Homework Assignment 5

Question 1: In class we looked at Principle Component Analysis as a way to do unsupervised dimensionality reduction and visualization of complex data. For this week I would like to turn you guys lose on using PCA on a new genetics dataset – again this could be excellent training for your final project.

Lior Pachter, a computational biologist at Cal Tech, wrote an excellent blog post on racism, population genetics, PCA, and the "perfect human". For this week's problem set I want you to:

1. Go and read Lior's blog post (link above)
2. Reproduce the PCA figures in that post. I've downloaded the SNP data and the meta data for you. You can find them in `data/pachter_human_snps.txt` and `data/pachter_meta.txt`. *note:* the first row of the SNP data is for Pachter's "perfect human". If you include that in your PCA plot you should change it's marker or color or somehow indicate that it is different