# flowML

Yao-Chung

8/10/2022

# Table of contents

**Preface** **3**

**1 Introduction** **4**
1.1 flowCore intro . . . . . 4
1.2 reading an FCS file into a flowFrame . . . . . 5

**2 Summary** **11**

**References** **12**

# Preface

This book includes the notes I learn to do machine learning on flow cytometry (FC) data.

The source code can be found at https://github.com/ferygood/flowML.

# 1 Introduction

## 1.1 flowCore intro

There is no better way to learn than to do. Here we learn how to analyze FC data using a popular bioconductor package called flowCore.

In this chapter, we will try to cover all the major concepts from How-To-flowCore.

flowCore package is that of using a standardized representation that will insure compatibility with existing technologies for data analysis and will support collaboration and interoperability of new methods as they are developed. flowCore adapts `expressionSet` and `AnnotateDataFrame` structures which are familiar to most Bioconductor users.

In the following articles, I try to cover the main steps of preprocessing: compensation, transformation, filtering.

```
BiocManager::install("flowCore")
```

```
Bioconductor version 3.15 (BiocManager 1.30.18), R 4.2.0 (2022-04-22)
```

```
Old packages: 'bslib', 'callr', 'devtools', 'DT', 'evaluate', 'fontawesome',
  'gert', 'htmltools', 'MASS', 'nlme', 'pillar', 'processx', 'rlang',
  'roxygen2', 'sass', 'shiny', 'stringi', 'survival', 'tibble', 'xfun'
```

Load required library

```
library(flowCore)
```

## 1.2 reading an FCS file into a flowFrame

```
file.name <- system.file("extdata", "0877408774.B08",
                         package="flowCore")
x <- read.FCS(file.name, transformation=FALSE) # default is linearize transformation
summary(x)
```

```
           FSC-H     SSC-H    FL1-H     FL2-H    FL3-H     FL1-A     FL4-H
Min.      85.0000   11.0000   0.0000    0.0000   0.0000    0.0000    0.0000
1st Qu.  385.0000  141.0000 233.0000  277.0000  90.0000    0.0000  210.0000
Median   441.0000  189.0000 545.5000  346.0000 193.0000   26.0000  279.0000
Mean     491.9644  277.9105 439.1023  366.1567 179.7122   34.0766  323.5306
3rd Qu.  518.0000  270.0000 610.0000  437.0000 264.0000   51.0000  390.0000
Max.    1023.0000 1023.0000 912.0000 1023.0000 900.0000 1023.0000 1022.0000
          Time
Min.      1.00
1st Qu. 122.00
Median  288.00
Mean    294.77
3rd Qu. 457.50
Max.    626.00
```

```
head(x)
```

```
     FSC-H SSC-H FL1-H FL2-H FL3-H FL1-A FL4-H Time
[1,]   382    77   618     0   225    55   286    1
[2,]   628   280   245   431   259     0   371    1
[3,]  1023   735   699   448   215   143   638    1
[4,]   373   128   202   354    94     0   149    1
[5,]  1023  1023   618   742   408    61   866    1
[6,]   489   292   179   374   154     0   363    1
```

```
str(x)
```

```
Formal class 'flowFrame' [package "flowCore"] with 3 slots
  ..@ exprs      : num [1:10000, 1:8] 382 628 1023 373 1023 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : NULL
```

```
.. .. ..$ : Named chr [1:8] "FSC-H" "SSC-H" "FL1-H" "FL2-H" ...
.. .. .. ..- attr(*, "names")= chr [1:8] "$P1N" "$P2N" "$P3N" "$P4N" ...
.. ..- attr(*, "ranges")= int [1:8] 1024 1024 1024 1024 1024 1024 1024 1024
..@ parameters :Formal class 'AnnotatedDataFrame' [package "Biobase"] with 4 slots
.. .. ..@ varMetadata      :'data.frame': 5 obs. of  1 variable:
.. .. .. ..$ labelDescription: chr [1:5] "Name of Parameter" "Description of Parameter" "Ra
.. .. ..@ data             :'data.frame': 8 obs. of  5 variables:
.. .. .. ..$ name    : 'AsIs' Named chr [1:8] "FSC-H" "SSC-H" "FL1-H" "FL2-H" ...
.. .. .. .. ..- attr(*, "names")= chr [1:8] "$P1N" "$P2N" "$P3N" "$P4N" ...
.. .. .. ..$ desc    : 'AsIs' Named chr [1:8] "FSC-H" "SSC-H" NA NA ...
.. .. .. .. ..- attr(*, "names")= chr [1:8] "$P1S" "$P2S" "$P3S" "$P4S" ...
.. .. .. ..$ range   : num [1:8] 1024 1024 1024 1024 1024 ...
.. .. .. ..$ minRange: num [1:8] 0 0 0 0 0 0 0 0
.. .. .. ..$ maxRange: num [1:8] 1023 1023 1023 1023 1023 ...
.. .. ..@ dimLabels        : chr [1:2] "rowNames" "columnNames"
.. .. ..@ .__classVersion__:Formal class 'Versions' [package "Biobase"] with 1 slot
.. .. .. .. ..@ .Data:List of 1
.. .. .. .. .. ..$ : int [1:3] 1 1 0
.. .. .. .. ..$ names: chr "AnnotatedDataFrame"
..@ description:List of 147
.. ..$ FCSversion               : chr "2"
.. ..$ $BYTEORD                 : chr "4,3,2,1"
.. ..$ $DATATYPE                : chr "I"
.. ..$ $NEXTDATA                : chr "0"
.. ..$ $SYS                     : chr "Macintosh System Software 9.2.2"
.. ..$ CREATOR                  : chr "CELLQuest<aa> 3.3"
.. ..$ $TOT                     : chr "10000"
.. ..$ $MODE                    : chr "L"
.. ..$ $PAR                     : chr "8"
.. ..$ $P1N                     : chr "FSC-H"
.. ..$ $P1R                     : chr "1024"
.. ..$ $P1B                     : chr "16"
.. ..$ $P1E                     : chr "0,0"
.. ..$ $P2N                     : chr "SSC-H"
.. ..$ $P2R                     : chr "1024"
.. ..$ $P2B                     : chr "16"
.. ..$ $P2E                     : chr "0,0"
.. ..$ $P3N                     : chr "FL1-H"
.. ..$ $P3R                     : chr "1024"
.. ..$ $P3B                     : chr "16"
.. ..$ $P3E                     : chr "4,1"
.. ..$ $P4N                     : chr "FL2-H"
.. ..$ $P4R                     : chr "1024"
```

```
.. ..$ $P4B                    : chr "16"
.. ..$ $P4E                    : chr "4,1"
.. ..$ $P5N                    : chr "FL3-H"
.. ..$ $P5R                    : chr "1024"
.. ..$ $P5B                    : chr "16"
.. ..$ $P5E                    : chr "4,1"
.. ..$ $P1S                    : chr "FSC-H"
.. ..$ $P2S                    : chr "SSC-H"
.. ..$ $P3S                    : chr(0)
.. ..$ $P4S                    : chr(0)
.. ..$ $P6N                    : chr "FL1-A"
.. ..$ $P6R                    : chr "1024"
.. ..$ $P6B                    : chr "16"
.. ..$ $P6E                    : chr "0,0"
.. ..$ TIMETICKS               : chr "50"
.. ..$ $P7N                    : chr "FL4-H"
.. ..$ $P7R                    : chr "1024"
.. ..$ $P7E                    : chr "4,1"
.. ..$ $P7B                    : chr "16"
.. ..$ $P8N                    : chr "Time"
.. ..$ $P8R                    : chr "1024"
.. ..$ $P8E                    : chr "0,0"
.. ..$ $P8B                    : chr "16"
.. ..$ $P8S                    : chr "Time (51.20 sec.)"
.. ..$ SAMPLE ID               : chr "Default Patient ID"
.. ..$ $SRC                    : chr "Default Patient Name"
.. ..$ CASE NUMBER             : chr "Default Case Number"
.. ..$ $CYT                    : chr "FACSCalibur"
.. ..$ CYTNUM                  : chr "E5451"
.. ..$ $BTIM                   : chr "13:13:54"
.. ..$ $ETIM                   : chr "13:14:27"
.. ..$ BD$AcqLibVersion        : chr "3.1"
.. ..$ BD$NPAR                 : chr "7"
.. ..$ BD$P1N                  : chr "FSC-H"
.. ..$ BD$P2N                  : chr "SSC-H"
.. ..$ BD$P3N                  : chr "FL1-H"
.. ..$ BD$P4N                  : chr "FL2-H"
.. ..$ BD$P5N                  : chr "FL3-H"
.. ..$ BD$P6N                  : chr "FL1-A"
.. ..$ BD$P7N                  : chr "FL4-H"
.. ..$ BD$WORD0                : chr "24"
.. ..$ BD$WORD1                : chr "474"
.. ..$ BD$WORD2                : chr "594"
```

```
  .. ..$ BD$WORD3                  : chr "530"
  .. ..$ BD$WORD4                  : chr "700"
  .. ..$ BD$WORD5                  : chr "402"
  .. ..$ BD$WORD6                  : chr "401"
  .. ..$ BD$WORD7                  : chr "401"
  .. ..$ BD$WORD8                  : chr "402"
  .. ..$ BD$WORD9                  : chr "401"
  .. ..$ BD$WORD10                 : chr "299"
  .. ..$ BD$WORD11                 : chr "287"
  .. ..$ BD$WORD12                 : chr "557"
  .. ..$ BD$WORD13                 : chr "0"
  .. ..$ BD$WORD14                 : chr "483"
  .. ..$ BD$WORD15                 : chr "612"
  .. ..$ BD$WORD16                 : chr "546"
  .. ..$ BD$WORD17                 : chr "724"
  .. ..$ BD$WORD18                 : chr "104"
  .. ..$ BD$WORD19                 : chr "100"
  .. ..$ BD$WORD20                 : chr "100"
  .. ..$ BD$WORD21                 : chr "100"
  .. ..$ BD$WORD22                 : chr "100"
  .. ..$ BD$WORD23                 : chr "1"
  .. ..$ BD$WORD24                 : chr "1"
  .. ..$ BD$WORD25                 : chr "0"
  .. ..$ BD$WORD26                 : chr "0"
  .. ..$ BD$WORD27                 : chr "0"
  .. ..$ BD$WORD28                 : chr "136"
  .. ..$ BD$WORD29                 : chr "77"
  .. ..$ BD$WORD30                 : chr "52"
  .. ..$ BD$WORD31                 : chr "52"
  .. ..$ BD$WORD32                 : chr "52"
  .. ..$ BD$WORD33                 : chr "52"
  .. ..$ BD$WORD34                 : chr "6"
  .. ..$ BD$WORD35                 : chr "300"
  .. .. [list output truncated]
```

```
keyword(x, c("$P1E", "$P2E", "$P3E", "$P4E"))
```

```
$`$P1E`
[1] "0,0"

$`$P2E`
```

```
[1] "0,0"

$`$P3E`
[1] "4,1"

$`$P4E`
[1] "4,1"
```

The default "linearize" transformation option will convert these to, effectively, have a $PnE value of "0,0":

```
summary(read.FCS(file.name))
```

```
            FSC-H     SSC-H      FL1-H      FL2-H       FL3-H     FL1-A
Min.      85.0000   11.0000    1.00000    1.00000    1.000000    0.0000
1st Qu.  385.0000  141.0000    8.13123   12.07901    2.246790    0.0000
Median   441.0000  189.0000  135.16485   22.46790    5.674221   26.0000
Mean     491.9644  277.9105  157.79417  105.98637    8.464880   34.0766
3rd Qu.  518.0000  270.0000  241.44182   50.93675   10.746078   51.0000
Max.    1023.0000 1023.0000 3651.74127 9910.45856 3278.121151 1023.0000
            FL4-H   Time
Min.      1.00000   1.00
1st Qu.   6.61169 122.00
Median   12.29826 288.00
Mean    140.39784 294.77
3rd Qu.  33.37625 457.50
Max.   9821.71889 626.00
```

scale transformation

```
summary(read.FCS(file.name, transformation="scale"))
```

```
             FSC-H      SSC-H        FL1-H       FL2-H        FL3-H      FL1-A
Min.    0.08308895 0.01075269 0.0000000000 0.000000000 0.0000000000 0.00000000
1st Qu. 0.37634409 0.13782991 0.0007131943 0.001108012 0.0001246915 0.00000000
Median  0.43108504 0.18475073 0.0134178268 0.002147005 0.0004674689 0.02541544
Mean    0.48090362 0.27166227 0.0156809848 0.010499687 0.0007465626 0.03331046
3rd Qu. 0.50635386 0.26392962 0.0240465869 0.004994175 0.0009747053 0.04985337
Max.    1.00000000 1.00000000 0.3651106383 0.991044961 0.3277448896 1.00000000
              FL4-H         Time
```

```
Min.    0.0000000000 0.0009775171
1st Qu. 0.0005612251 0.1192570870
Median  0.0011299392 0.2815249267
Mean    0.0139411784 0.2881427175
3rd Qu. 0.0032379485 0.4472140762
Max.    0.9821701062 0.6119257087
```

Another parameter of interest is the alter.names parameter, which will convert the parameter names into more "R friendly" qauivalents, usually by replacing "-" with ".";

```
read.FCS(file.name, alter.names=TRUE)
```

```
flowFrame object '0877408774.B08'
with 10000 cells and 8 observables:
      name               desc     range  minRange  maxRange
$P1  FSC.H              FSC-H      1024   0.00000      1023
$P2  SSC.H              SSC-H      1024   0.00000      1023
$P3  FL1.H                 NA      1024   1.00904     10000
$P4  FL2.H                 NA      1024   1.00904     10000
$P5  FL3.H                 NA      1024   1.00904     10000
$P6  FL1.A                 NA      1024   0.00000      1023
$P7  FL4.H                 NA      1024   1.00904     10000
$P8   Time Time (51.20 sec.)      1024   0.00000      1023
164 keywords are stored in the 'description' slot
```

# 2 Summary

In summary, this book has no content whatsoever.

# References