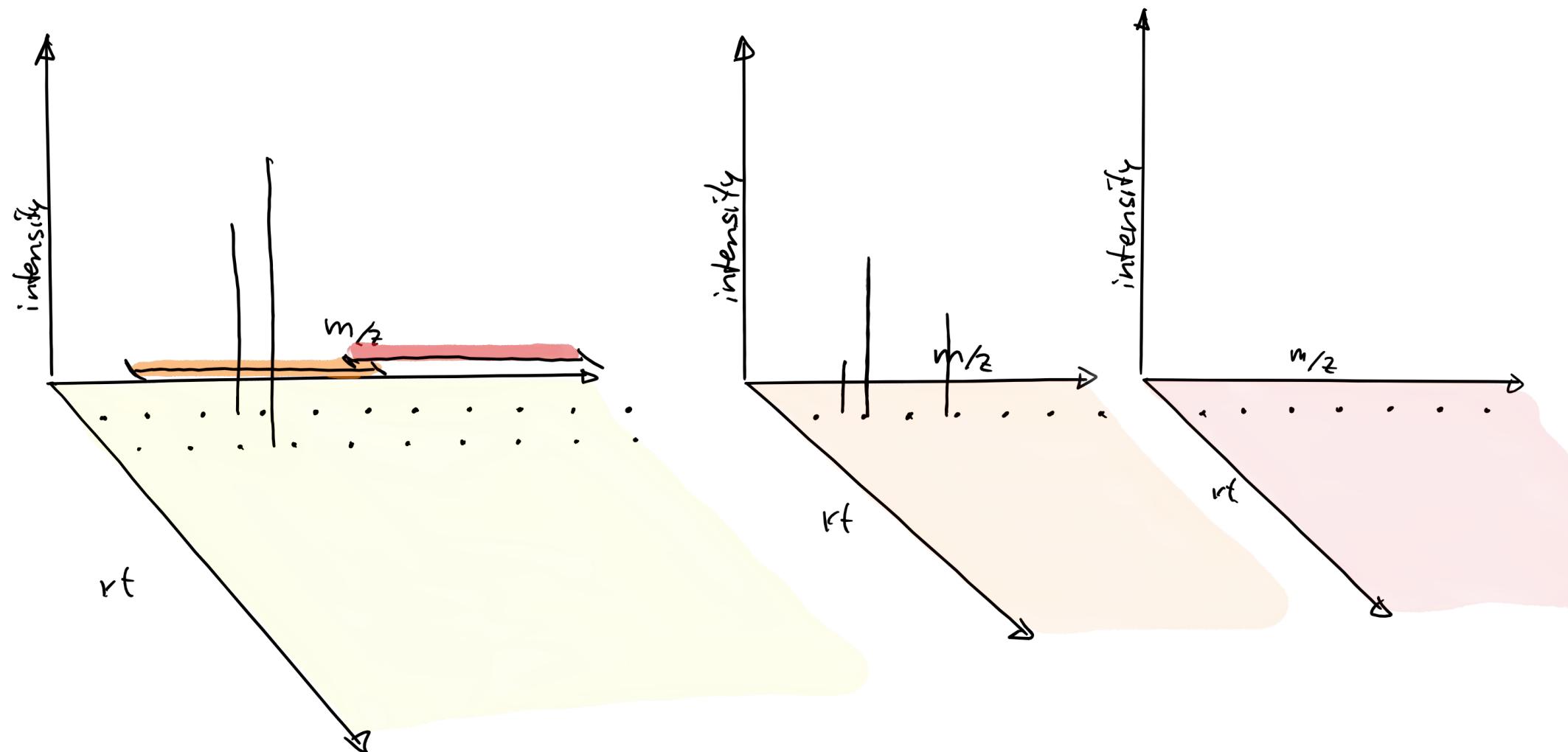
SWATH data



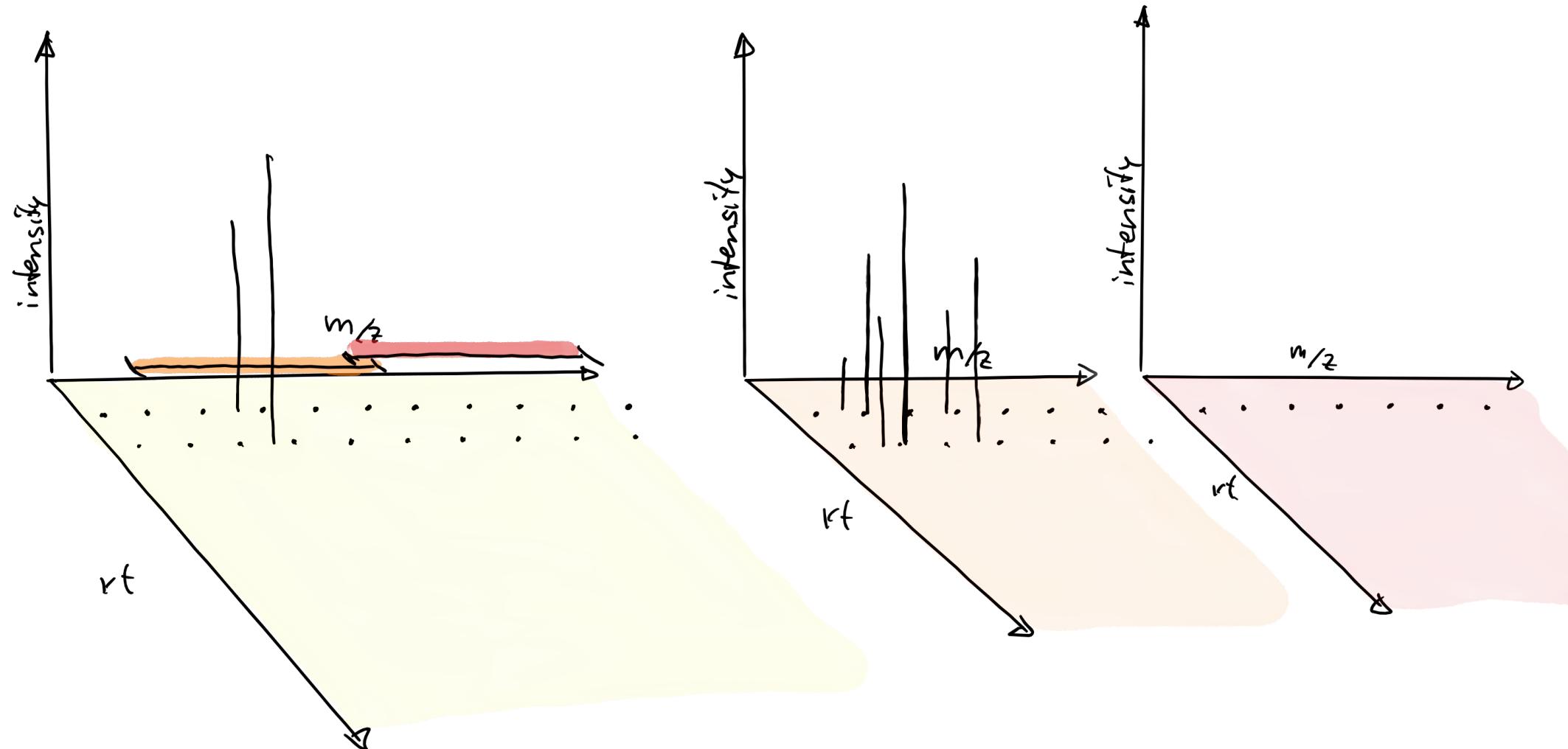
SWATH data



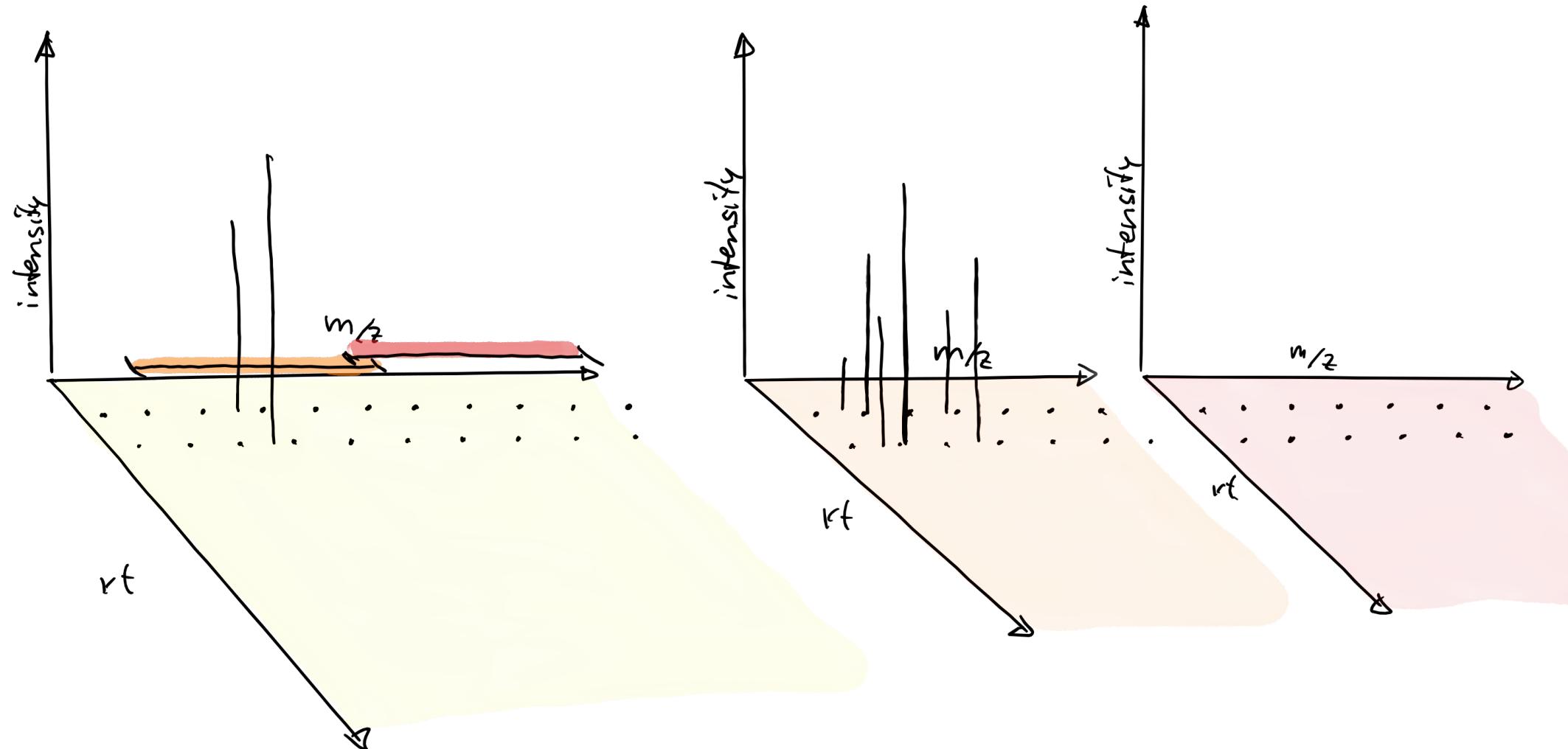
SWATH data



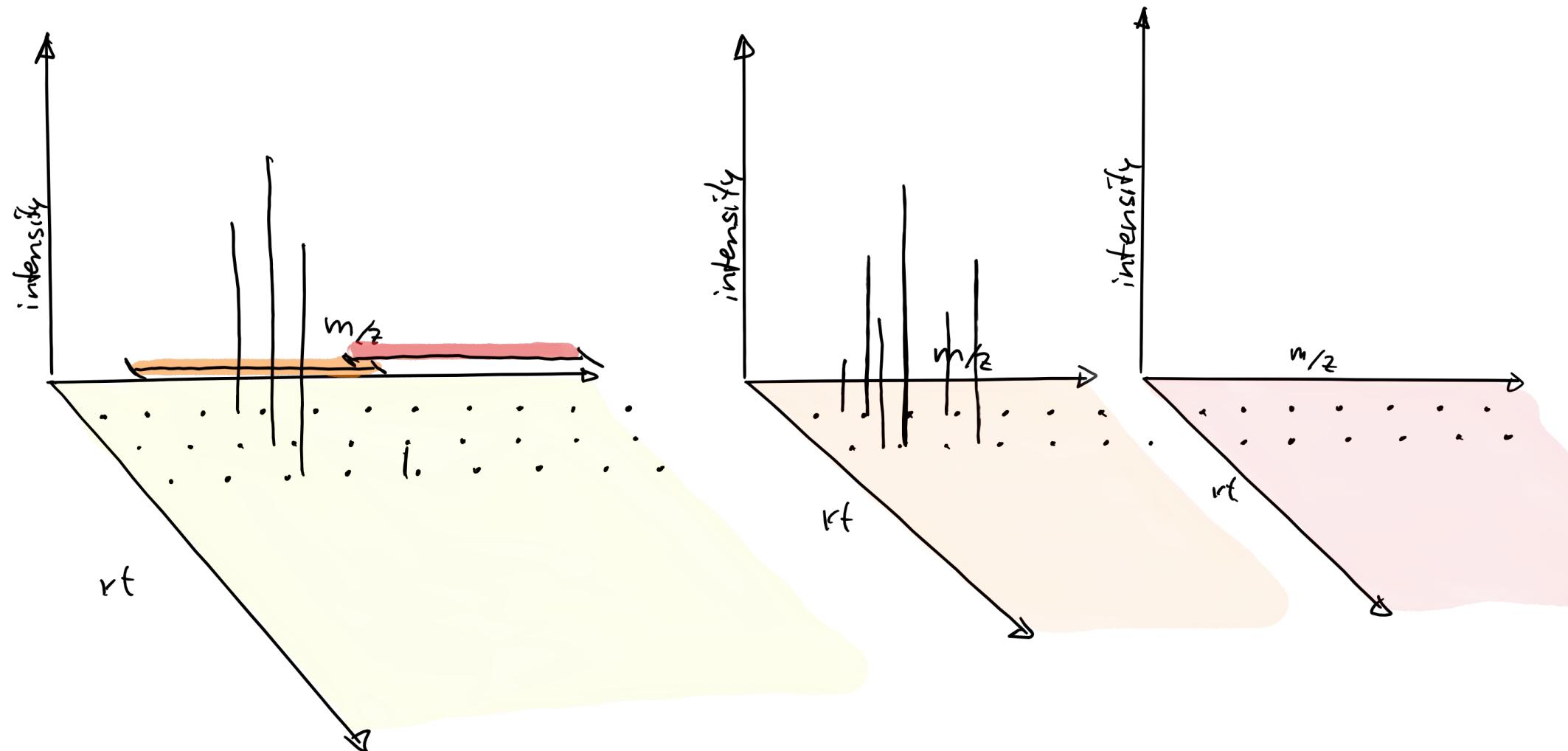
SWATH data



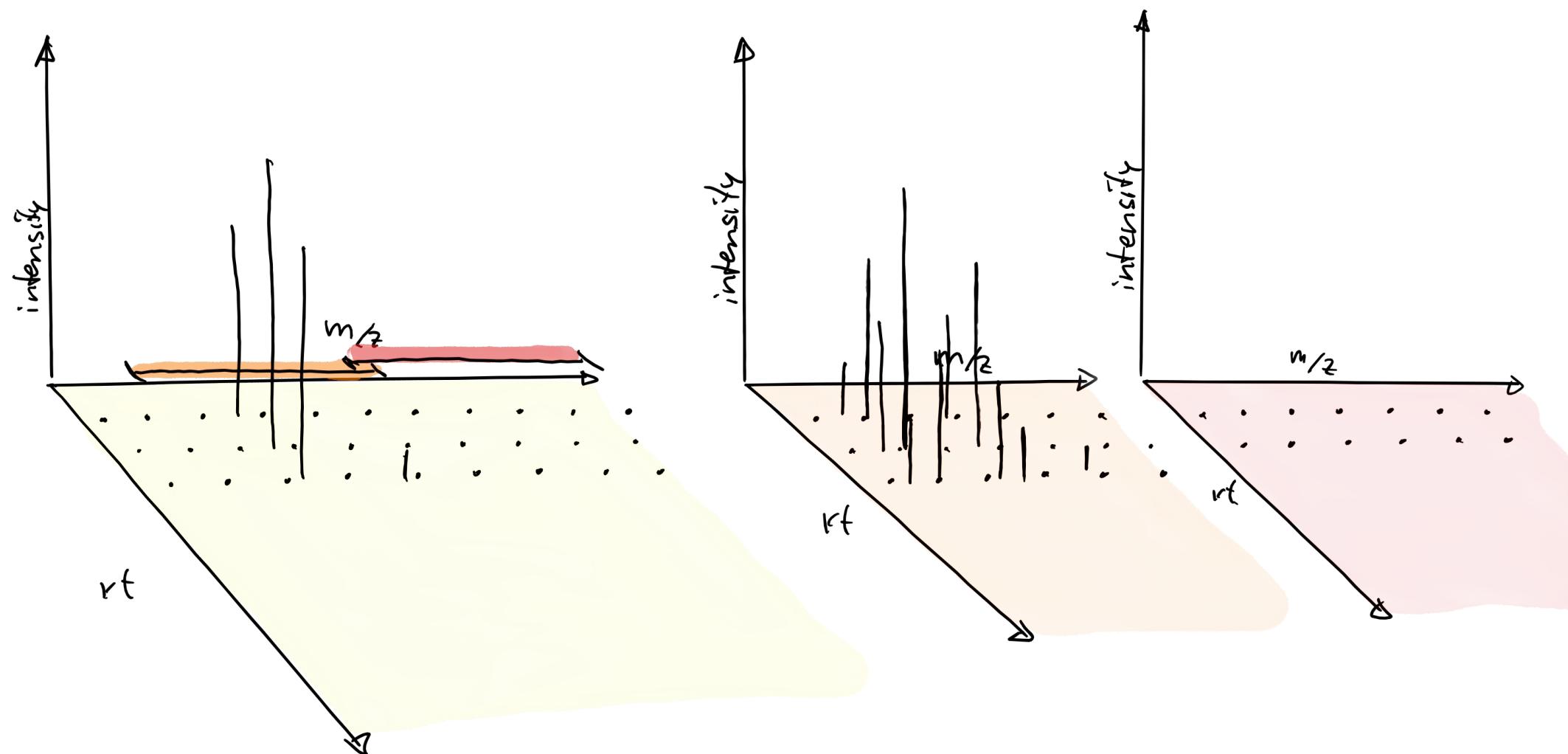
SWATH data



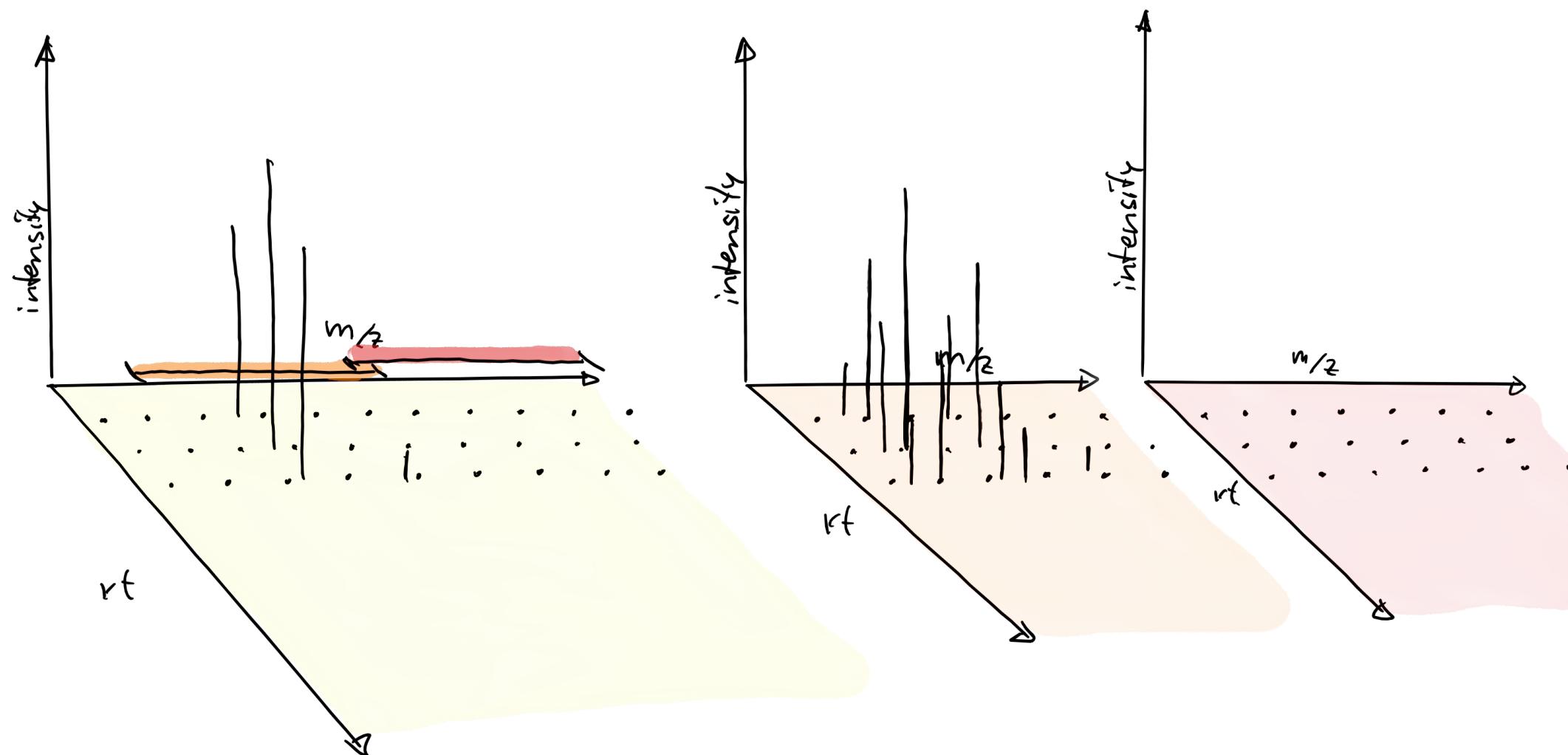
SWATH data



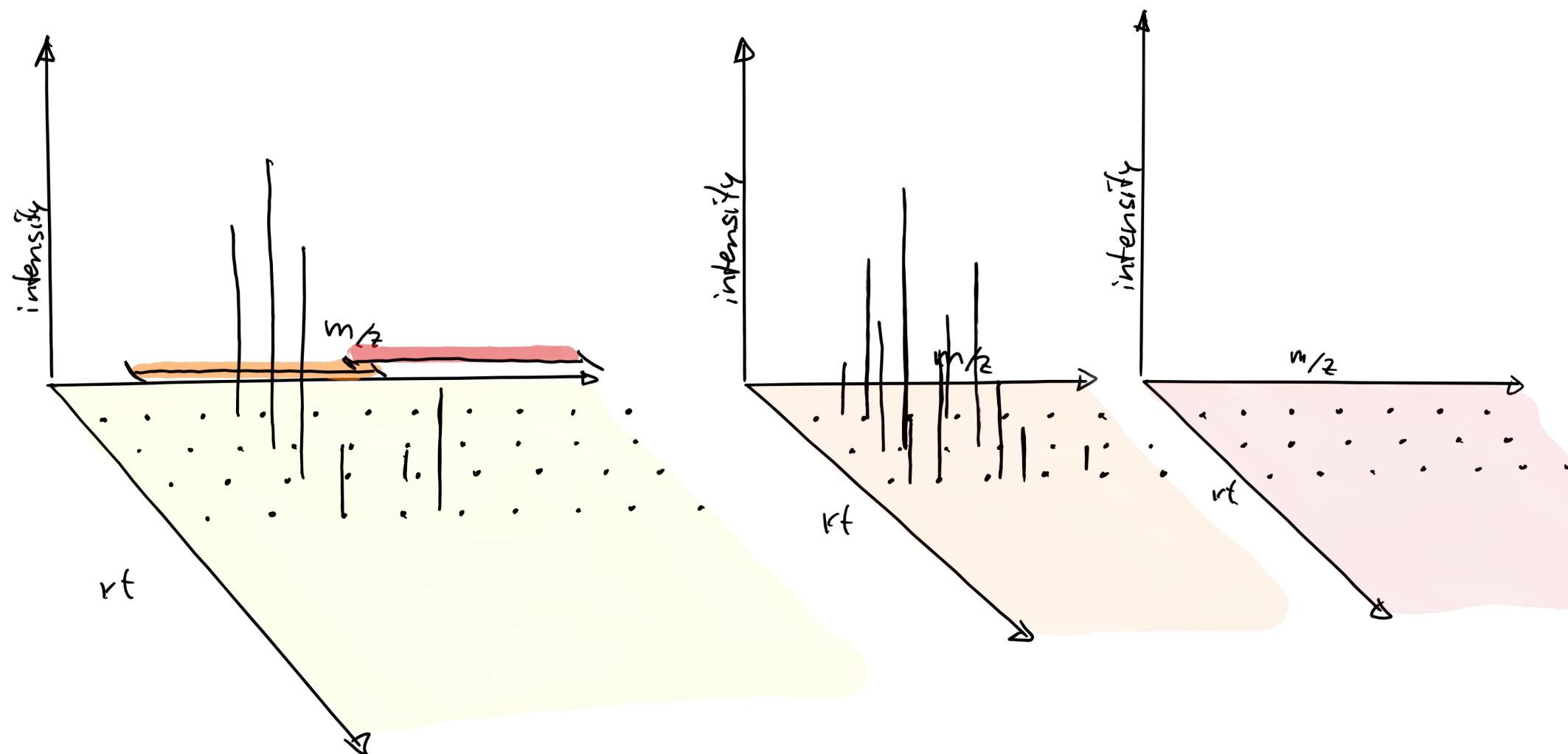
SWATH data



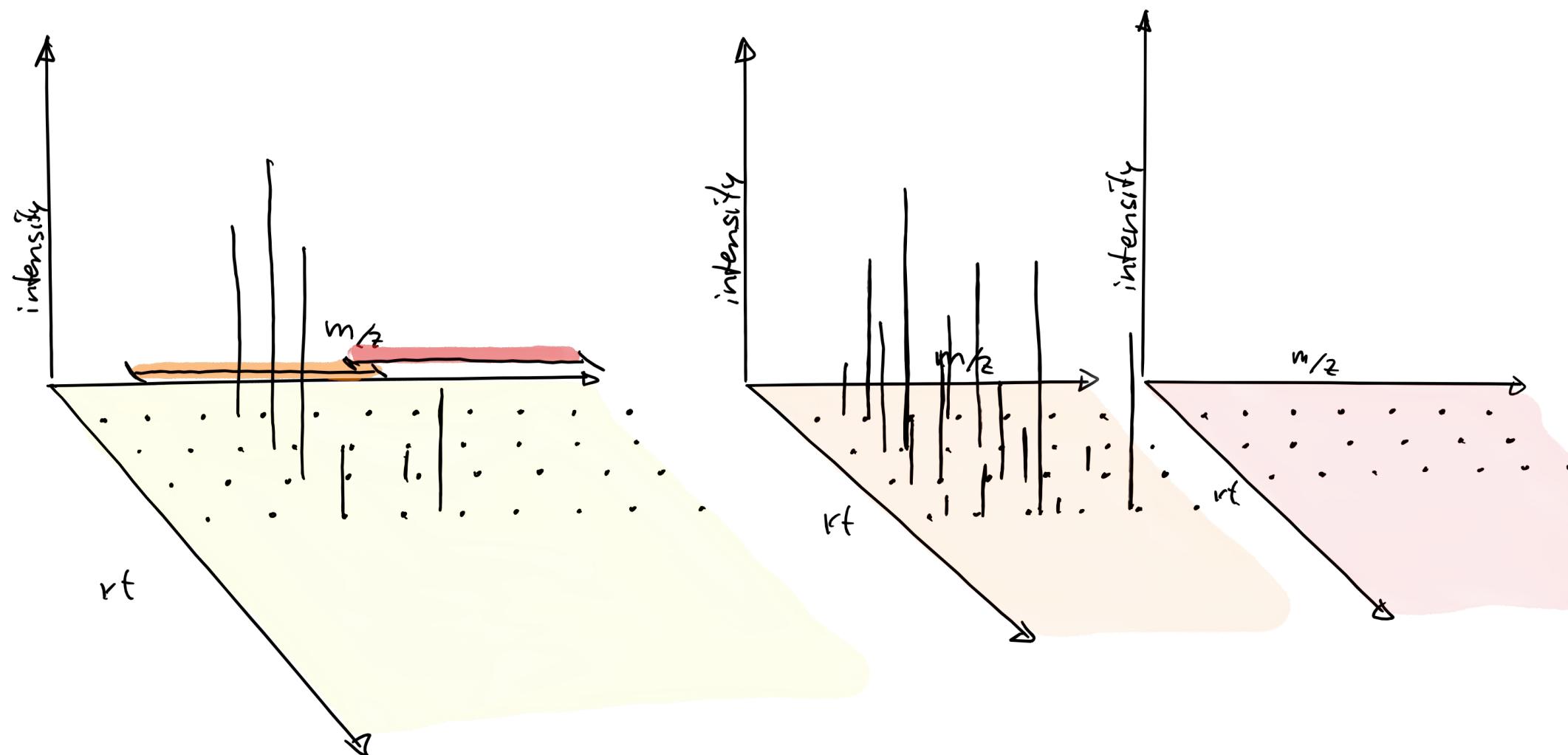
SWATH data



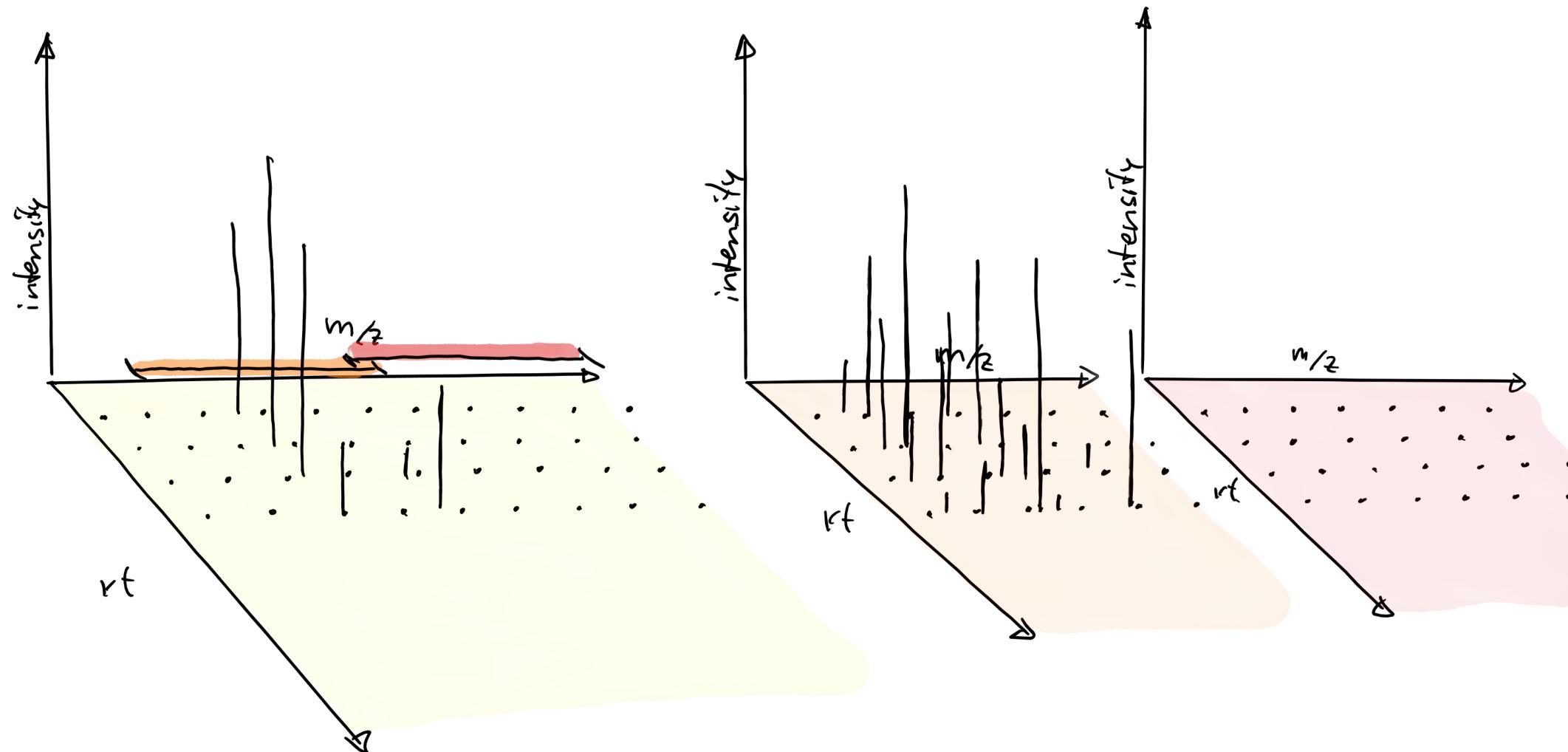
SWATH data



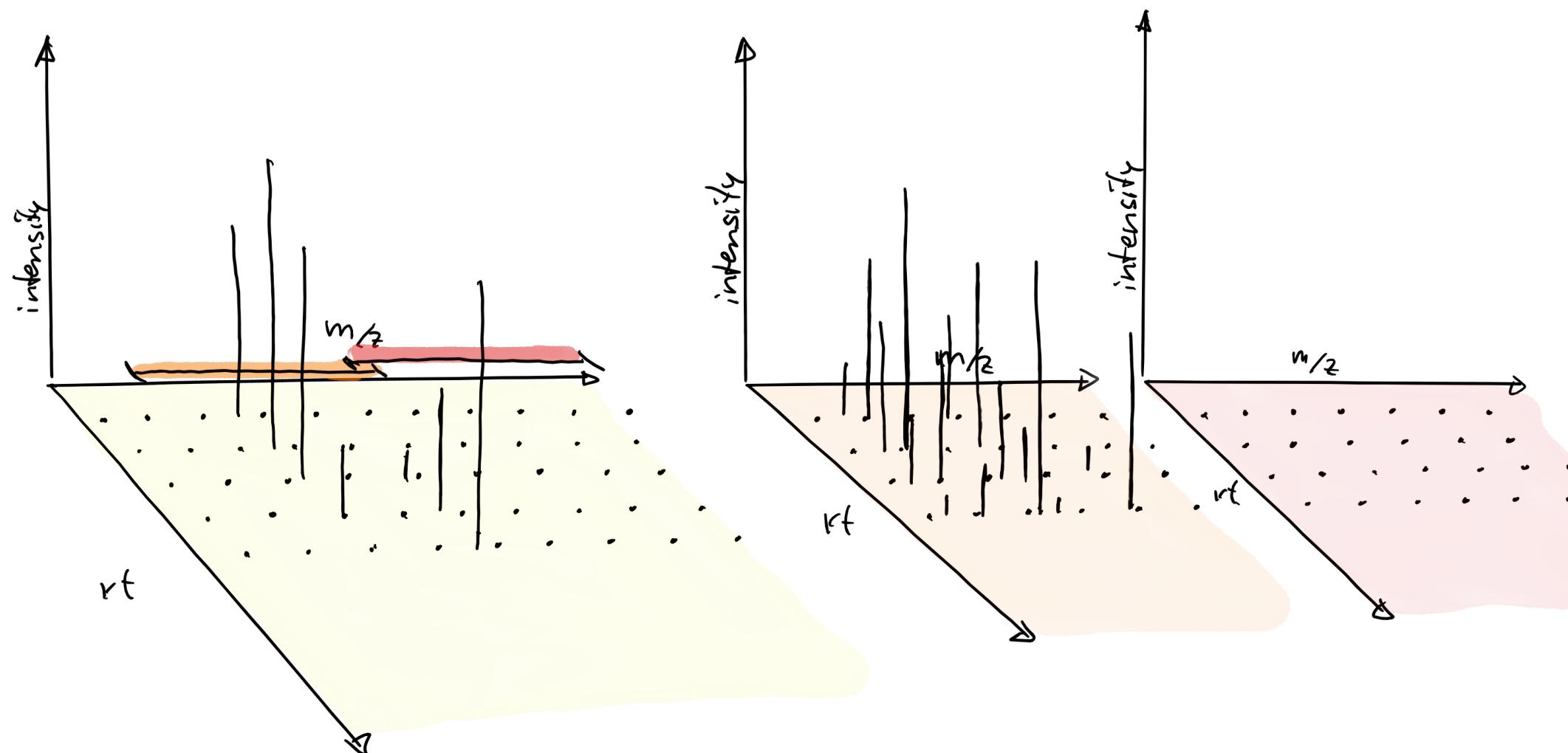
SWATH data



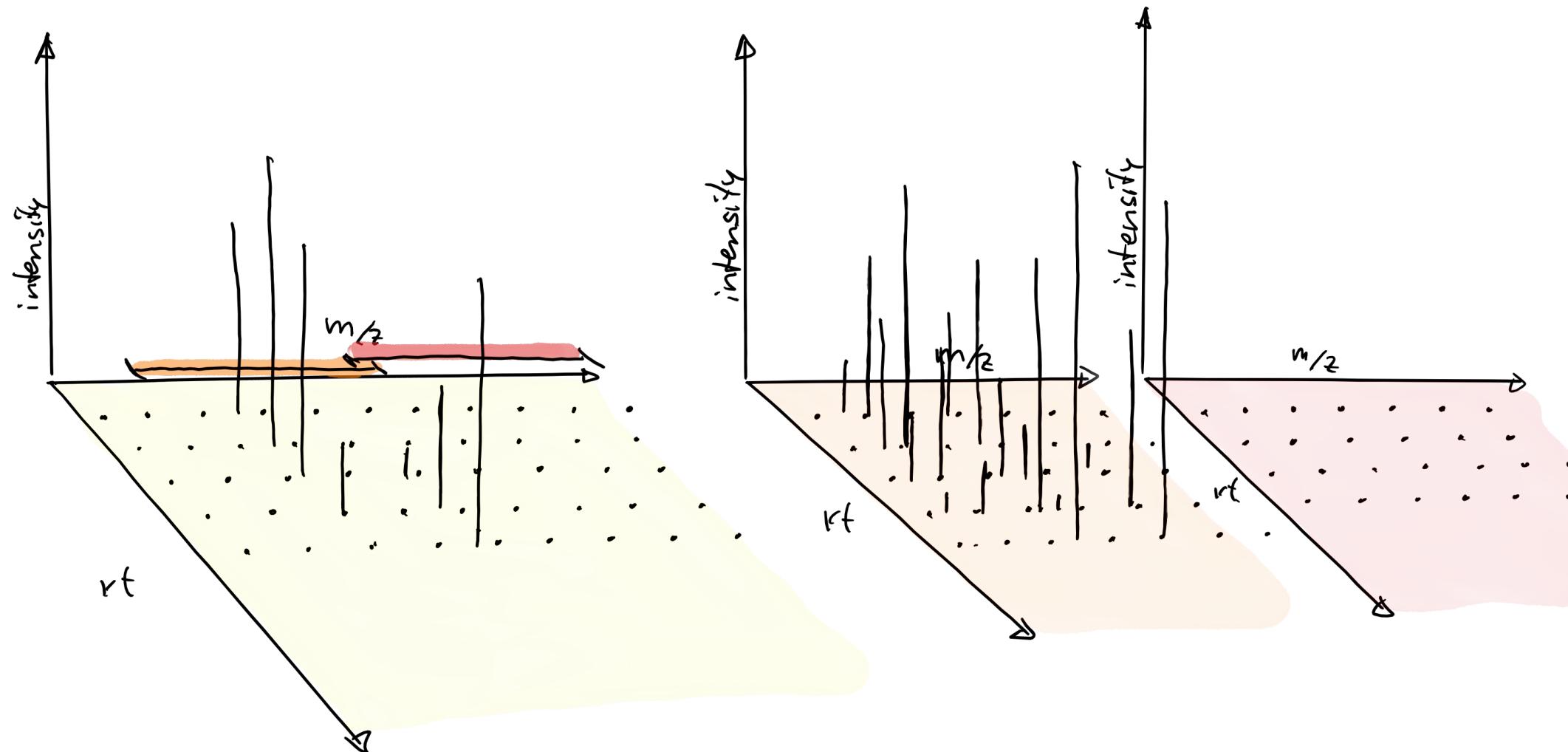
SWATH data



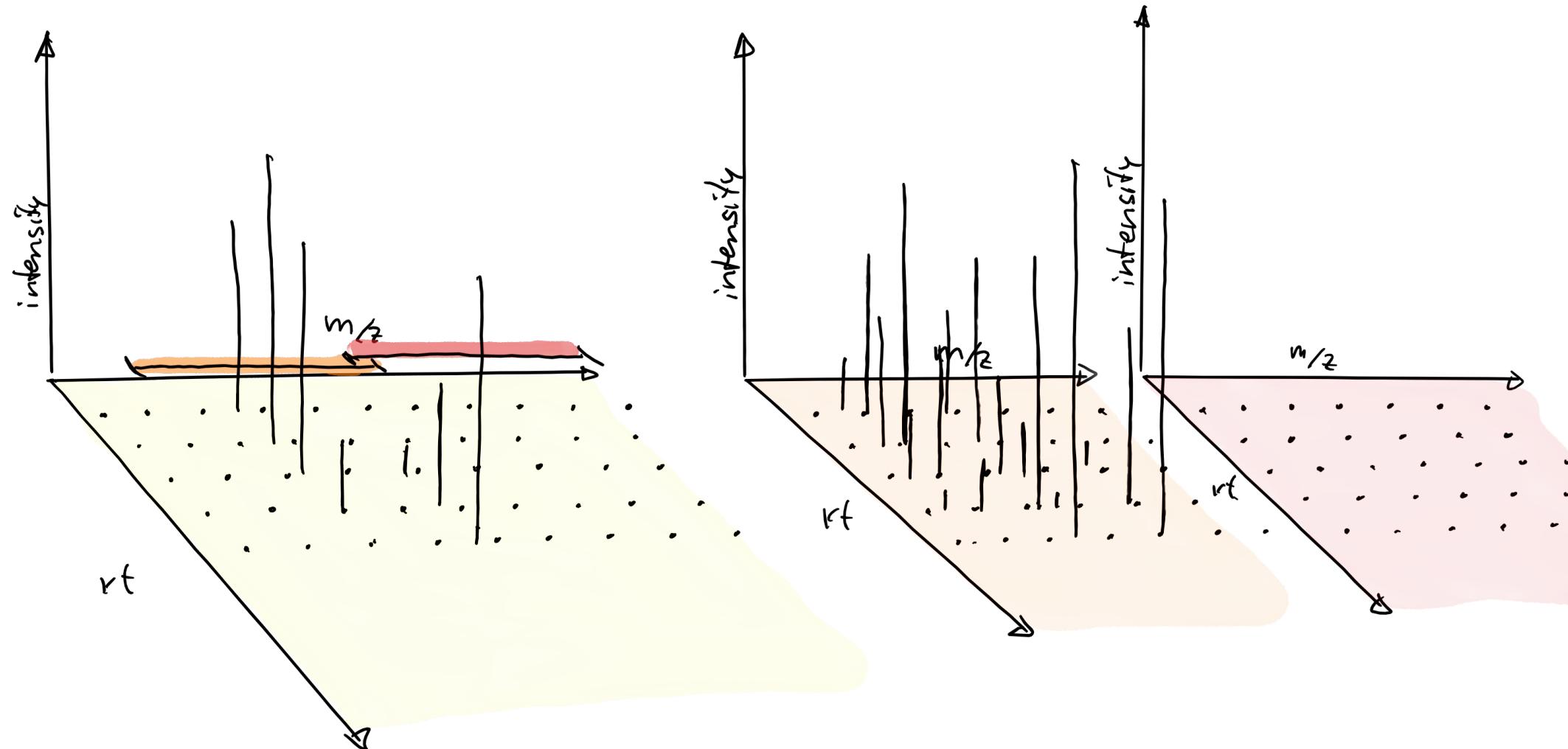
SWATH data



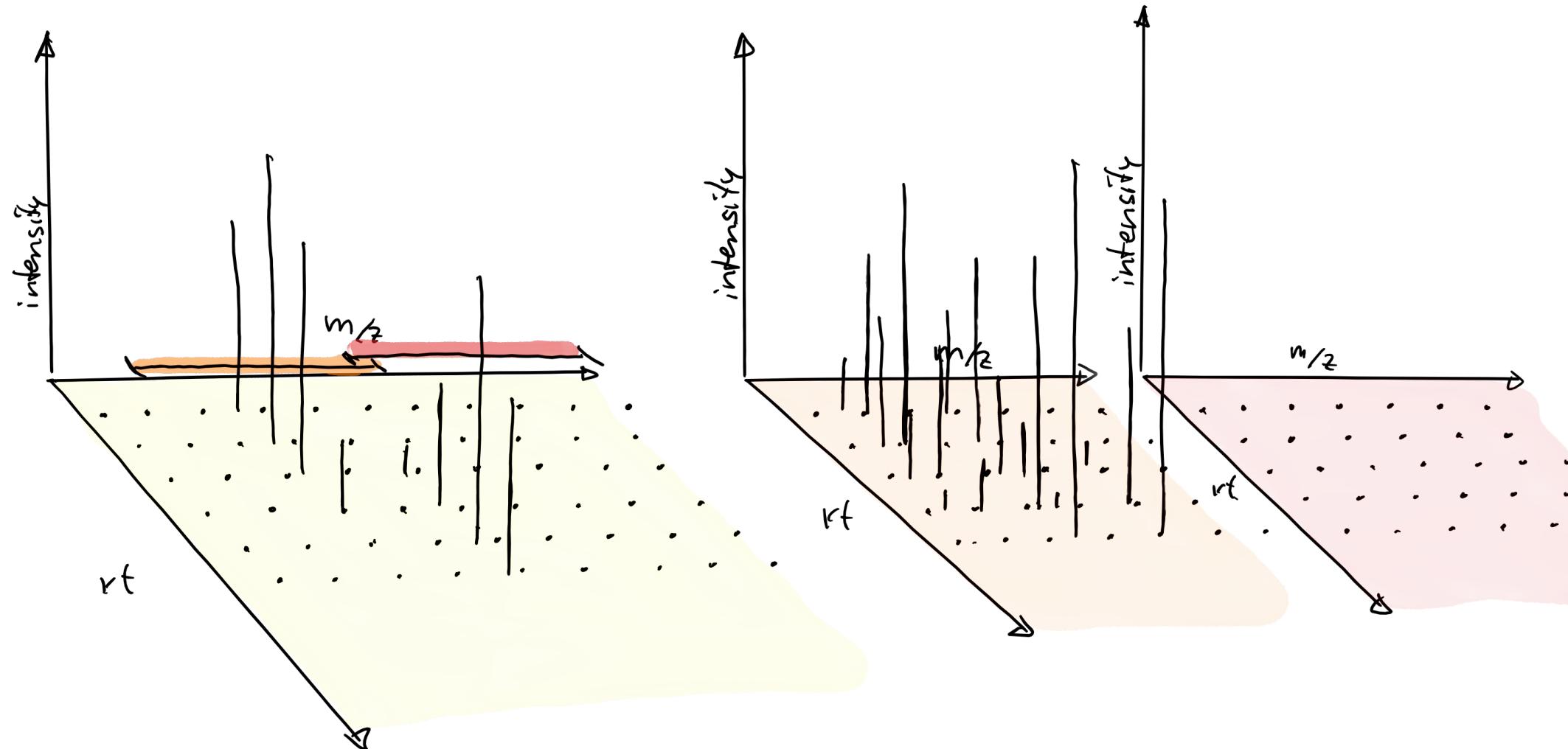
SWATH data



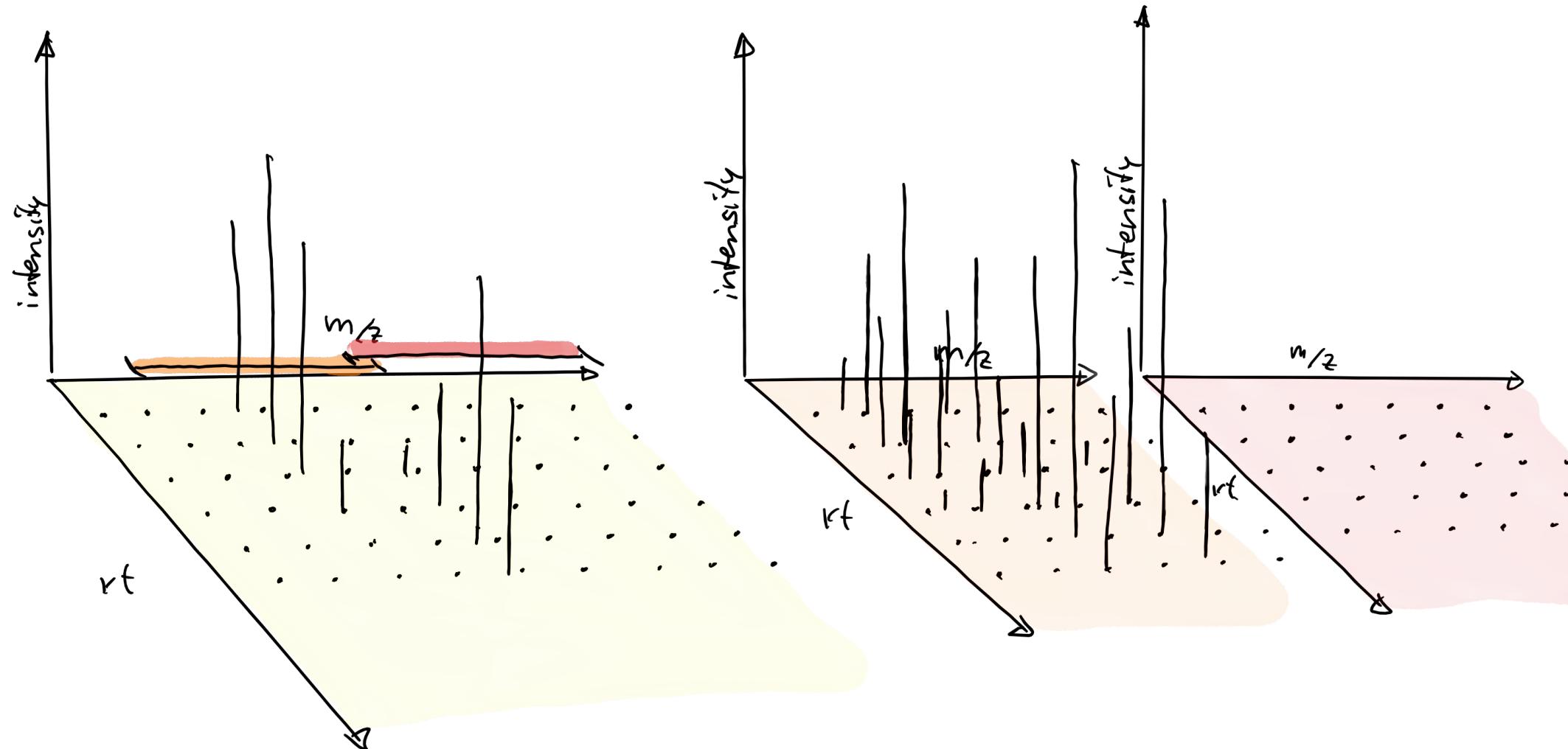
SWATH data



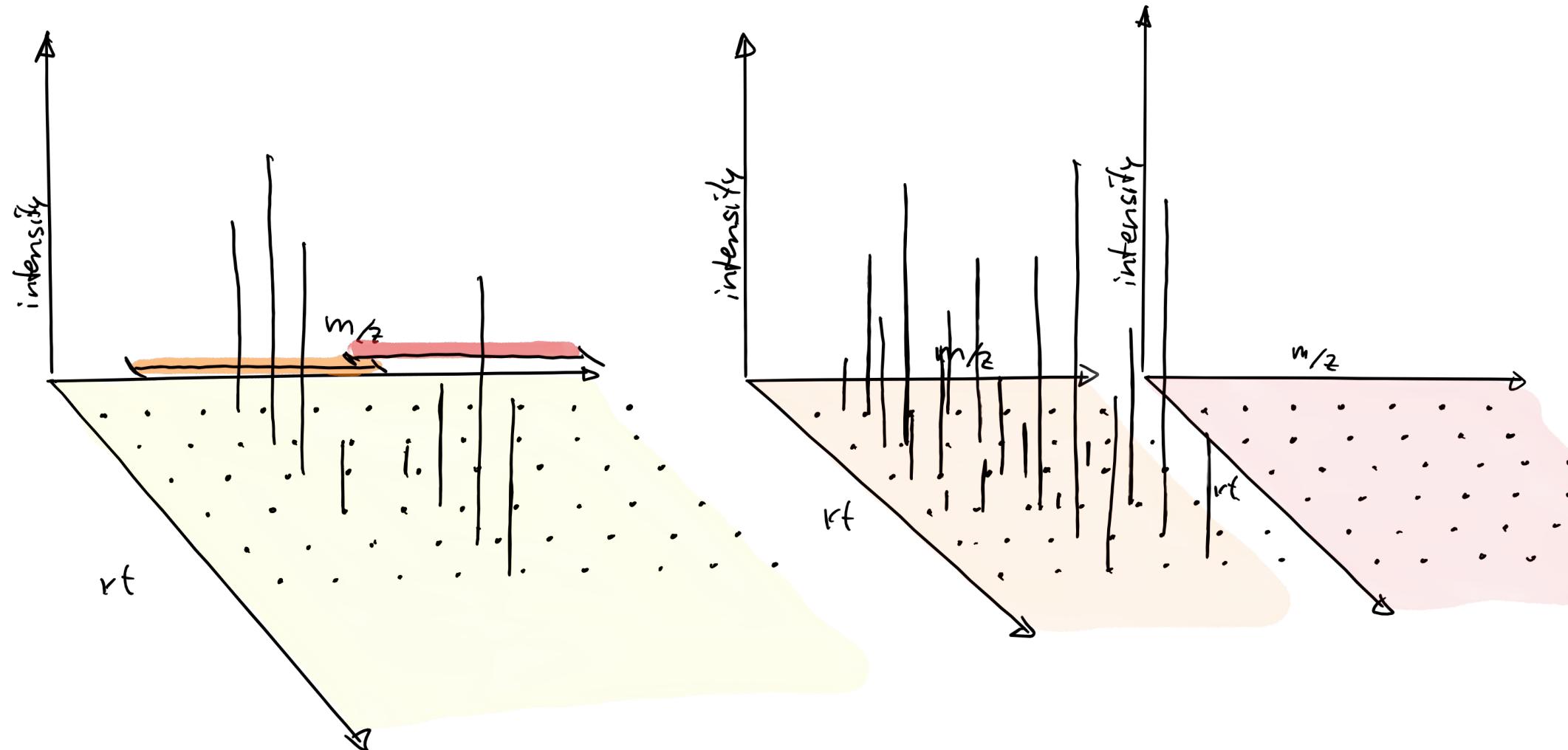
SWATH data



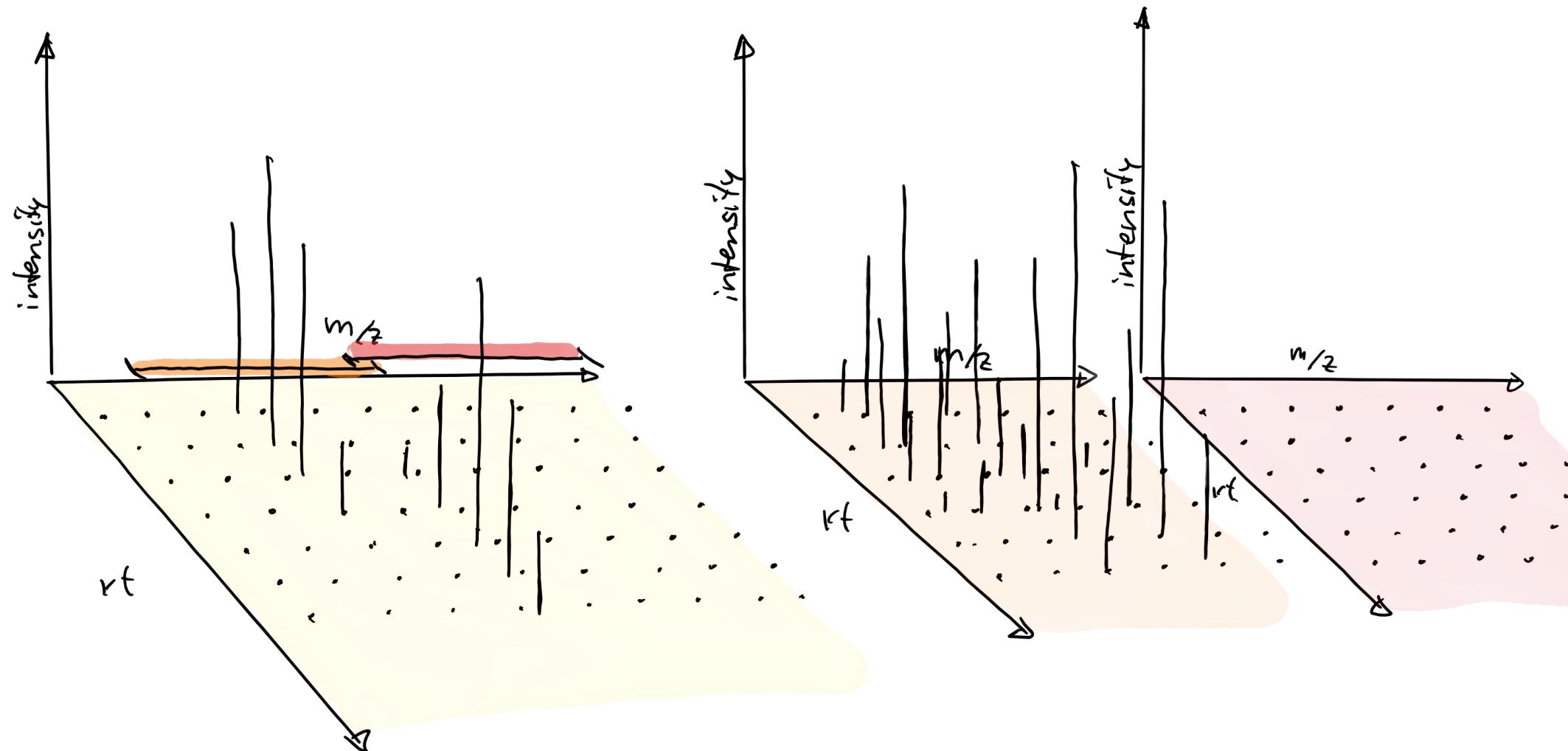
SWATH data



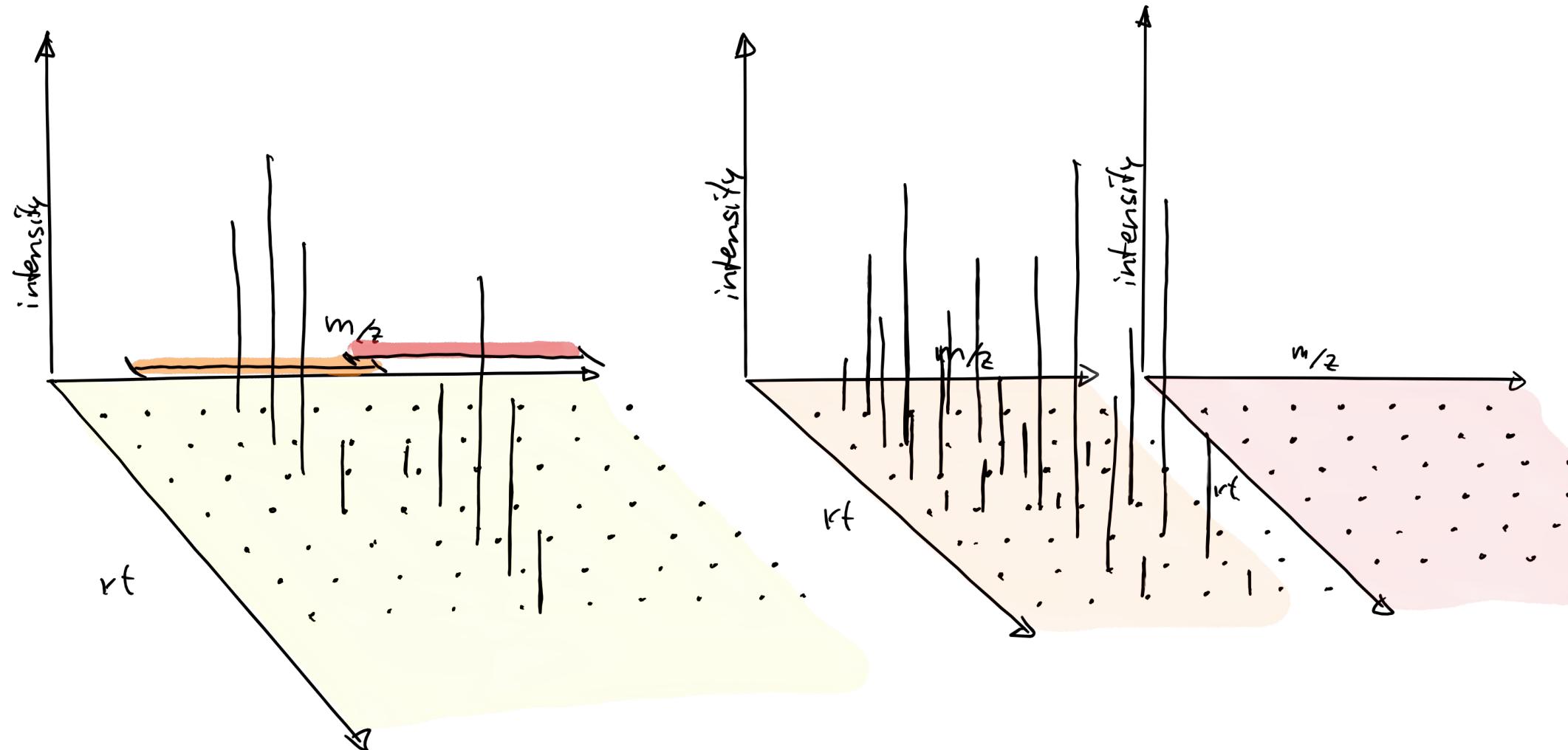
SWATH data



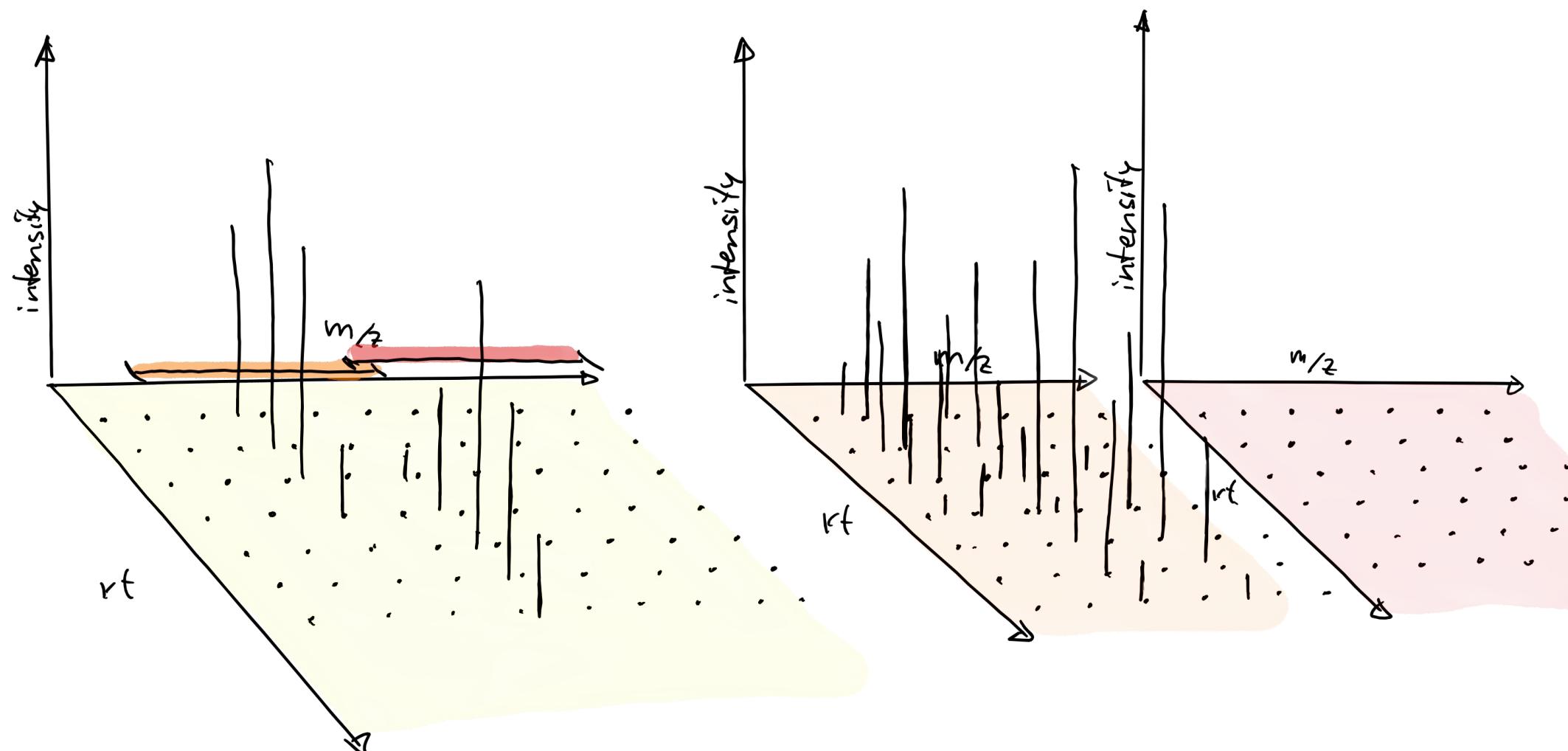
SWATH data



SWATH data



SWATH data

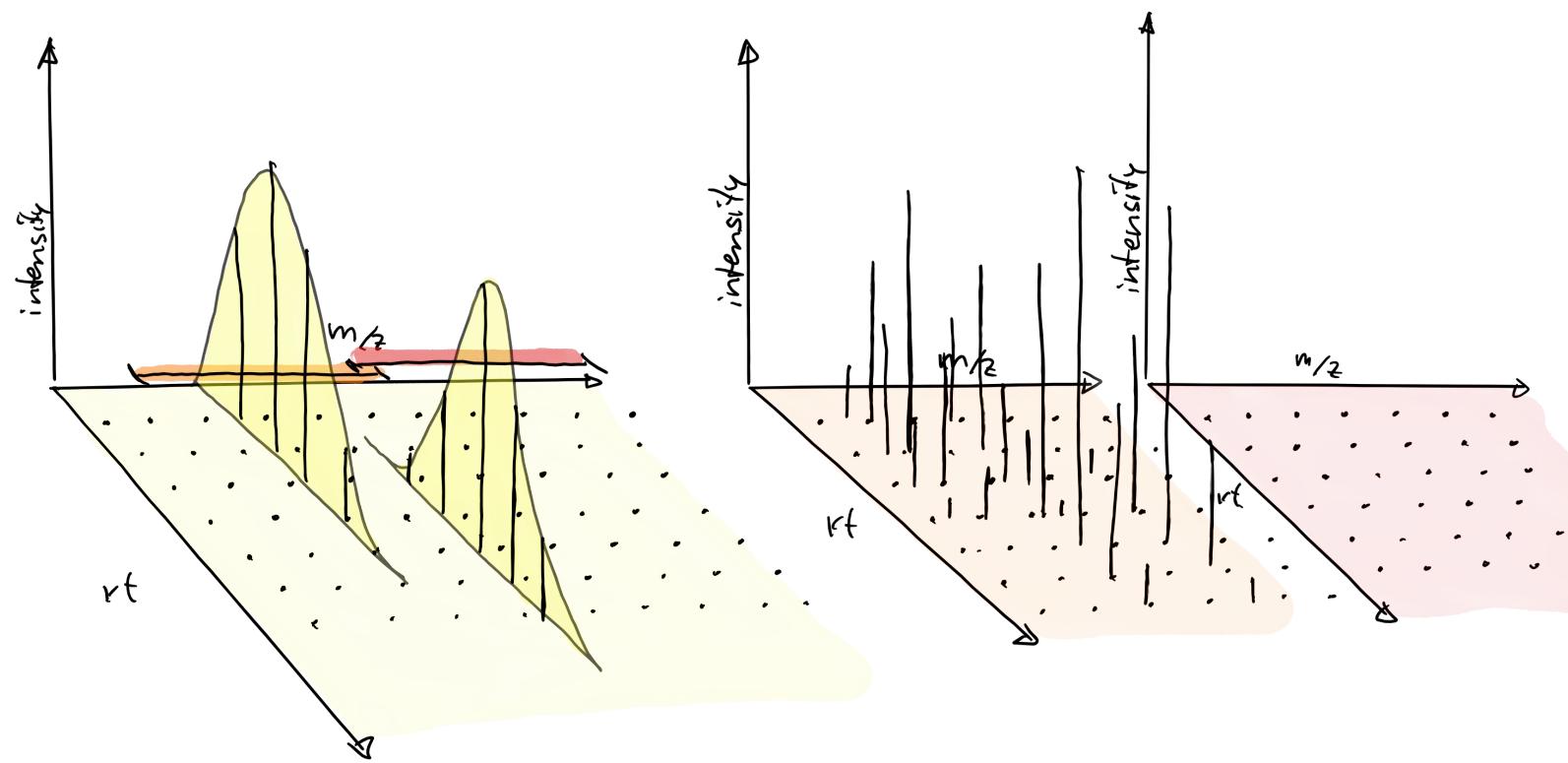


```
swath_data <- readMSData("PestMix1_SWATH.mzML", mode = "onDisk")
```

Analyzing SWATH data with **xcms**

- Chromatographic peak detection in MS1.

```
cwp <- CentWaveParam(snthresh = 5, noise = 100, ppm = 10,  
                      peakwidth = c(3, 30))  
swath_data <- findChromPeaks(swath_data, param = cwp)
```



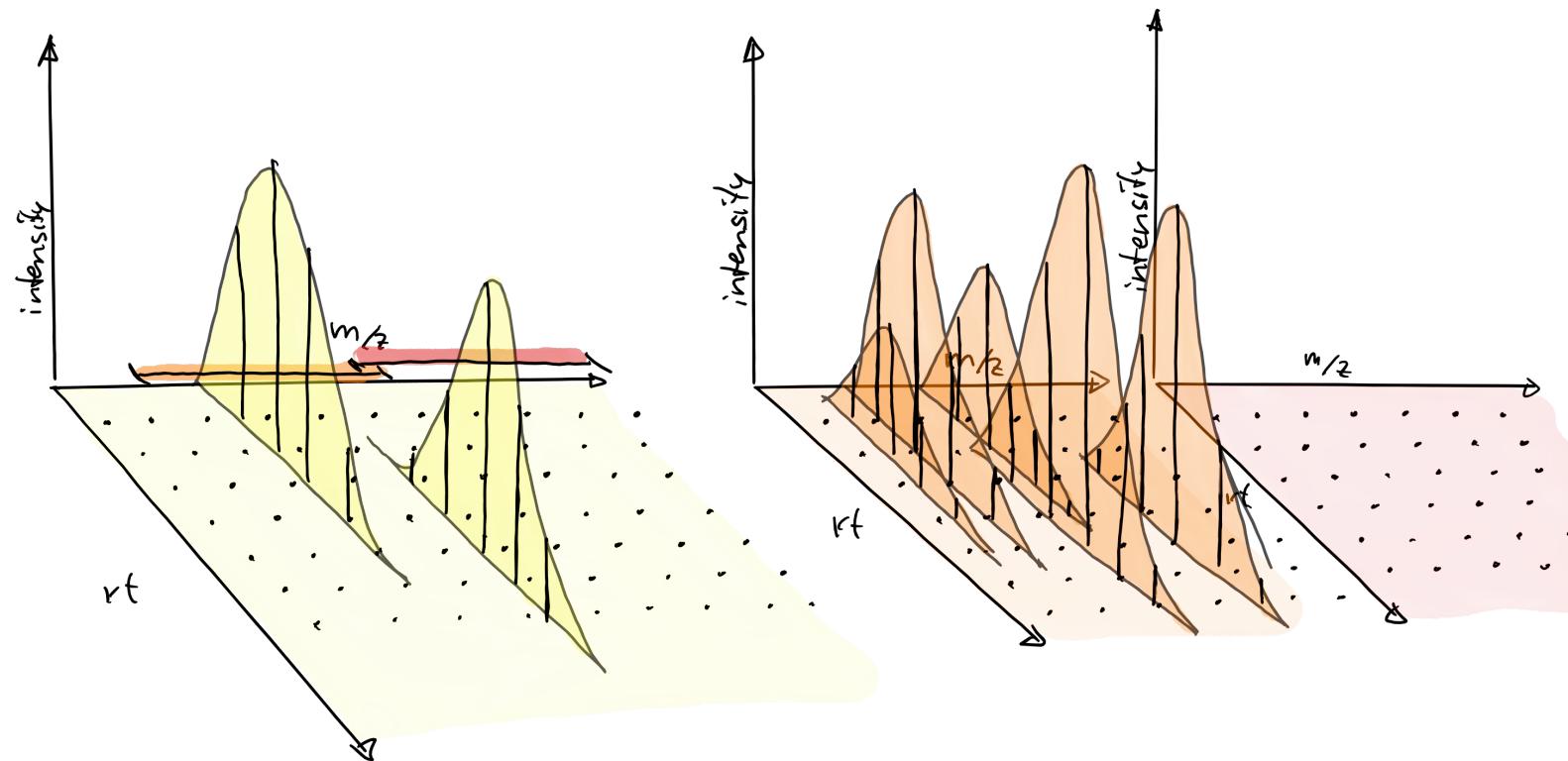
Analyzing SWATH data with **xcms**

- Chromatographic peak detection in MS1.

```
cwp <- CentWaveParam(snthresh = 5, noise = 100, ppm = 10,  
                      peakwidth = c(3, 30))  
swath_data <- findChromPeaks(swath_data, param = cwp)
```

- Chromatographic peak detection in MS2 (within each isolation window).

```
swath_data <- findChromPeaksIsolationWindow(swath_data, param = cwp)
```



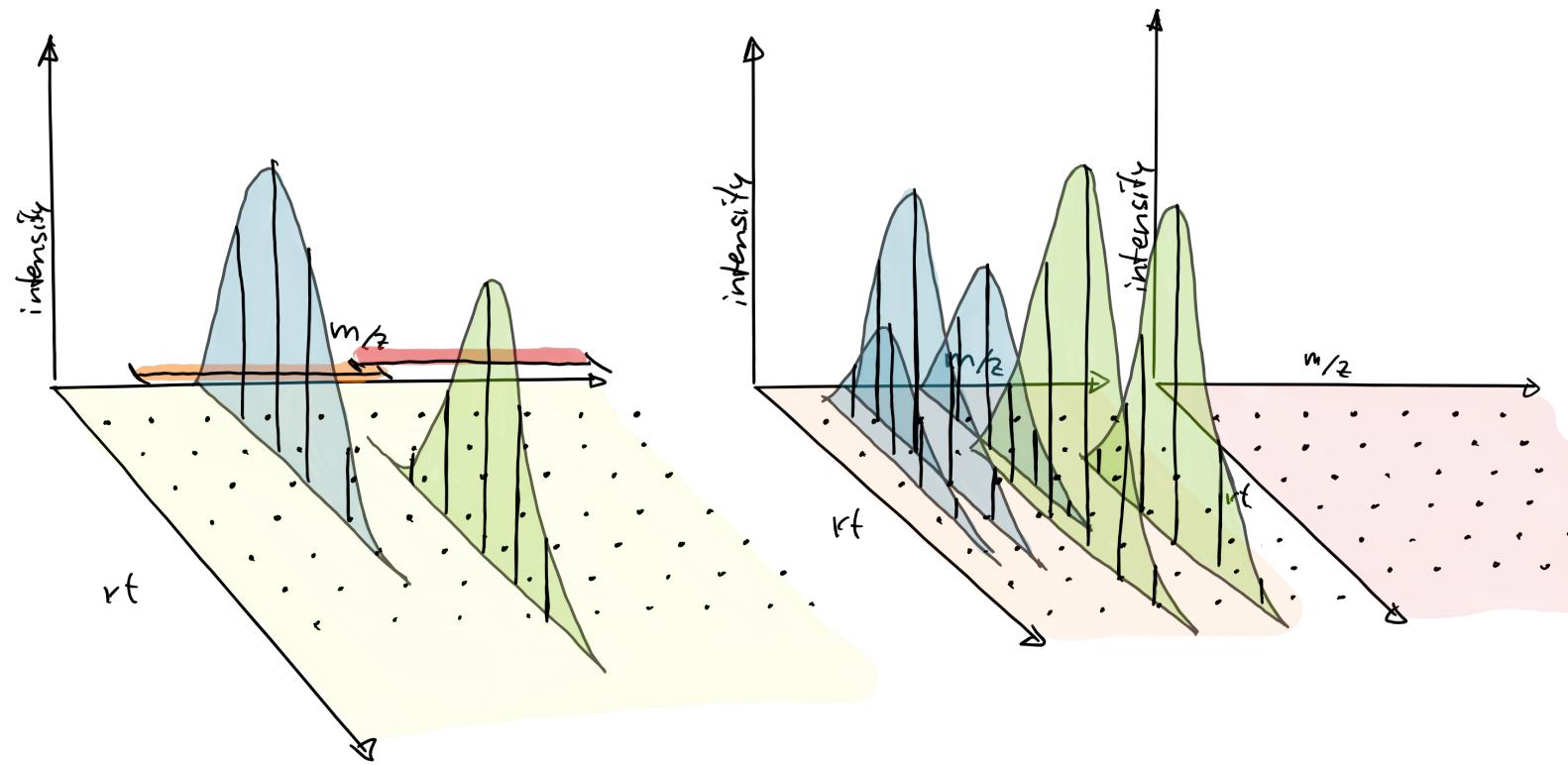
Analyzing SWATH data with **xcms**

- Chromatographic peak detection in MS1.

```
cwp <- CentWaveParam(snthresh = 5, noise = 100, ppm = 10,  
                      peakwidth = c(3, 30))  
swath_data <- findChromPeaks(swath_data, param = cwp)
```

- Chromatographic peak detection in MS2 (within each isolation window).

```
swath_data <- findChromPeaksIsolationWindow(swath_data, param = cwp)
```



Analyzing SWATH data with **xcms**

- Reconstructing MS2 spectrum from SWATH data:

```
swath_spectra <- reconstructChromPeakSpectra(swath_data, minCor = 0.9)
```

- For each MS1 chromatographic peak:
 - Find MS2 peaks (within the correct isolation window) with similar retention time.
 - Correlate peak shape of MS1 and candidate MS2 peaks.
 - Reconstruct the MS2 spectra based on matching MS2 peaks' m/z and intensity.

Analyzing SWATH data with **xcms**

- Example: reconstructed MS2 spectrum for Fenamiphos.

```
chromPeaks(swath_data, mz = 304.1131, ppm = 20, msLevel = 1L)
```

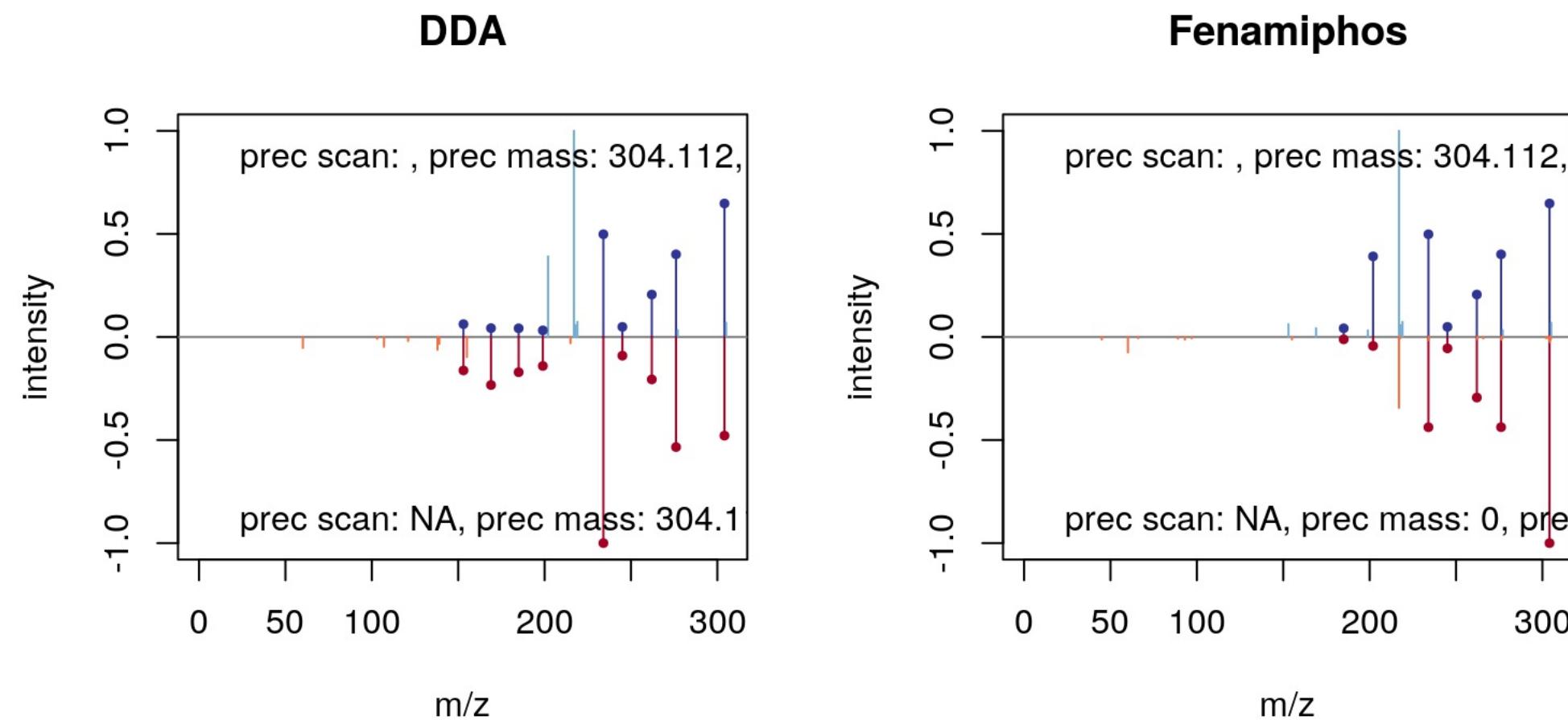
```
##          mz      mzmin      mzmax       rt      rtmin      rtmax      into      intb
## CP35 304.1124 304.1121 304.1126 423.945 419.445 428.444 10697.34 10688.34
##          maxo     sn sample
## CP35 2401.849 618       1
```

```
swath_sp <- swath_spectra[mcols(swath_spectra)$peak_id == "CP35"]
swath_sp
```

```
## Spectra with 1 spectra and 3 metadata column(s):
##      msLevel      rtime peaksCount |           ms2_peak_id
##      <integer> <numeric> <integer> |           <CharacterList>
##      1          2        NA         15 | CP205,CP207,CP217,....
##                                         ms2_peak_cor      peak_id
##                                         <NumericList> <character>
##      1 0.999787099100833,0.964862731008839,0.980288982893133,...      CP35
```

Analyzing SWATH data with **xcms**

```
par(mfrow = c(1, 2))
plot(swath_sp[[1]], ex_spectrum, main = "DDA", tolerance = 40e-6)
plot(swath_sp[[1]], fenamiphos, main = "Fenamiphos", tolerance = 40e-6)
```



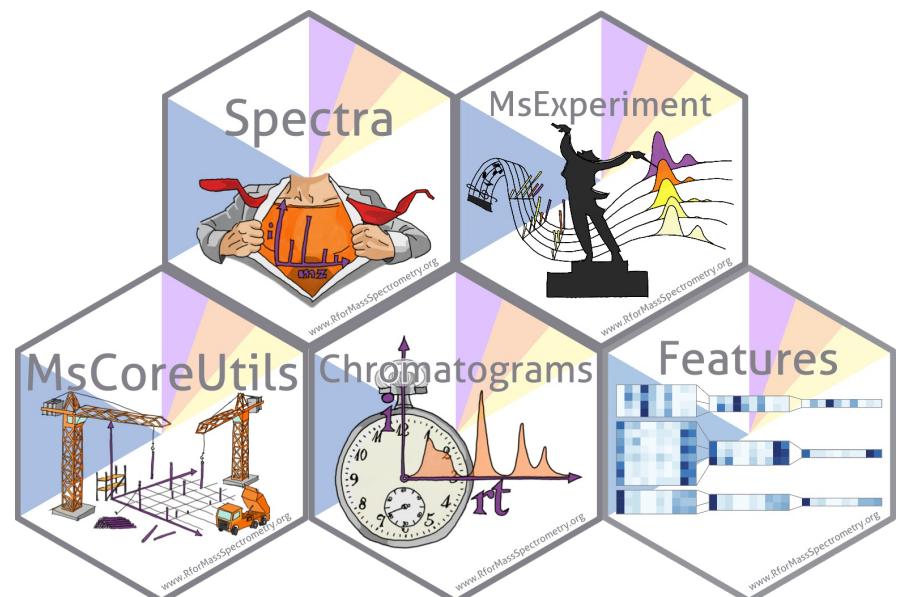
Annotating MS2 spectra

- Compare `Spectra` against reference `Spectra`:
 - import from mgf file(s).
 - ... (future developments...)
- Export spectra and peak/feature intensities for [Feature-Based Molecular Networking](#)

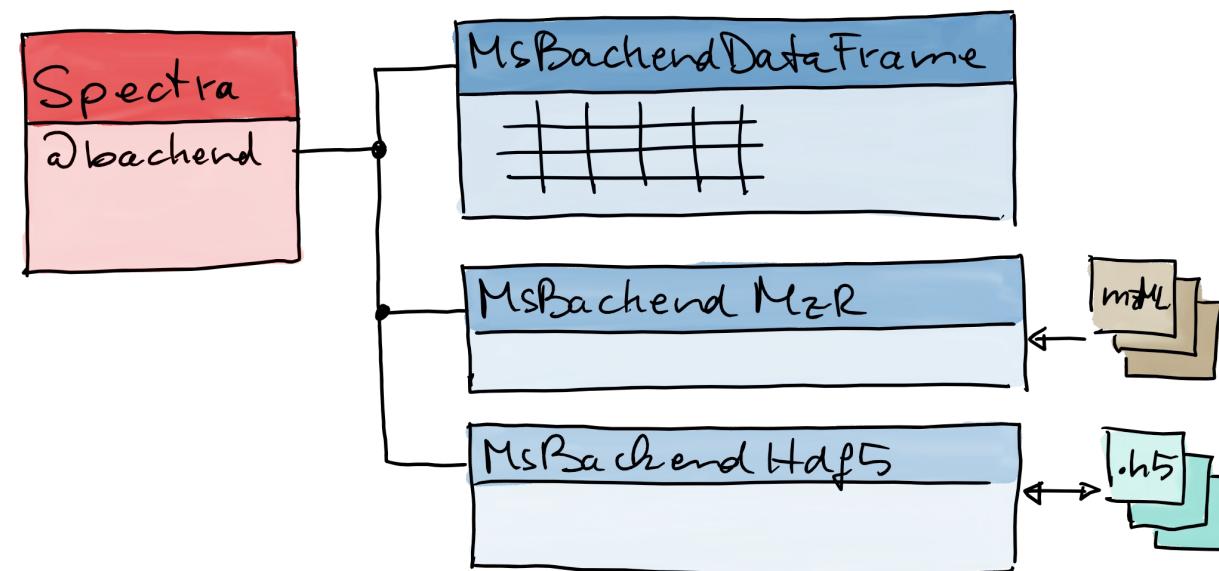


Future developments

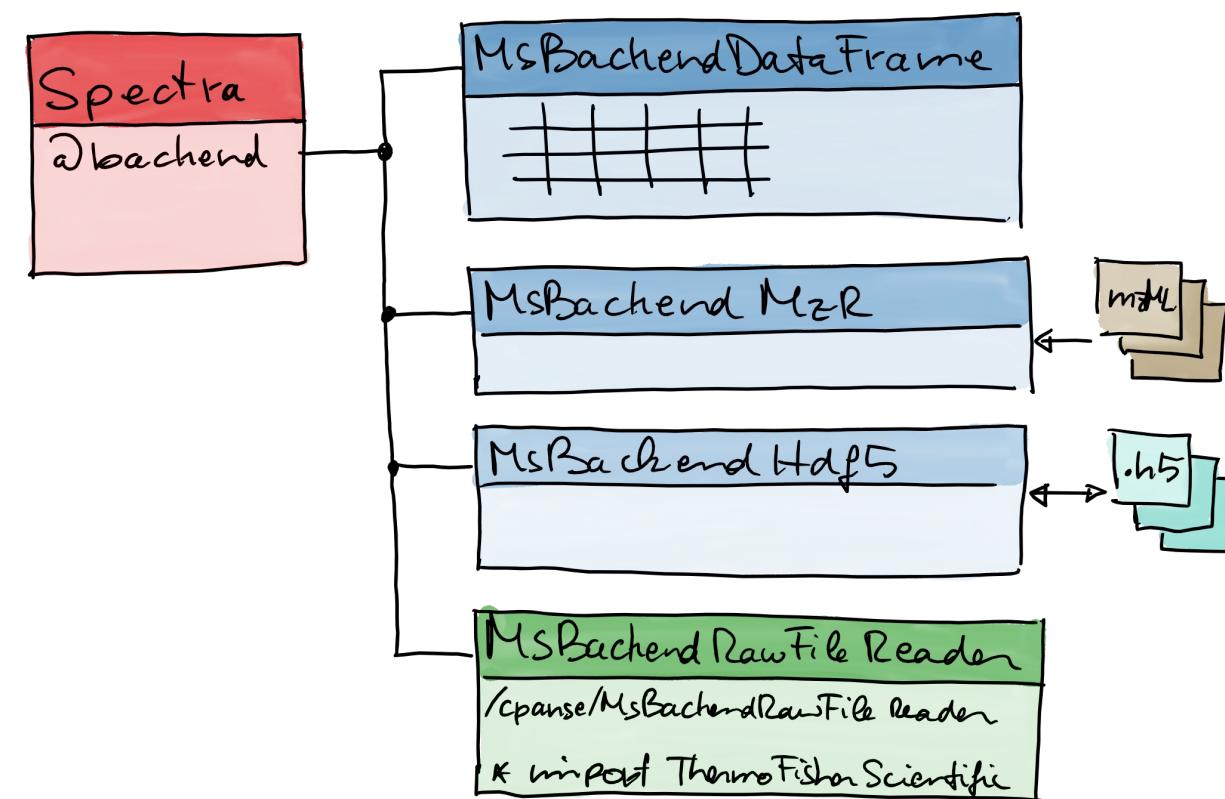
- <https://RforMassSpectrometry.org>
- Define an efficient, flexible and well documented infrastructure for Mass Spectrometry data in R.
- Provide core functionality.
- Provide core data representations.
- Reusable in other packages.



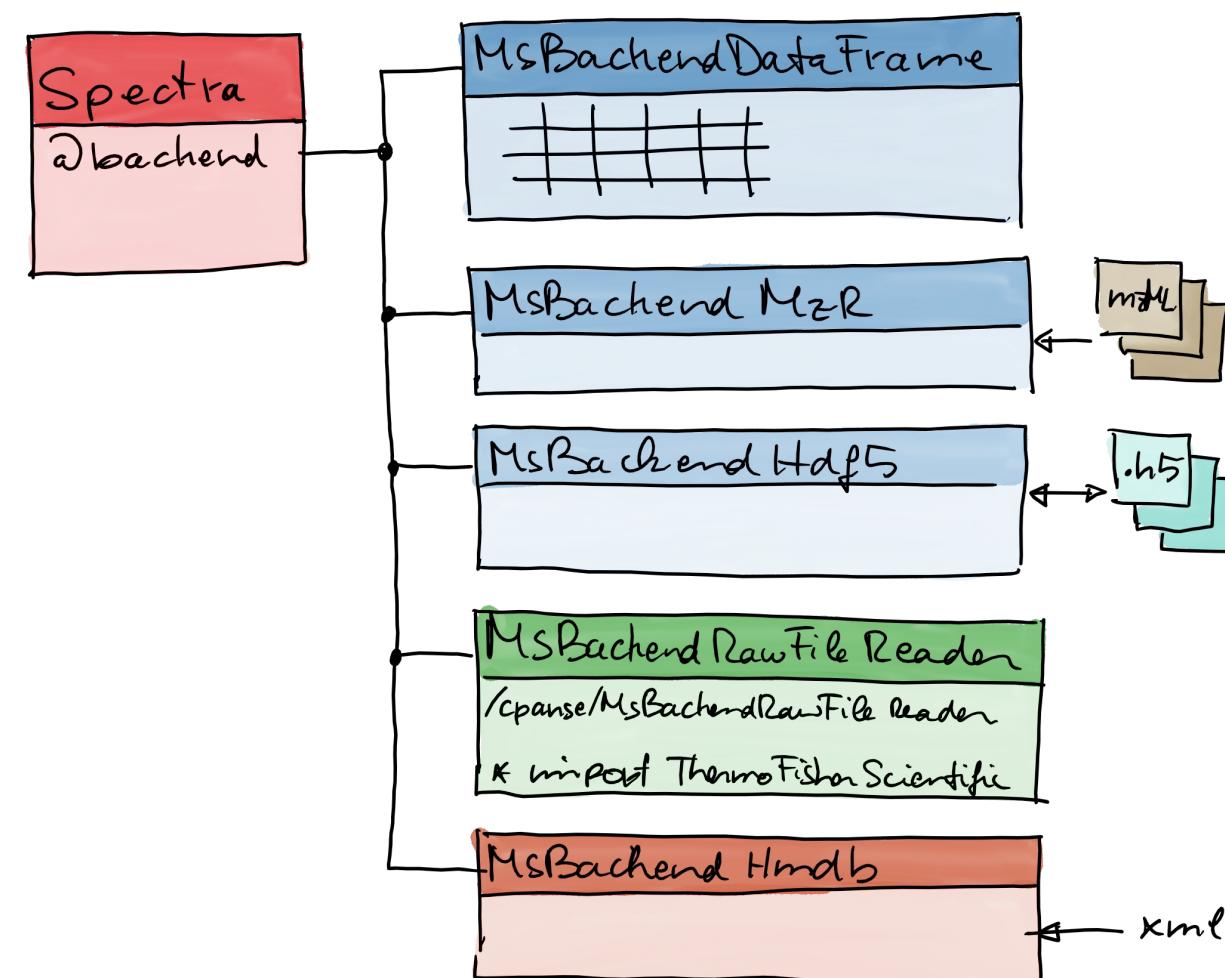
- Introduction of *backends*: independence between MS functionality and data origin/storage.



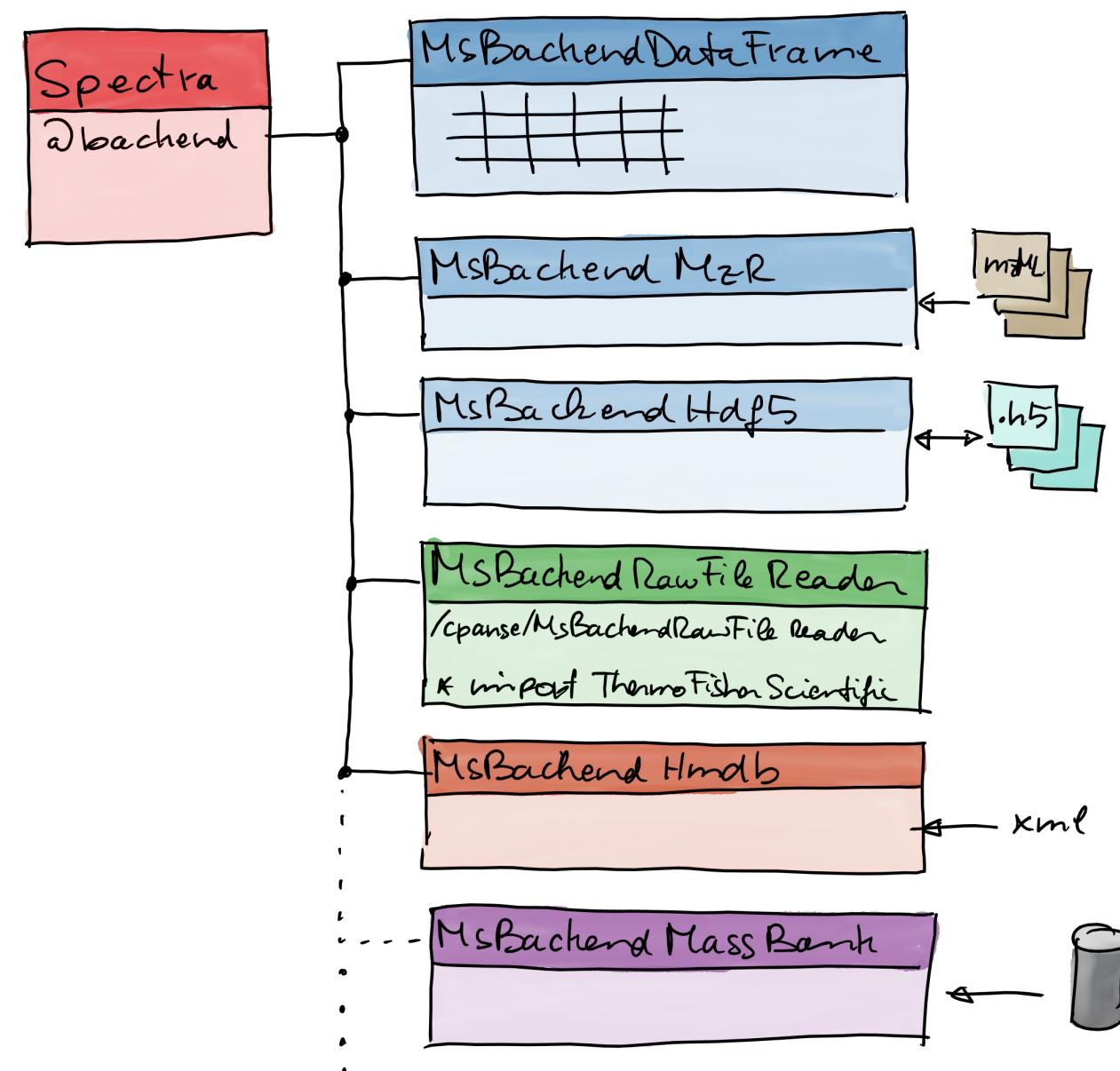
- Introduction of *backends*: independence between MS functionality and data origin/storage.



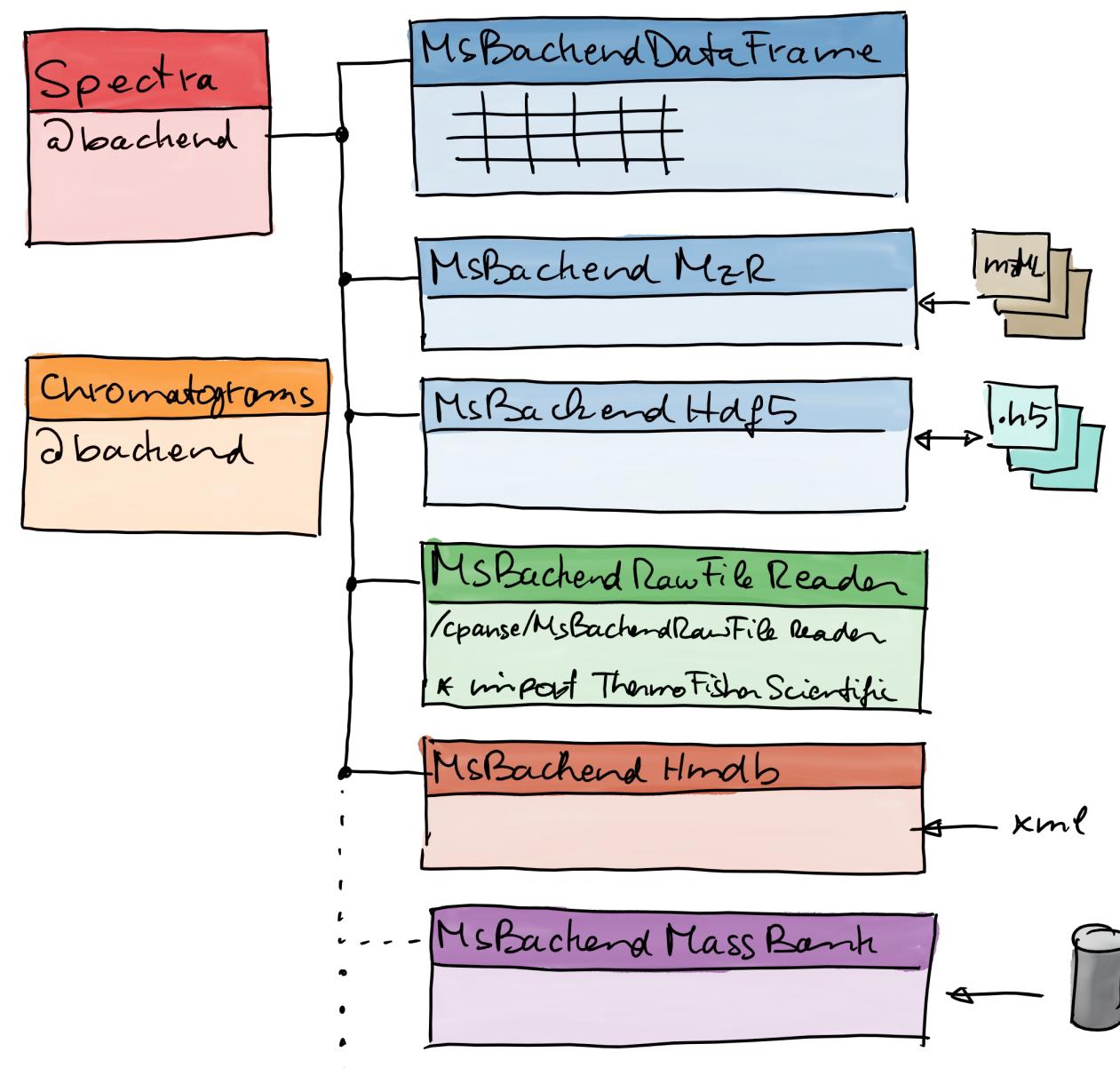
- Introduction of *backends*: independence between MS functionality and data origin/storage.



- Introduction of *backends*: independence between MS functionality and data origin/storage.



- Introduction of *backends*: independence between MS functionality and data origin/storage.



Thank you for your attention

Also thanks to: Micheal Witting, Jan Stanstrup, Steffen Neumann, Sebastian Gibb, Laurent Gatto