***Path218 Homework 4:***

Problem set philosophy for the Path 218.

0. Don't panic *(citation: Douglas Adams... "A Hitchhikers Guide to the Galaxy)*

1. The goal of this problem set is to provide practice in coding and dealing with tabular data

2. If you fully succeed in whatever time you allot to this, wonderful (and tell us the answers)

3. If you don't fully succeed in whatever time you allot, send us what you have, pasting answers and code into a copy of this document as far as you get. If you're code isn't working, add a few comment lines explaining what you think is going wrong (you can also paste any 'error' statement after the code.

4. The problems are designed to be addressed using coding tools from the 7/16 lecture and 7/21 discussion sections. But if you are familiar or want to use other *Pythonic* tools not covered in the lectures/handouts/sections, please feel free to use these (there are many solutions to any problem with Python... we're happy with any of them).

5. This homework will be due Monday, August 3rd at 9:42pm. Email your homework (as a modified version of this file) to **path218homework@gmail.com. Please title the file as follows: "YourName\_PS4.docx"**

We'll start with a program slightly modified from the lecture on 07/16:

**QuadGate.py** uses data from sorting of individual cells labeled with a variety of molecular probes (Behbehani GK, Bendall SC, Clutter MR, Fantl WJ, Nolan GP., (2012) to generate "map" of cell population as a function of CD4 and CD8 levels. A few extras have been added to the original script from the lecture. First, two lines are drawn (a horizontal line at CD4 = 10 and a vertical line at CD8 = 20), separating the plot into 4 quadrants. In flow cytometry, this demarcation is known as a quadrant “gate", where a gate determines the boundary of a cell population. Second (to clean up the data a bit) we’ve modified the code to ignore cells with CD4 or CD8 values of <0.1. We'll focus on two populations (with the cutoffs being CD4=10 and CD8=20):

i) CD4-CD8+ (upper left), expressing CD8 but not CD4 and

ii) CD4+CD8- (bottom right) expressing CD4 but not CD8.

## QuadGate.py: Compare CD4 to CD8 signals from a group of cells from NormalMarrowFlow.txt, plot a heat map

from VSG\_Module import \*

class cell:

def \_\_init\_\_( self, H , V ):

for i in range( len(H) ):

setattr( self, H[i] , max(0.1 , float(V[i])) )

F=open('NormalMarrowFlow.txt',mode='rU')

Header\_List=F.next().strip().split('\t')

D1={}

for L0 in F:

L1=L0.strip().split('\t')

c=cell(Header\_List,L1)

if c.CD3>5.0 and c.CD4>0.1 and c.CD8>0.1:

x\_key=int(log(c.CD4)\*10) ## Calculating the integer part of log(x)\*10 for each x splits x values into bins,

y\_key=int(log(c.CD8)\*10)

if not((x\_key,y\_key) in D1):

D1[(x\_key,y\_key)]=0

D1[(x\_key,y\_key)]+=1

for (x\_key,y\_key) in D1:

xp=e\*\*(x\_key/10.0) ## This arithmetic converts the index value x\_key back into the original (pre-log) x value for each bin

yp=e\*\*(y\_key/10.0)

vrect( xc=10\*x\_key , yc=10\*y\_key, r=5 , fill=D1[(x\_key,y\_key)] , xg=xp , yg=yp)

vgrid( gylog=True, gxlog=True, gxlabel='CD4', gylabel='CD8', gtitle='QuadGate.py')

## plot some lines for the desired quadrant gate

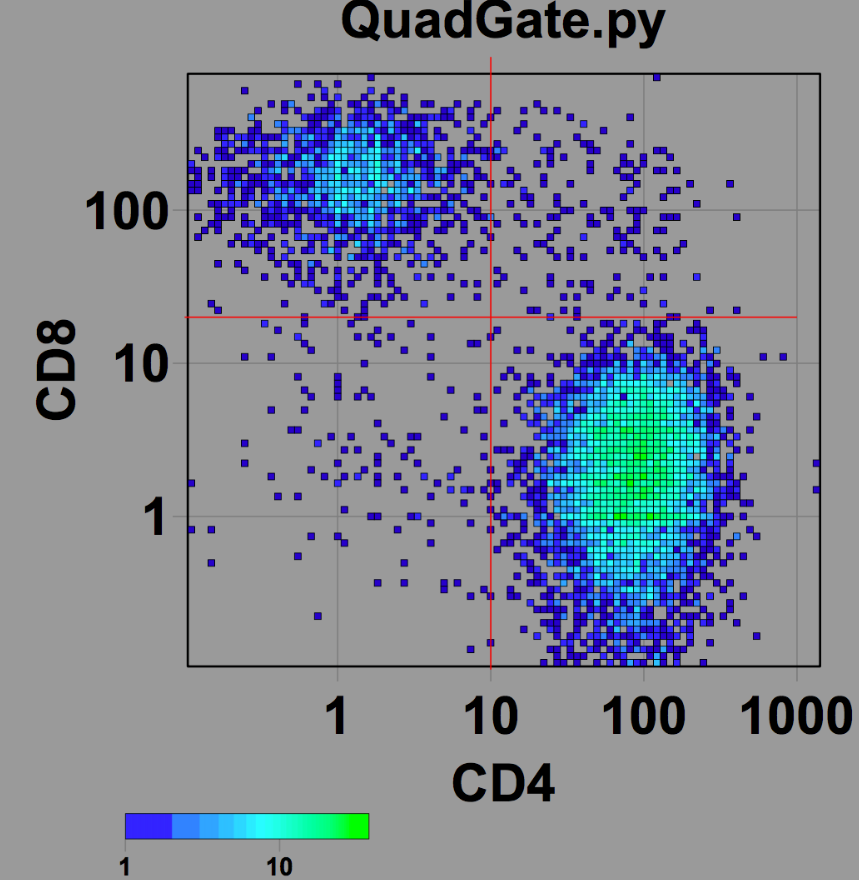
vline(x1=100\*log(0.1),y1=100\*log(20),x2=100\*log(1000),y2=100\*log(20),stroke=red, strokewidth=2)

vline(x1=100\*log(10),y1=100\*log(0.1),x2=100\*log(10),y2=100\*log(1000),stroke=red, strokewidth=2)

vcolorkey() ## vcolorkey() displays something like a color key for color values that have been set using VSG's numerical spectrum

vdisplay()

F.close()



**1a) Determine the number of cells in the two populations**

## Paste your code here

CD4+CD8- cells: \_\_\_\_ [Hint.. to check if you are on the right track, this should be a four digit number ending in 95]

CD4-CD8+ cells: \_\_\_\_

**1b)** For the CD4+CD8- and CD4-CD88+ populations, find the mean value of the CD4 and CD8 markers in both. Mark the position of the mean on the heat map (e.g. with a red circle of radius 10).

## Paste your code here

CD4+CD8- cells: mean CD4=\_\_\_\_ mean CD8=\_\_\_\_

CD4-CD8+ cells: mean CD4=\_\_\_\_ mean CD8=\_\_\_\_

**1c)** The strong benefit of Cytof is the ability to measure many attributes at the same time. Let’s explore whether there are any additional markers that differentiate the CD4+CD8- and CD4-CD8+ populations. For each column of data print (i) the mean value for the CD4+CD8- cells, the mean value for CD4-CD8+ cells, and the ratio between these two means.

• A tip here (and in 1b)... the function **getattr(*obj*,"*attr"*)** gets a specific attribute ***attr*** by name for object ***obj***.

## Paste your code here

## Paste your results here

**Attribute CD4+CD8- mean CD4-CD8+ mean ratio**

**1d)** Based on the results in 1c, choose 4-6 markers you might pursue as potential alternatives to CD4 and CD8 to distinguish the two populations of T cells.

## Paste your vote here