**Additional Background**

**The project idea in non-science terms:**

- Imagine there are different types of balls you want to sort out from a pile.

- What available are different buckets that can collect balls of different colors.

- You can given a subset of paints that you can use to paint each type of balls.

- The goal is to select a different paint for each type of ball so you can separate them into different buckets.

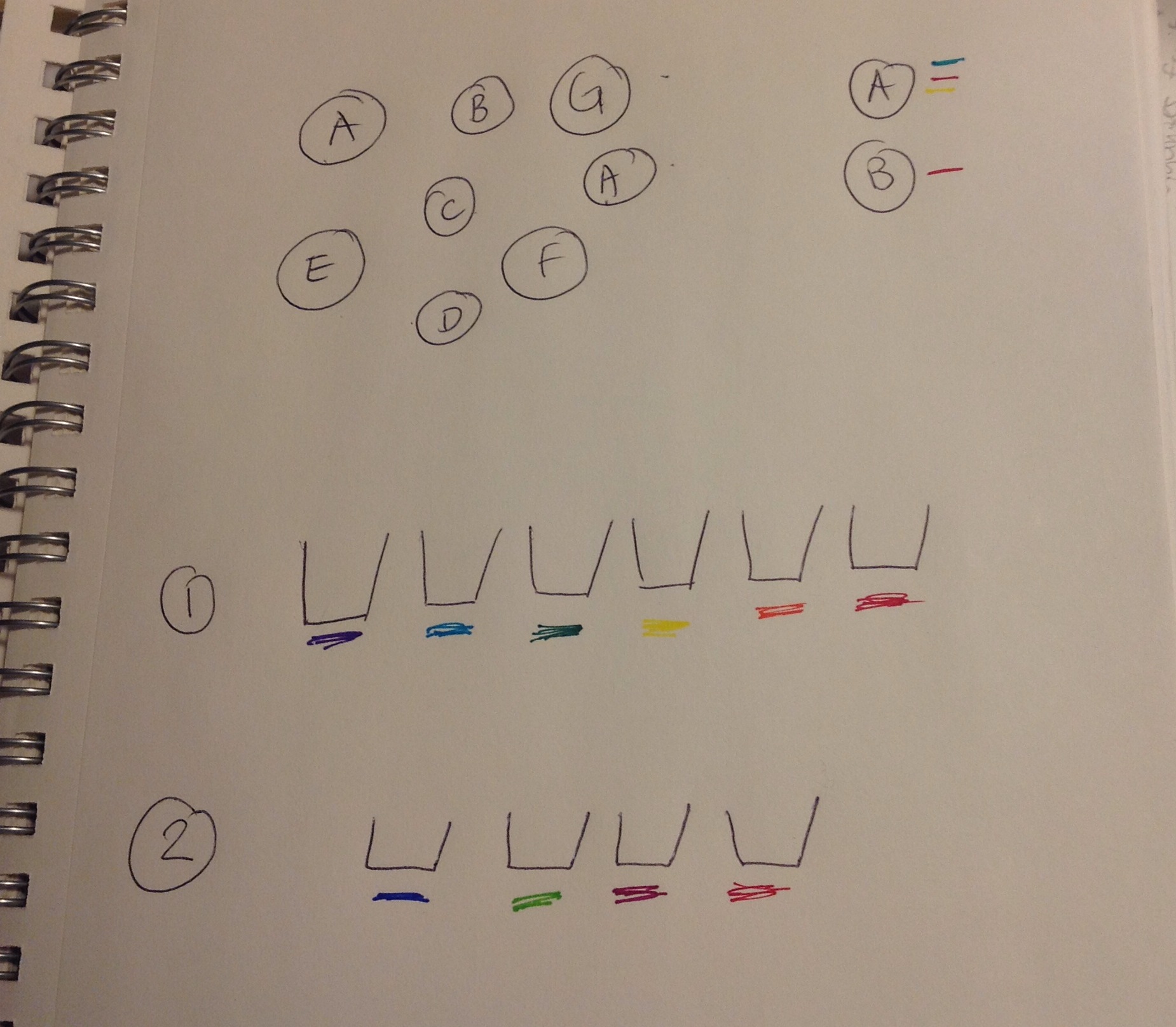
Analogy:

Balls = different types of cells

Paint = colorful antibodies that can label the cells

Colors = fluorescence

Buckets = detection filters of an instrument



Now let’s make it a little more complicated.

- The paint colors sometimes overlap. Ideally we want to choose paints that have minimal overlap.

- The colors are not discrete. Instead they follow different spectra and one color may be picked up by more than 1 bucket.

- There are different rows of buckets. Only certain colors can go into certain rows.

Analogy:

Color overlap = fluorescent signal spillover to another detector

Rows of buckets = detectors off different laser paths. The laser determines how much fluorescent signal is emitted.

**Target Audience:**

Biologists planning a multi-color fluorescent labeling experiment.

The goal is to use visualization to streamline the workflow of selecting antibodies.

**Prior art:**

Major reagent companies include BDBiosciences, Life Technologies, Beckmen Coulter etc.. Typically one would start by going to an antibody catalog to search for the antibodies that are available to label that cell type of interest. The catalogs are typically in table form.

Typical antibody catalogs in table format

- what’s cool: has all the info

- what’s not so cool: no integration with fluorescent spectra and instrument info

BD Biosciences:

<http://www.bdbiosciences.com/nvCategory.jsp?action=SELECT&form=formTree_catBean&item=744882>

Life Technologies:   
<http://www.lifetechnologies.com/us/en/home/life-science/cell-analysis/antibodies-and-secondary-detection/primary-antibodies.html>

One would make a note of the “paints” (fluorescent conjugates) that are available to label the cells. Then one would consult a fluorescence spectral viewer to check the color distributions to identify colors that don’t overlap significantly.

<http://www.lifetechnologies.com/us/en/home/life-science/cell-analysis/labeling-chemistry/fluorescence-spectraviewer.html>

One will also have to determine if the measurement instrument itself has the filters for detection (i.e. does it contains the appropriate collection buckets).

**Proposal Feedback from TF: rancis Kei Masuda**

Background:

I think your project idea is fantastic! Having dealt with the problem you state, I can definitely see the potential for this visualization for your final project (and if you continue to refine it after class it could

possibly be commercialized…). Make sure to expand and elaborate your background/motivation section a bit for the final project process book.

Objectives:  Clear and well laid out.

Data: I would definitely not underestimate the amount of time it will take to scrape and access all of the data you need to put together for this project. The link you provided me for the bdbiosciences website gave me an error message saying that the pertinent section of the website was temporarily down/broken. Hopefully the website will be back up by the time you need this website or you already have scraped this data.

Data Processing:  Again, I would not underestimate the amount of time the data processing will take. You seem to have your steps well thought out and organized, so I’m not worried

:-)

Visualization: I think you have your visualization well thought out and sketched. I am especially a fan of the current detailed view. I’m not 100% sure a stacked line graph is the best option for the overview selector, so I’d make a quick sketch of one or two other options before implementing anything. If you decide that that is your best visualization after a bit more brainstorming, I’m all for it! I’m excited to see your final project!

Must Haves: Solid list. Good job.

Optionals: Excellent list. My hunch at this point is that addition instrument configuration will be beyond the scope of the final project, but I’d definitely consider it if you feel like you have extra time/want to develop this tool for something beyond CS171.

Schedule: Good schedule, but definitely don’t underestimate the amount of time it will take to data scrape/data process.

Some general comments:

Don’t underestimate the amount of time data processing takes!

Look out for an email from me to schedule an appointment for your post-prototype meeting.

**Data Collection:**

In my proposal I planned to use the catalog from BD Biosciences. However I realized I could not find the fluorescent spectra details for many of the patented new dyes that the company released in the past few years. I decided to use the Life Technologies catalog instead for the final project as the fluorescence spectra information is more readily available.

**Catalog Data:**

* Obtain .cvs files from the website. There has a number of item data size limit so I did multiple searched by major fluorophore types to scrap out the data.
* Write Python script to parse the raw .csv data, integrate the info I need for this visualization in JSON

Data structure:

{ Species: {Name: {Clone1: {Conjugate1 (fluorescent): {catalog no: […, …]

target gene: …. ,

clonality: e.g. monoclonal,

reactivity:[…, … , … ].

host:…… }

}

{Conjugate2 (fluorescent): {catalog no: […, …]

target gene: …. ,

clonality: e.g. monoclonal,

reactivity:[…, … , … ].

host:…… }

}

{Clone2:

……………..

Relevant files:

catalogParse.py

catalog.json (JSON with indentation for readability)

catalog-min.json (JSON with minimal whitespace to save space)

I get the set of fluorescence conjugates from Python. The next time is to cross reference them to spectral data.

**Spectral Data:**

I got the data from <http://www.fluorophores.tugraz.at/substance/>

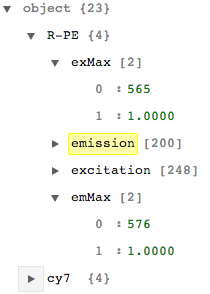
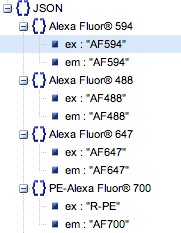
Unfortunately there is no easy way to automatically query for all the fluorescent molecules together. I had to do the search one by one. The raw spectral data were saved in CSV format.

A Python script was written to parse out the excitation and emission max values and integrate all the CSV files into one JSON file (spectra.json /spectra-min.json).

The names of the fluorescent conjugates as obtained from Life Technologies’ website often contains special characters like the trademake symbol. To reference each fluorophore to its spectral data, I manually created a data object that links the trademark names to its name for spectral data. I saved that it as a SON file: fluoro\_spectra\_lookup.json

Among the fluorescent conjugates, some are known as tandem dyes in which 2 fluorescent molecules are linked together as A-B (e.g. APC-Cy7). There was not spectral info available info directly for most of those dyes. However, knowing that the dyes absorb with the first dye and emit with the second dye, their spectral data was a merge of its two components. That info is also integraded in fluoro\_spectra\_lookup.json

Example data structure: (left: fluoro\_spectra\_lookup.json , right: spectra.json)



Each item is an array:

[0]: wavelength

[1]: normalized absorption/emission

Tandem dye example ->

**Instrument Data:**

The instrument configuration file at my workplace is available as an Excel spreadsheet. I will use the data from two instruments in this project: the Aria (5 lasers, 17 fluorescent detectors, analogy: 5 rows of 17 total buckets) and the Astrios (7 lasers, 28 fluorescent detectors, analogy: 7 rows of 28 total buckets). The data in the excel spreadsheet is parsed and saved as aria.json and astrios.json (aria-min.json and astrios-min.json for min size versions).

Data structure format:

{Laser1: {bandpass filter1: [low range, high range],

{bandpass filter2: [low range, high range],

……

Laser2: {bandpass filter1: [low range, high range],

{bandpass filter2: [low range, high range],

**Visualization**

Initial sketch idea (see next page):

1. Overview section to allow users select the targets of interest
   1. Dropdown menu or Entry box to type in the target they want
   2. Selection by hierarchical tree?
   3. View of whole catalog by color? (TF doesn’t think it’s such a great idea).

All will be linked to keyword filtering e.g. human, mouse, immunology, cell surface etc..

Users select targets, which are blown up in detailed view

1. Detailed view section to allow to view panel and spectra

Color code antibodies and spectra according to the visual spectrum.

Panel view:

for each target, how many clones,

for each clones, what colors are available

Spectral viewer:

Select instrument to use

Which laser/filter combinations can detect the color

How much spectral overlap

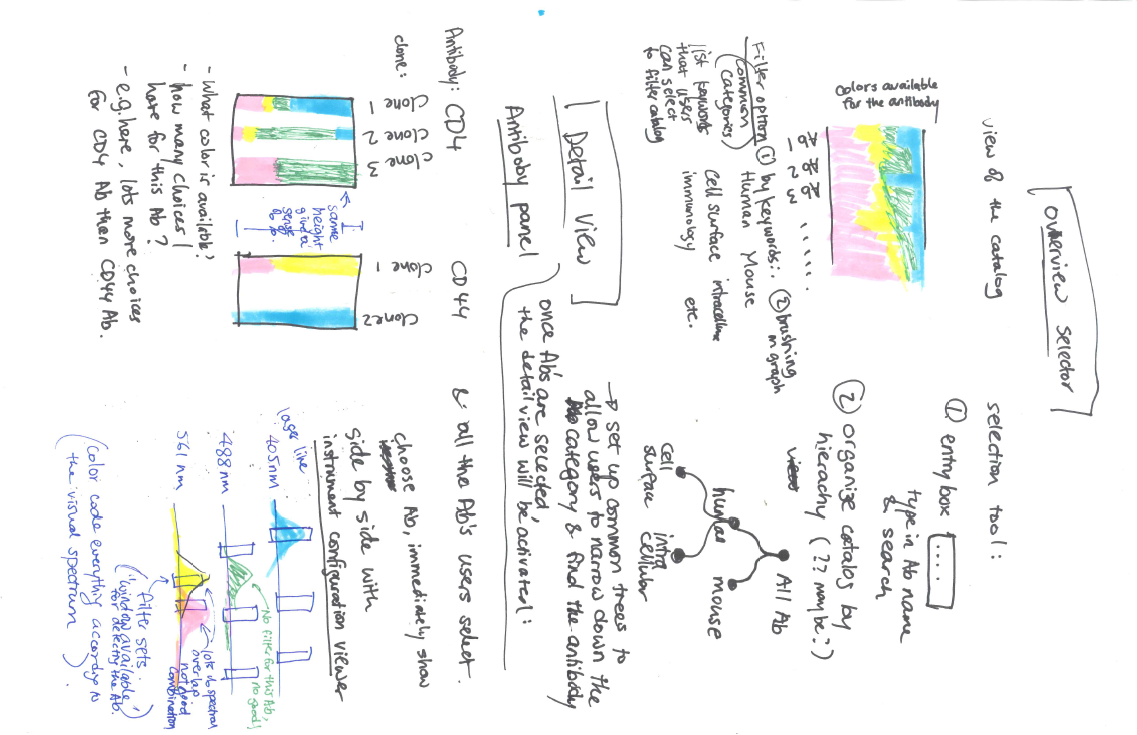
Interaction:

Mouseover to browse

Select antibody

Order of work:

* focus on the detailed view first as this is the meat of the visualization
* the color coding is key, get it to work

asd

Initial sketch for proposal:

**Detailed View**

**Fluorescence spectra view:**

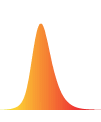
Reference: creating a linear area gradient view

<http://bl.ocks.org/mbostock/1086421>

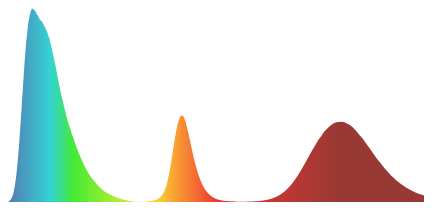
First color spectrum created:



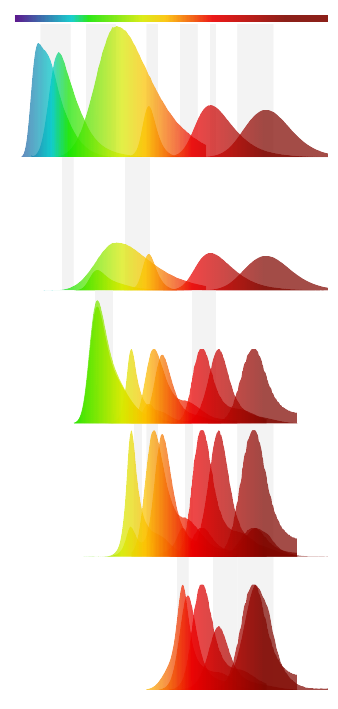
Each spectrum is a clip path drawn on a rectangle



Height is adjusted by the normalized excitation by the laser in each path



Add in filter sets (grey rects), each row is a laser path and its filter. Spectra per path are sorted by size (draw bigger ones first) so every spectrum is reachable by mouse.



Interactions at this point:

- Quick and dirty mouseover tooltip to show name of color

- Mouseover highlights the selected fluorophore and dim the rest (note: one color may be detected in more than my laser path)

**Panel View:**

I will keep the spectral view on the left side and the panel view on the right. The reason is the spectral view will have a constant width but the panel view will depend on the number of antibodies selected.

I am thinking my initial idea of keeping the height of the stacked bar constant may not be the best. My initial thought of keeping main benefit of the same height is to give a sense of the proportion of the different colors –but that info is not critical.