# Module 2 code: Exploring FCM data in R
# Author: Radina Droumeva

# First, a short refresher to R
# Variable Assignment
x <- 5
y <- 10
x + y

# Vectors (one dimensional) and Matrices (two dimensional)
x <- c(1, 4, 5)
x
y <- seq(from = 1, to = 3, by = 1)
y
x + y

# Get help on how to construct a matrix
?matrix
A <- matrix(c(1, 2, 3, 11, 12, 13), nrow = 2, ncol = 3, byrow = TRUE)
A
rbind(x, y)
A + rbind(x, y)
A[1, 3]
A[2, ]
A[, 1]

# Lists and names
mylist <- list(`first` = x, `second` = y)
mylist
mylist[[1]]
mylist[["first"]]
mylist[[3]]
length(mylist)
length(x)
dim(A)

# More advanced functionality: which, intersect, union
?rnorm
a <- rnorm(20)
a
which(a > 0)
a[1]
a[which(a > 0)]
which(a < -1)
intersect(which(a > 0), which(a < -1))
combined <- union(which(a > 0), which(a < -1))
combined
length(combined)

# Simple plotting
plot(a, col = "red")
plot(density(a))
hist(a)
plot(density(rnorm(1000)))

#########################################################

# Flow Cytometry data
# Look at the help files to search for a function:
?read.FCS
??read.FCS
# Load the package which extends the functionality of R to work with flow
... data
library(flowCore)
# Make sure R knows which directory our data will be read from
getwd()
setwd('/home/rguru/Documents/Workshop/data')
dir()
dir('fullFCS/')
# Read an FCS file
f <- read.FCS('fullFCS/100715.fcs')
f
# 'f' is a flowFrame object. See ?flowFrame for details and to see what
... you can do with it
# how many events the file has
nrow(f)
# the channel names:
colnames(f)

# Extract the expression values into a matrix
E <- exprs(f)

# The expression values are like a matrix -- each cell has a row of
# measurements - one for each channel. Here are the first 10 cells:
E[1:10,]

# Explore the meta data stored within the FCS file
f@description

names(f@description)
f@description$`TUBE NAME`
f@parameters@data
f@parameters@data[1, c("minRange", "maxRange")]

# Try a simple plot -- note the error R gives you. It says that you have
# to first load the 'flowViz' library before you can plot FCM files.
plot(f, c("FSC-A", "SSC-A"))

library(flowViz)

plot(f, c("FSC-A", "SSC-A"), ylim = c(0, 5000), smooth=FALSE)

# Note SSC-A is the third parameter (P3) and the meta data tells us it is
# to be viewed on a LOG scale:
colnames(f)[3] # See that this is SSC-A
f@description$`P3DISPLAY`

# Now read a flow set
fs <- read.flowSet(path = 'fullFCS', pattern = "*.fcs")

# You can see sample names as well as the channel names
sampleNames(fs)

length(fs)

colnames(fs)

# A flowSet object is similar to a list, a list of flowFrames
fs[["100715.fcs"]]

fs[[1]]

# Use fsApply to get cell counts for all samples
nrow(fs[[1]])

fsApply(fs, nrow)
# Use fsApply to extract the TUBE NAME keyword in all samples

```r
fsApply(fs,
    function(f) f@description$`TUBE NAME``)
```

### Plotting exercise

```r
plot(fs[[2]], c("FSC-A", "SSC-A"), ylim = c(0, 5000), smooth = FALSE)
```

# Plot the density of the forward scatter area values for the first sample:

```r
E <- exprs(fs[[1]])
fscValues <- E[, "FSC-A"]
fscValues[1:10] 
plot(density(fscValues))
```

# We can plot all 3 samples on one plot:

```r
par(mfrow = c(3, 1)) # This creates a plot region with a single column of 3 subplots
plot(fs[[1]], c("FSC-A", "SSC-A"), main = sampleNames(fs)[1], ylim = c(0, 5000), smooth=FALSE)
plot(fs[[2]], c("FSC-A", "SSC-A"), main = sampleNames(fs)[1], ylim = c(0, 5000), smooth=FALSE)
plot(fs[[3]], c("FSC-A", "SSC-A"), main = sampleNames(fs)[1], ylim = c(0, 5000), smooth=FALSE)
```