



Flexible infrastructure for mass spectrometry data

Laurent Gatto¹, Johannes Rainer², Sebastian Gibb³ *

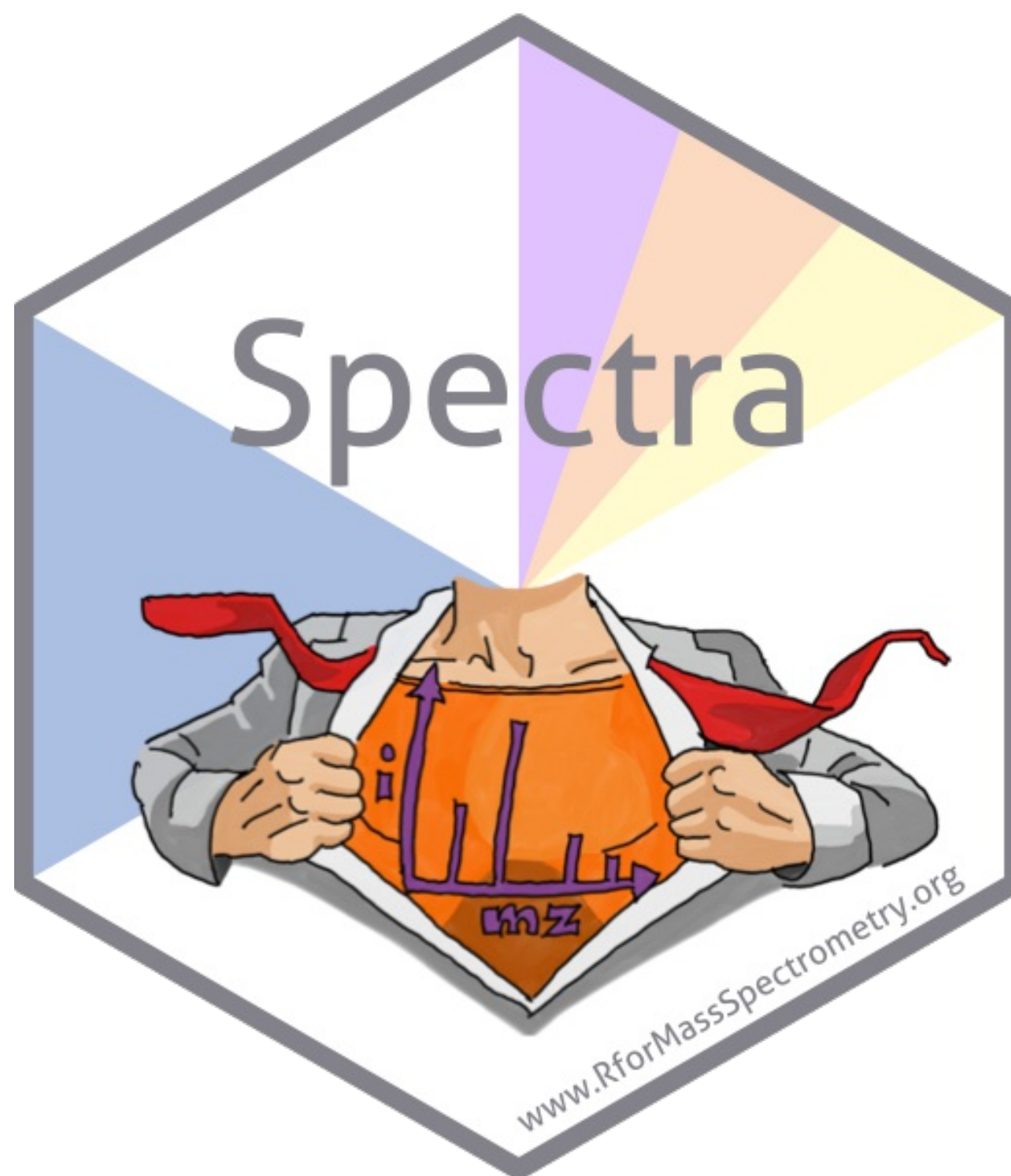
¹de Duve Institute, UCLouvain, Brussels, Belgium

²Institute for Biomedicine, Eurac Research, Italy

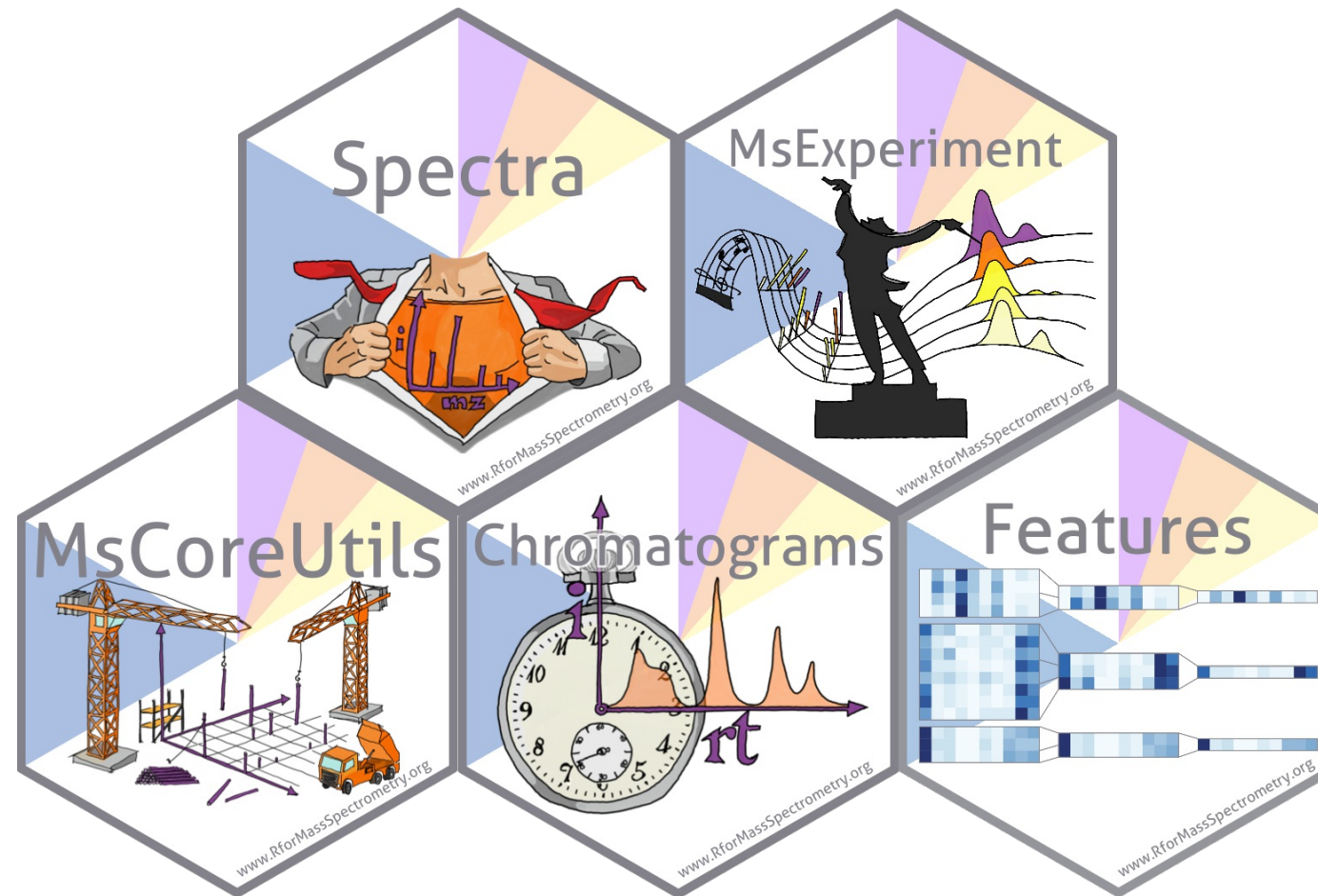
³Department of Anaesthesiology and Intensive Care, University Medicine Greifswald, Germany

*order of authors defined by `sample` with a random seed of `42`

<https://doi.org/10.5281/zenodo.3566699>

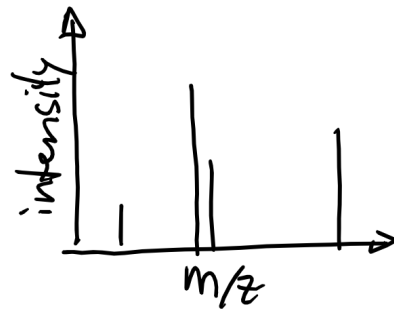


RforMassSpectrometry



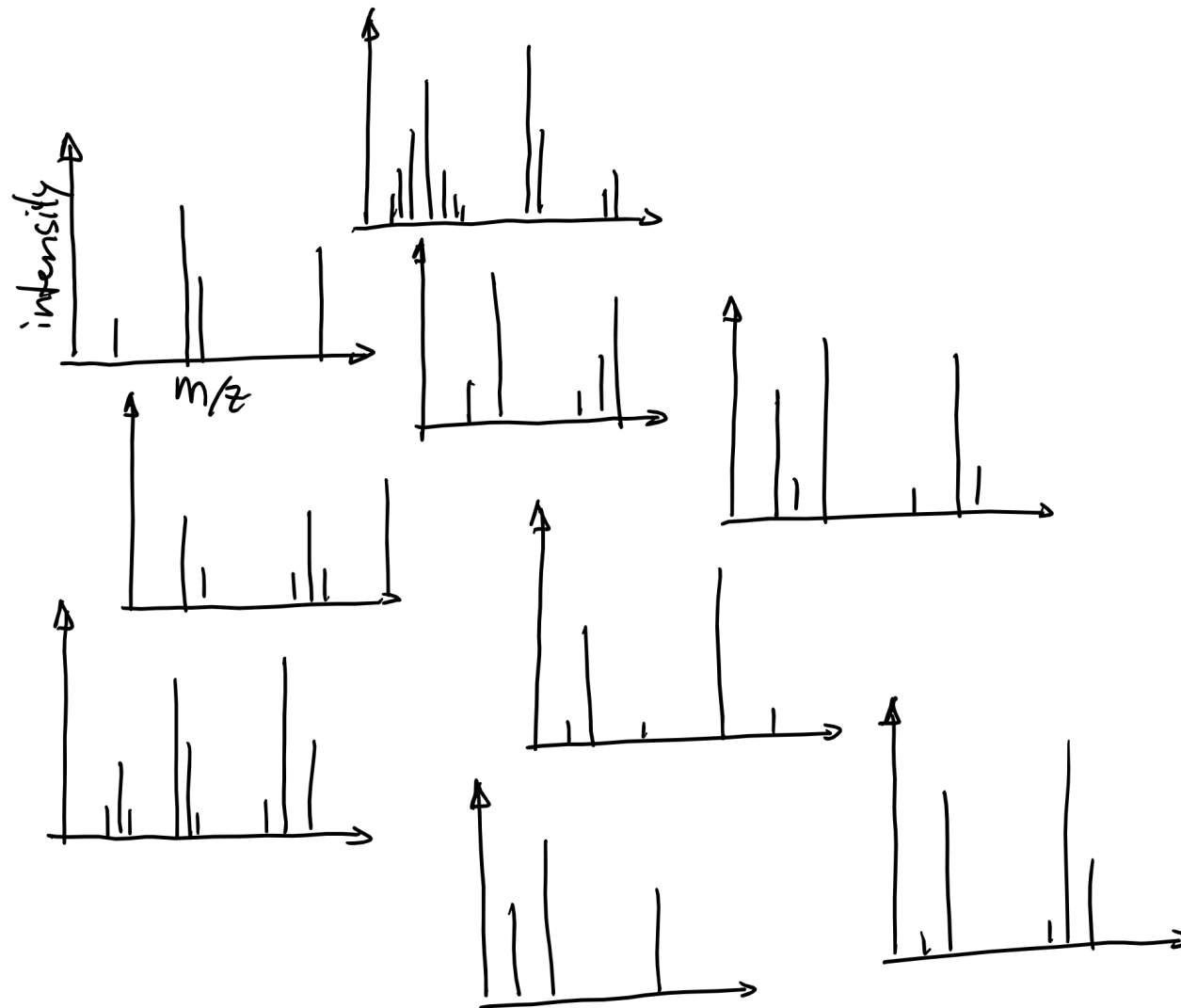
- [Initiative](#) to create a flexible and scalable infrastructure for MS data.
- See Laurent Gatto's [poster](#) for details.

Mass spectrum



- Spectrum:
 - 2 `numeric`: m/z and intensity values.
 - additional metadata information.
- `MSnbase`: `Spectrum` object for a single spectrum.

Mass spectrum



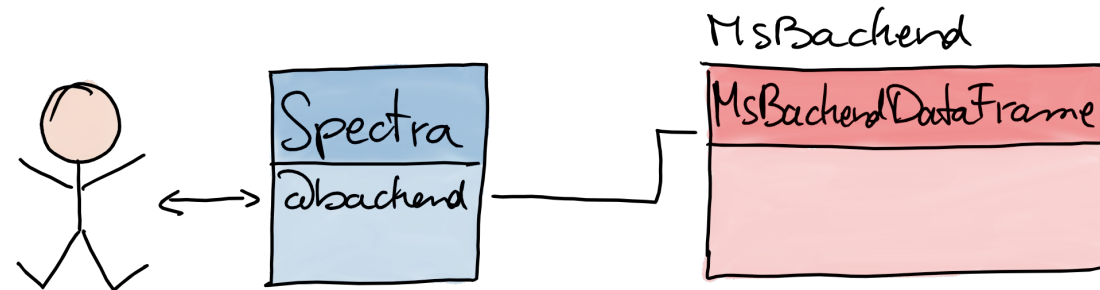
... but usually we deal with many spectra ...

Think bigger: Spectra



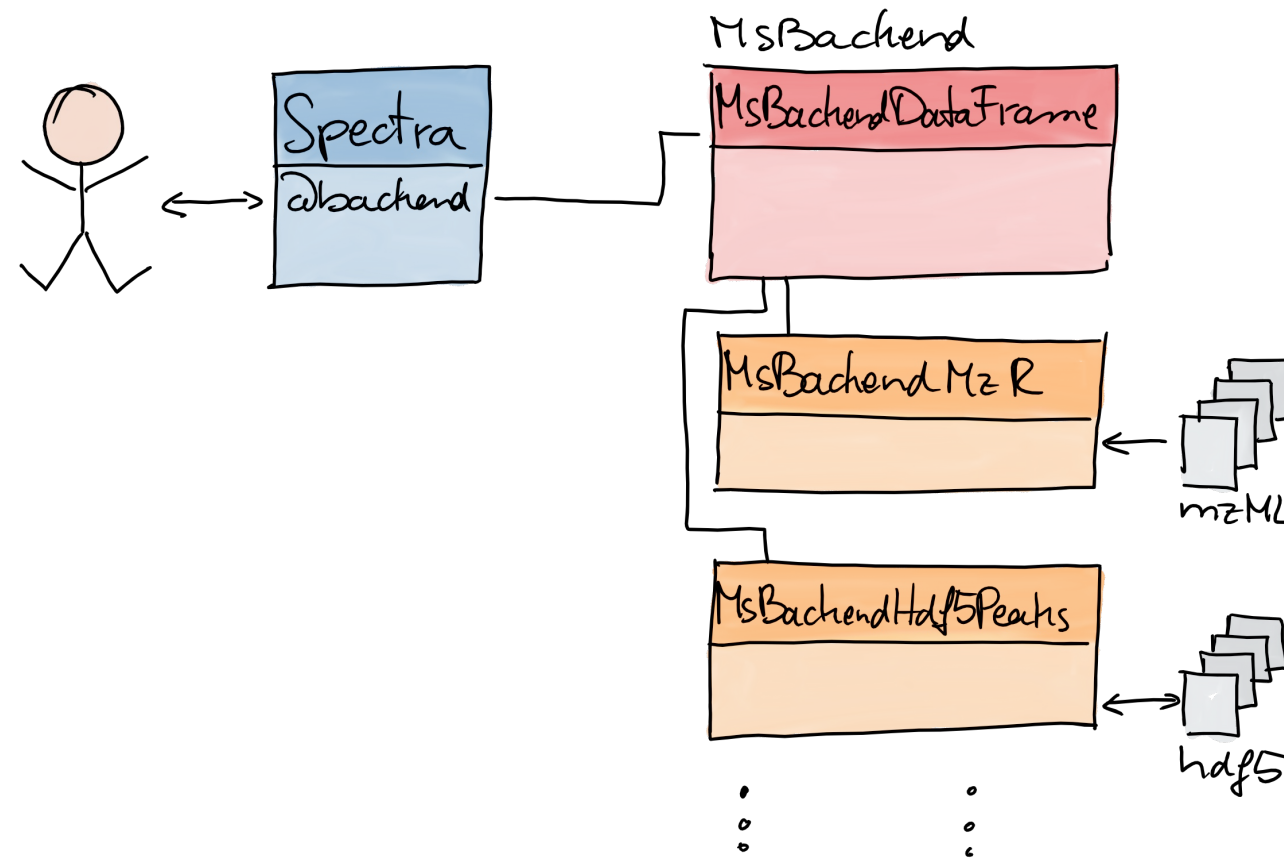
- One object to represent data from one or many spectra.

Think flexible: MsBackend



- Separate user functionality from data handling and storage.

Think flexible: MsBackend



- Separate user functionality from data handling and storage.
- Enables use of different *backends* (in-memory/on-disk, remote files, SQL-based, ...).

Example: data import

- Import data from an *mzML* file.

```
library(Spectra)
library(magrittr)

sps <- Spectra("data/20191107_Mix2_CE20.mzML", backend = MsBackendMzR())
sps
```

```
## MSn data (Spectra) with 1255 spectra in a MsBackendMzR backend:
##      msLevel      rtime scanIndex
##      <integer>    <numeric> <integer>
## 1           1    0.277000000002      1
## 2           1    0.580000000002      2
## ...      ...      ...      ...
## 1254         1 480.325000000002    1254
## 1255         1 480.688999999998    1255
## ... 33 more variables/columns.
##
## file(s):
## 20191107_Mix2_CE20.mzML
## Processing:
##
```

Example: data subsetting

- Select all MS2 spectra for a [M+H]⁺ ion of Histidine.

```
mz_hist <- 156.07675 # Histidine
ms2_hist <- sps %>%
  filterMsLevel(2) %>%
  filterPrecursorMz(mz = mz_hist + ppm(c(-mz_hist, mz_hist), 20))
ms2_hist
```

```
## MSn data (Spectra) with 1 spectra in a MsBackendMzR backend:
##      msLevel      rtime scanIndex
##      <integer> <numeric> <integer>
## 1           2    184.593         489
## ... 33 more variables/columns.
##
## file(s):
## 20191107_Mix2_CE20.mzML
## Processing:
## Filter: select MS level(s) 2 [Sat Dec  7 18:09:56 2019]
## Filter: select spectra with a precursor m/z within [156.073628465, 156.079871535] [Sat Dec  7 18:09:56 2019]
```

Example: data processing

```
ms2_hist <- ms2_hist %>%  
  pickPeaks() %>%  
  removePeaks(threshold = 500) %>%  
  clean(all = TRUE)  
ms2_hist
```

```
## MSn data (Spectra) with 1 spectra in a MsBackendMzR backend:  
##      msLevel      rtime scanIndex  
##    <integer> <numeric> <integer>  
## 1          2    184.593         489  
## ... 33 more variables/columns.  
##  
## file(s):  
## 20191107_Mix2_CE20.mzML  
## Lazy evaluation queue: 3 processing step(s)  
## Processing:  
## Filter: select MS level(s) 2 [Sat Dec  7 18:09:56 2019]  
## Filter: select spectra with a precursor m/z within [156.073628465, 156.079871535] [Sat Dec  7 18:09:56 2019]  
## Peak picking with MAD noise estimation, hws = 2, snr = 0 [Sat Dec  7 18:09:56 2019]  
## Signal <= 500 in MS level(s) 2 set to 0 [Sat Dec  7 18:09:56 2019]  
## Spectra of MS level(s) 2 cleaned. [Sat Dec  7 18:09:56 2019]
```

- Data manipulations applied *on-the-fly*.

Example: use custom backend

- Import reference spectra from [HMDB](#).

```
library(MsBackendHmdb)
fls <- dir("data/hmdb_all_spectra", full.names = TRUE, pattern = "ms_ms_")
hmdb <- Spectra(fls, source = MsBackendHmdbXml(), nonStop = TRUE)
hmdb
```

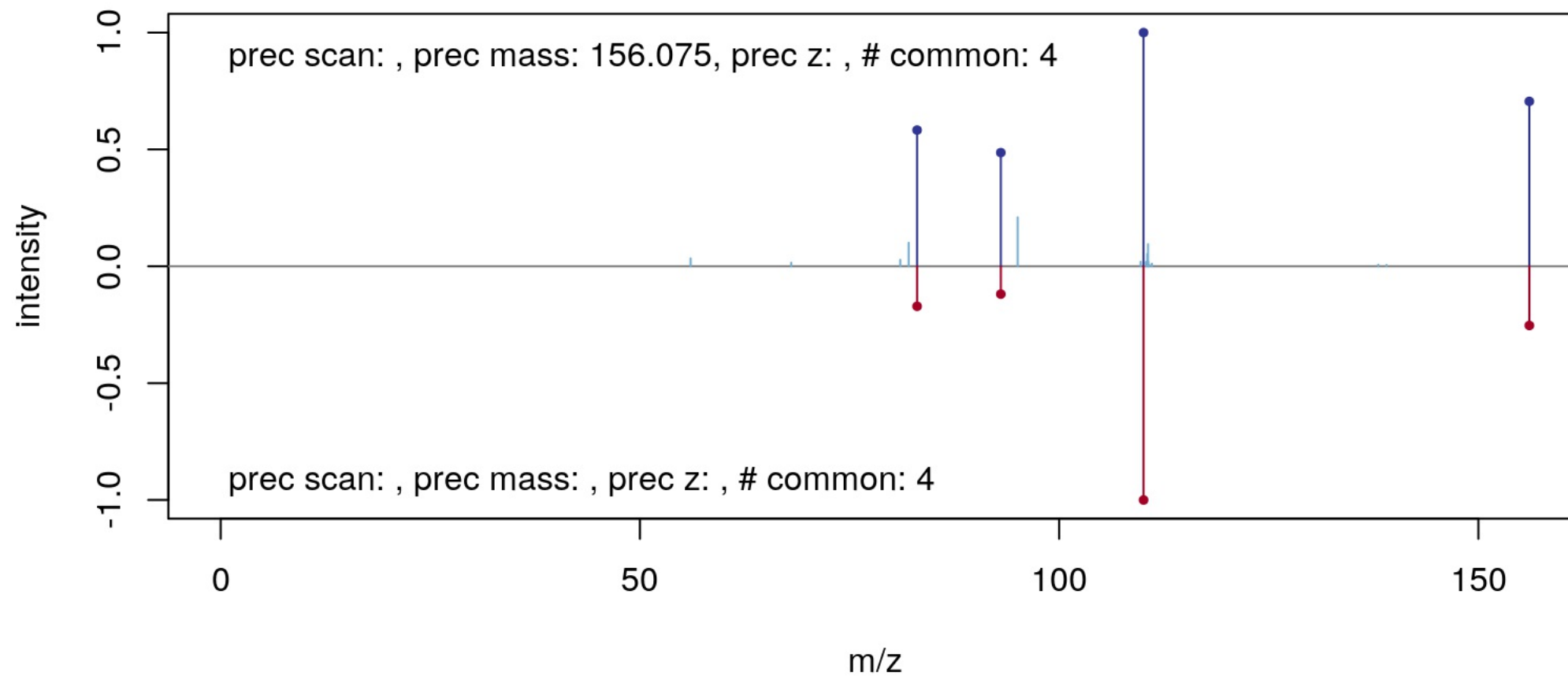
```
## MSn data (Spectra) with 458963 spectra in a MsBackendDataFrame backend:
##           msLevel      rtime scanIndex
##           <integer> <numeric> <integer>
## 1             2         NA         NA
## 2             2         NA         NA
## ...           ...         ...         ...
## 458962         2         NA         NA
## 458963         2         NA         NA
## ... 21 more variables/columns.
## Processing:
## Switch backend from MsBackendHmdbXml to MsBackendDataFrame [Tue Dec 3 16:19:53 2019]
```

Example: compare spectra

- Match spectrum against database.

```
res <- compareSpectra(ms2_hist, hmdb, ppm = 40)  
hmdb$compound_id[res > 0.7]
```

```
## [1] "HMDB0000177" "HMDB0000177"
```



See Sebastian Gibb's poster for details