Practical R

BIOF339

---Final project

Yafeng He Dec, 2018

Loading the raw data

```
library("MALDIquant")
datapath <- "/Users/hey9/course_project/B10/1Ref/fid"

s <- importBrukerFlex(datapath, verbose=FALSE)
summary(mass(s))
```

```
> summary(mass(s))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    700    1280    2033    2149    2960    4060
```

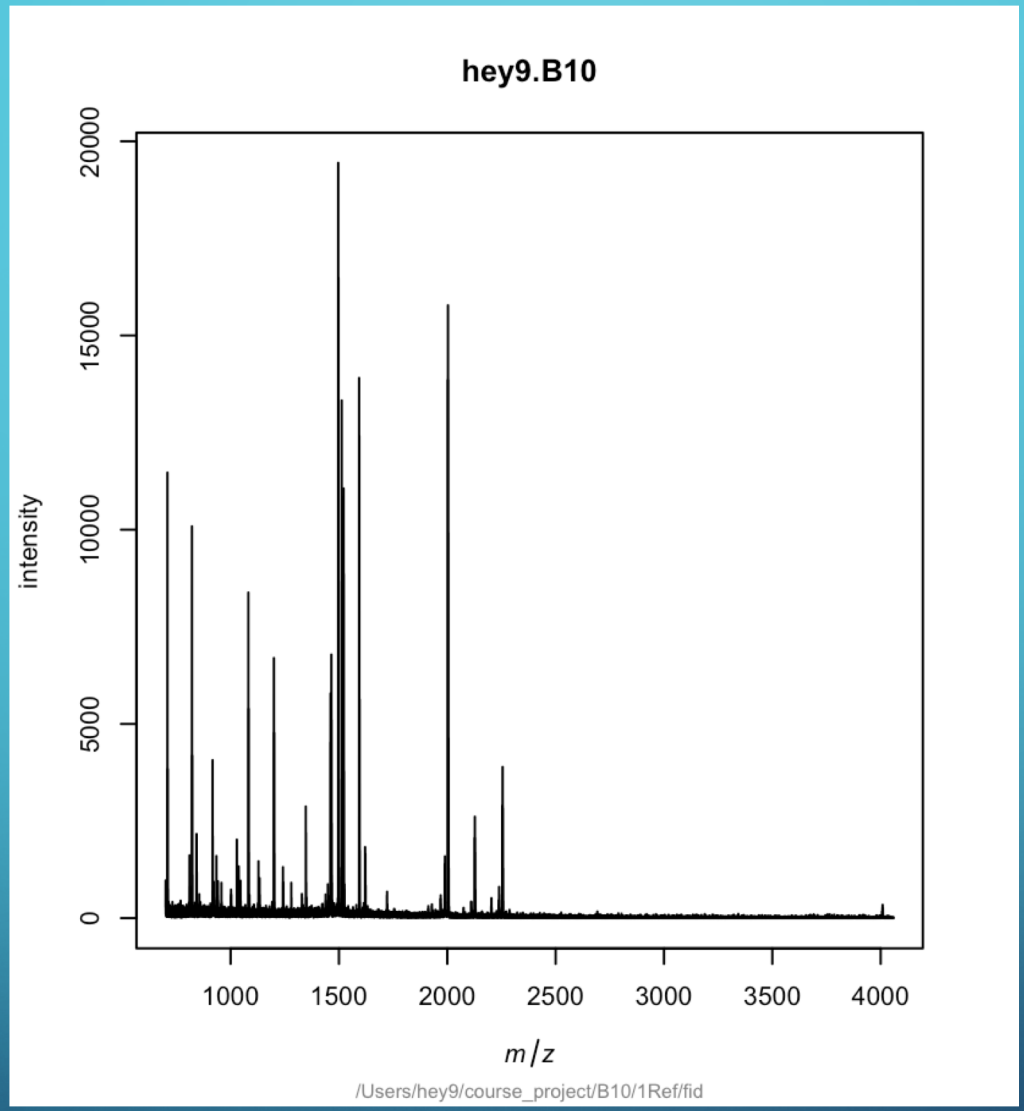
```
summary(intensity(s))
```

```
> summary(intensity(s))
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
    0.00    10.00    48.00    93.76   109.00 19443.00
```

```
head(as.matrix(s))
```

```
> head(as.matrix(s))
         mass intensity
[1,] 699.9932       334
[2,] 700.0158       313
[3,] 700.0383       282
[4,] 700.0609       318
[5,] 700.0835       460
[6,] 700.1060       445
```

```
plot(s)
```


hey9.B10
intensity
0
5000
10000
15000
20000
1000
1500
2000
2500
3000
3500
4000
m/z
/Users/hey9/course_project/B10/1Ref/fid

Transforming data

```
s2 <- transformIntensity(s, method="sqrt")
s2
```

```
> s2
[[1]]
S4 class type            : MassSpectrum
Number of m/z values     : 87373
Range of m/z values      : 699.993 - 4060.017
Range of intensity values: 0e+00 - 1.394e+02
Memory usage             : 1.343 MiB
Name                     : hey9.B10
File                     : /Users/hey9/course_project/B10/1Ref/fid
```

```
s3 <- smoothIntensity(s2, method="MovingAverage",
halfWindowSize=2)
s3
```

```
> s3
[[1]]
S4 class type            : MassSpectrum
Number of m/z values     : 87373
Range of m/z values      : 699.993 - 4060.017
Range of intensity values: 0e+00 - 1.267e+02
Memory usage             : 1.343 MiB
Name                     : hey9.B10
File                     : /Users/hey9/course_project/B10/1Ref/fid
```

plotting

```
s4 <- removeBaseline(s3, method="SNIP")
s4
```

```
> s4
[[1]]
S4 class type            : MassSpectrum
Number of m/z values     : 87373
Range of m/z values      : 699.993 - 4060.017
Range of intensity values: 0e+00 - 1.179e+02
Memory usage             : 1.343 MiB
Name                     : hey9.B10
File                     : /Users/hey9/course_project/B10/1Ref/fid
```

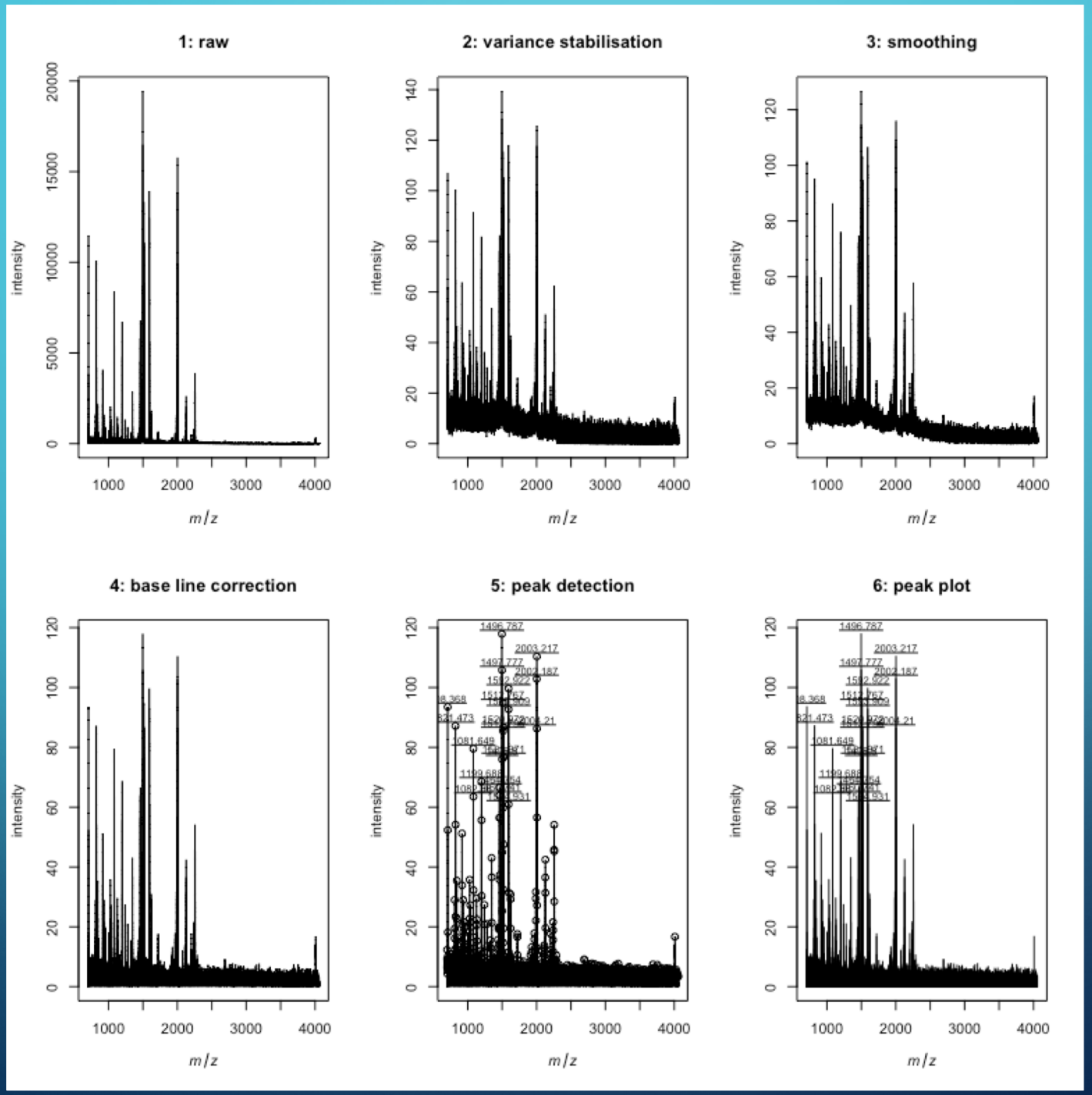
```
p <- detectPeaks(s4)

peak.data <- as.matrix(p)
par(mfrow=c(2,3))
xl <- range(mass(s))

points(p)
top20 <- intensity(p) %in% sort(intensity(p), decreasing=TRUE)[1:20]
summary(top20)
```

```
> summary(top20)
   Mode   FALSE    TRUE
logical    1987      20
```

```
labelPeaks(p, index=top20, underline=TRUE)
plot(p, sub="", main="6: peak plot", xlim=xl)
labelPeaks(p, index=top20, underline=TRUE)
```


1: raw
2: variance stabilisation
3: smoothing
4: base line correction
5: peak detection
6: peak plot
intensity
m/z
1000
2000
3000
4000
1496.787
2003.217
1497.777
2002.187
1512.922
1081.649
1199.688