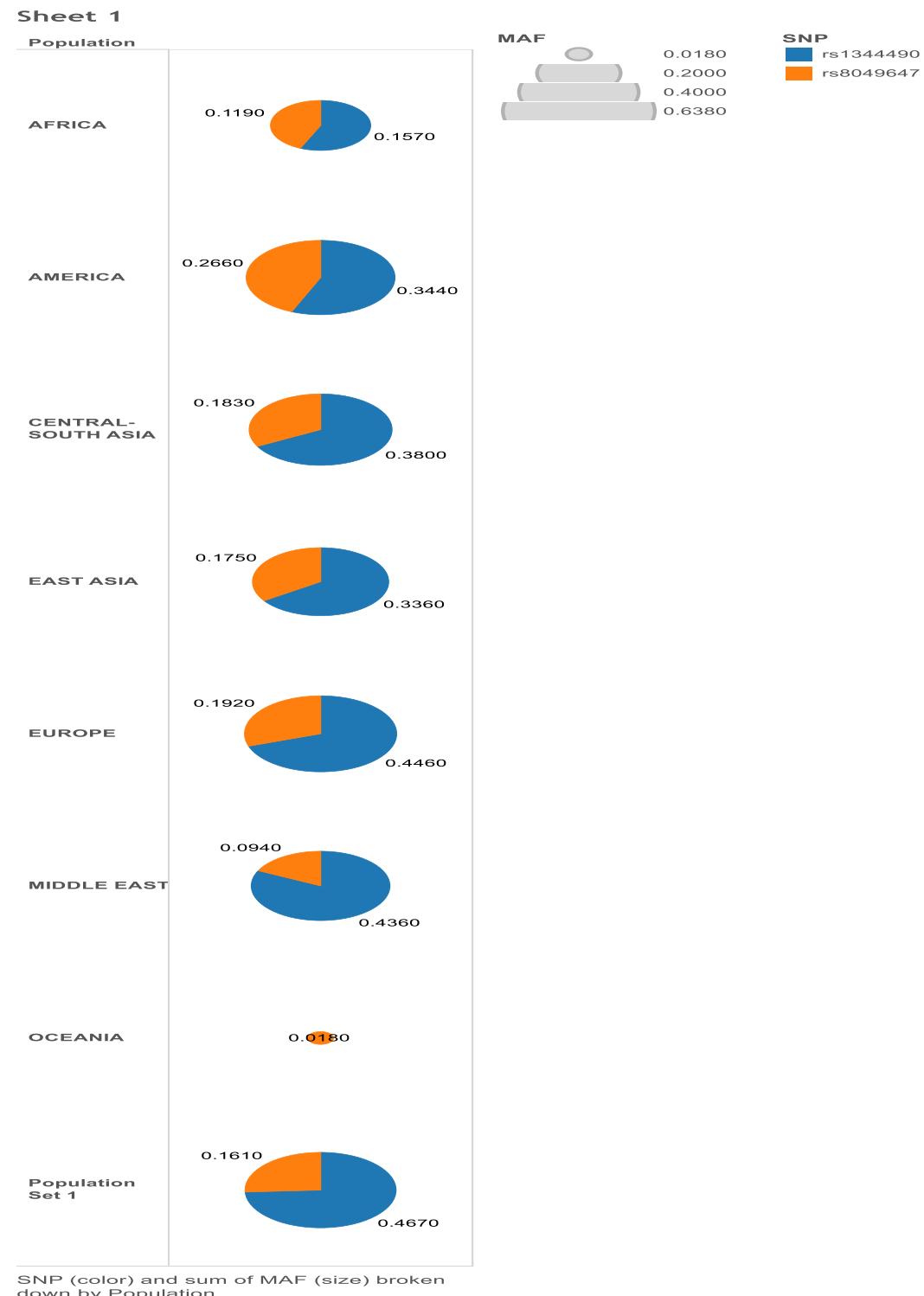


We were given the HGDP Stanford data for this particular visualization assignment. A group of scientist collected 1043 individual genomic DNA samples from all over the world, determining their genotypes at more than 650,000 SNP loci. To visualize such huge amount of data is always difficult, so we choose a subset of the data and try to visualize it. My objective of the assignment was to see which SNPs were associated with Chron's disease and possibly estimate the populations associated with the disease. Chron's disease is an autoimmune disorder of the ileum and the intestines. It is caused when the immune system attacks the cell of the intestines, causing inflammation and usually secondary infection in the secondary tissue. Various information browsers like SPSmart, dbSNP, 1000genomes were used to bridge out the information on which SNPs were associated with the disease and also the possible populations of the HGDP data associated with the disease. NOD2 is a gene involved in Chron's genetic susceptibility. This same gene might reduce innate and adaptive immunity in gastrointestinal tissue and impair the ability to resist infection by the MAP bacterium. For the visualizations I tried using JMP, R studio and Tableau. We first look at the SNPs associated with the disease and also carry out some statistical visualizations.

DESIGN1

The first design was to see the minor allele frequency associated with each SNP of the NOD2 gene among all the continents. This was done in tableau.



DESIGN2

The second design was based on possible HGDP populations that I estimated which are more susceptible to Chron's disease based on information derived from 1000 genomes browser and various other information browser. This was done in Tableau.

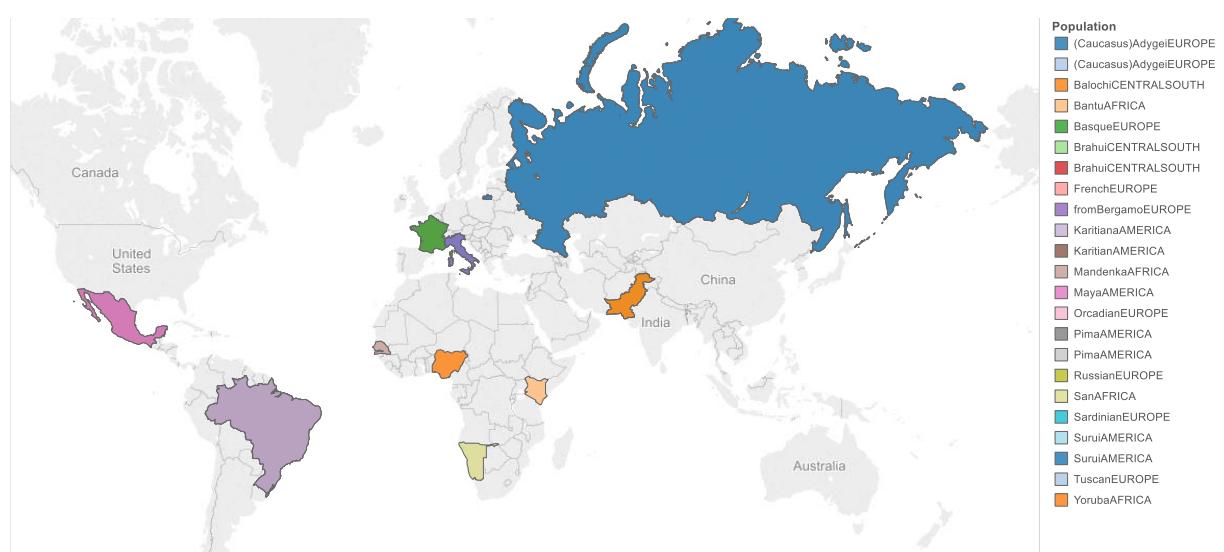
```

C:\Users\Abhi\Downloads\samples (5).txt - Notepad++
File Edit Search View Encoding Language Settings Macro Run Plugins Window ?
results (7).csv samples (5).txt
4 France BasqueEUROPE HGDP01362
5 France BasqueEUROPE HGDP01373
6 France BasqueEUROPE HGDP01363
7 France BasqueEUROPE HGDP01374
8 France BasqueEUROPE HGDP01370
9 France BasqueEUROPE HGDP01366
10 France BasqueEUROPE HGDP01377
11 France BasqueEUROPE HGDP01358
12 France BasqueEUROPE HGDP01360
13 France BasqueEUROPE HGDP01367
14 France BasqueEUROPE HGDP01378
15 France BasqueEUROPE HGDP01368
16 France BasqueEUROPE HGDP01379
17 France BasqueEUROPE HGDP01369
18 France BasqueEUROPE HGDP01380
19 France BasqueEUROPE HGDP01371
20 France BasqueEUROPE HGDP01359
21 France BasqueEUROPE HGDP01364
22 France BasqueEUROPE HGDP01375
23 France BasqueEUROPE HGDP01365
24 France BasqueEUROPE HGDP01357
25 France BasqueEUROPE HGDP01376
26 France FrenchEUROPE HGDP00521
27 France FrenchEUROPE HGDP00533
28 France FrenchEUROPE HGDP00522
29 France FrenchEUROPE HGDP00534
30 France FrenchEUROPE HGDP00511
31 France FrenchEUROPE HGDP00523
32 France FrenchEUROPE HGDP00535
33 France FrenchEUROPE HGDP00512

```

Normal text file length : 13743 lines : 309 Ln : 1 Col : 1 Sel : 0 | 0 UNIX UTF-8 INS

Windows taskbar: Chrome, File Explorer, da-S..., Data..., Exam..., Table..., Table..., C:\Us..., 11:49 AM, 29/5/2016



The problem with this visualization was that it was not a clear representation of the data. A much better visualization was indeed needed.

DESIGN3

A much better visualization was based on a map that had the countries as well as the different population belonging to the country represented as well. A bubble plot representing each of the populations has been represented so that it is clear which country they belong to. This was done in Tableau

Sheet 1



Insights

Keeping in mind Getsalt's laws and the 5 design sheet methodology, my main objective was to create such a map where whoever sees the visualization can understand what the data means and the data can be clear from the cognitive point of view. My main objective was to look at the populations which are more likely to be susceptible to Chrohn's disease and represent them according to the country they belong to. From the map we can conclude that populations in Brazil, Pakistan and Russia are at risk with getting infected with the disease. The results are the best possible estimates of population which have been represented in the HGDP Stanford database.

Screencast

A screencast can be viewed at the following link

<https://www.youtube.com/watch?v=2A7hx-C1pfk>