

Tufte Handout

John Smith

August 13th, 2014

Average MS spectra

Load sample spectra. All from one peak at 14.54 min of the file:

WEB346_A_3_pMS2,HCD_317,303.mzXML.

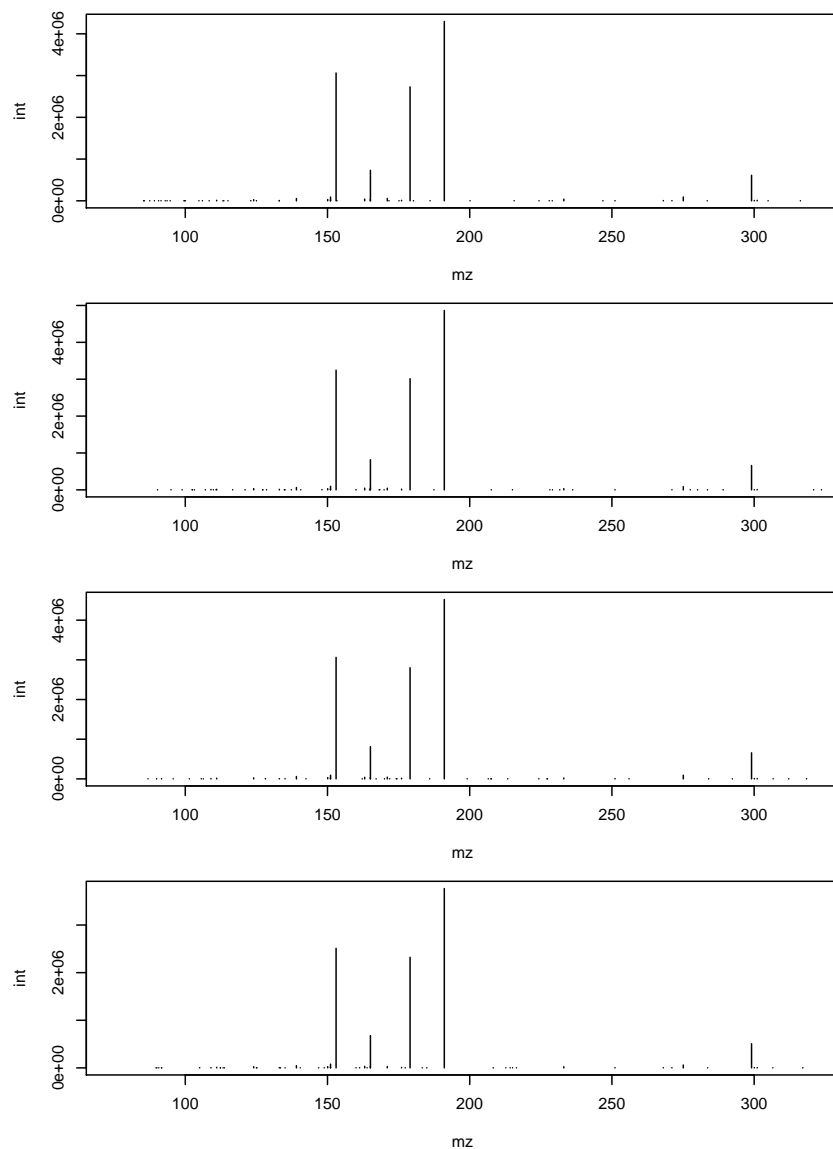


Figure 1: Four MS2 spectra (CID.45.eV@mz.317) of peak at 15.54 min from file WEB346_A_3_pMS2,HCD_317,303.mzXML.

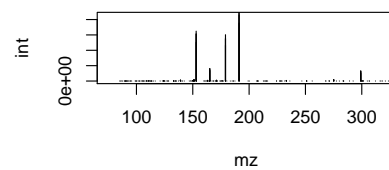


Figure 2: All spectra overlaid on one plot.

Algorithm for averaging

- 1) bin individual spectra in bins with fixed size (e.g. 0.001 m/z)
- 2) create spectramatrix, fill empty bins with 0 (no ions detected means 0 intensity, not NA)
- 3) average the columns to obtain mean spectrum

```

moveAVG <- function(spectra, cuts = 0.001, MS = F) {
  ## calculate bins from spectra and cuts size
  mzmax <- lapply(spectra, function(x) x[, 1]) %>%
    unlist
  mzmax <- mzmax[!is.na(mzmax)]
  mzmin <- mzmax %>% min %>% floor
  mzmax %<>% max %>% ceiling
  breaks <- seq(from = mzmin, to = mzmax, cuts)

  ## bin spectra and combine into matrix
  mat <- lapply(spectra, function(x) {
    n <- .bincode(x[, 1], breaks = breaks,
      include.lowest = T)

    if (MS)
      tmp <- matrix(data = c(breaks, rep(0,
        length(breaks))), ncol = 2, byrow = F) else tmp <- matrix(data = c(breaks, rep(NA,
        length(breaks))), ncol = 2, byrow = F)

    tmp[, 2][n[!is.na(n)]] <- x[, 2][!is.na(n)]
    return(tmp[, 2])
  }) %>% unlist %>% matrix(., nrow = length(spectra),
    byrow = T)

  ## calculate column means
  mat <- colMeans(mat, na.rm = T)
  if (MS)
    IND <- -which(mat == 0) else IND <- -which(is.na(mat))

  ## return spectrum, remove data that is NA
  return(matrix(c(breaks[IND], mat[IND]), ncol = 2))
}

```

test algorithm

Several smoothed spectra show, that a cutsize of 0.001 is sufficient to bin the split peaks at m/z 191.0701 and 153.018.

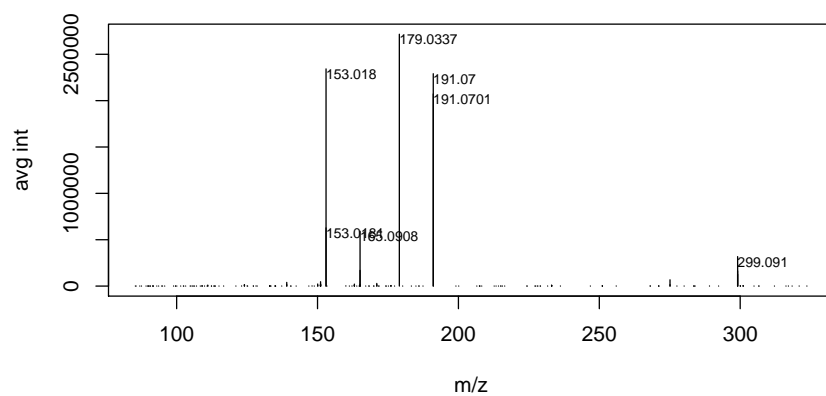


Figure 3: binsize: 0.0001

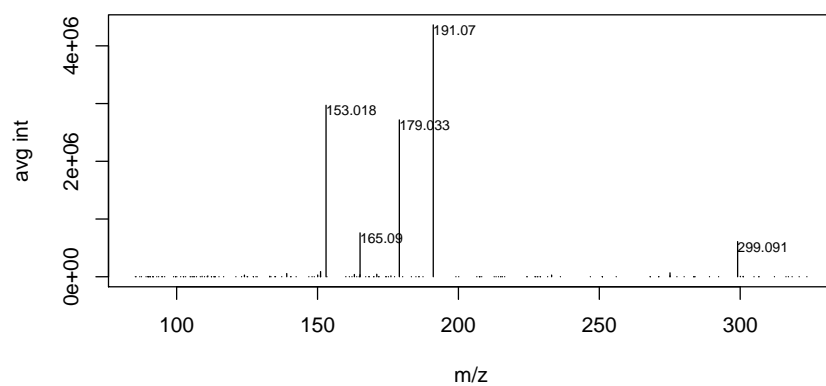


Figure 4: binsize: 0.001

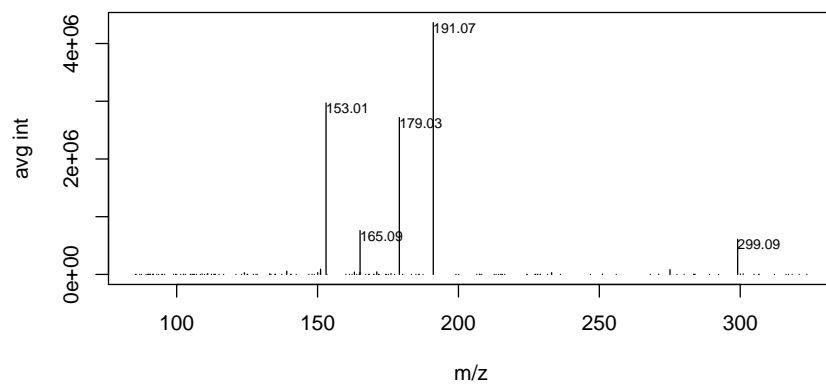


Figure 5: binsize: 0.01

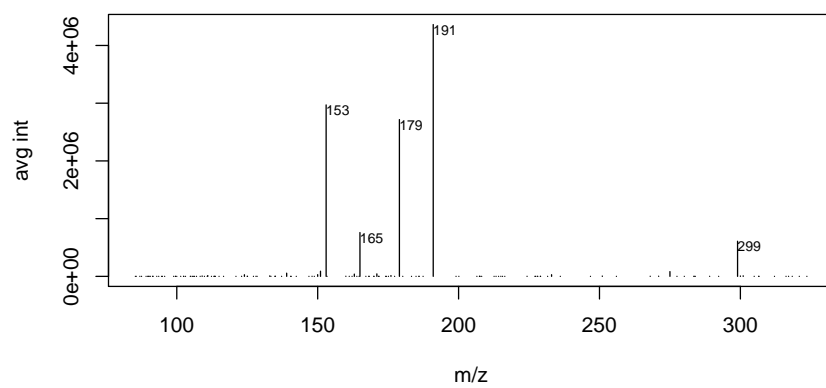


Figure 6: binsize: 0.1