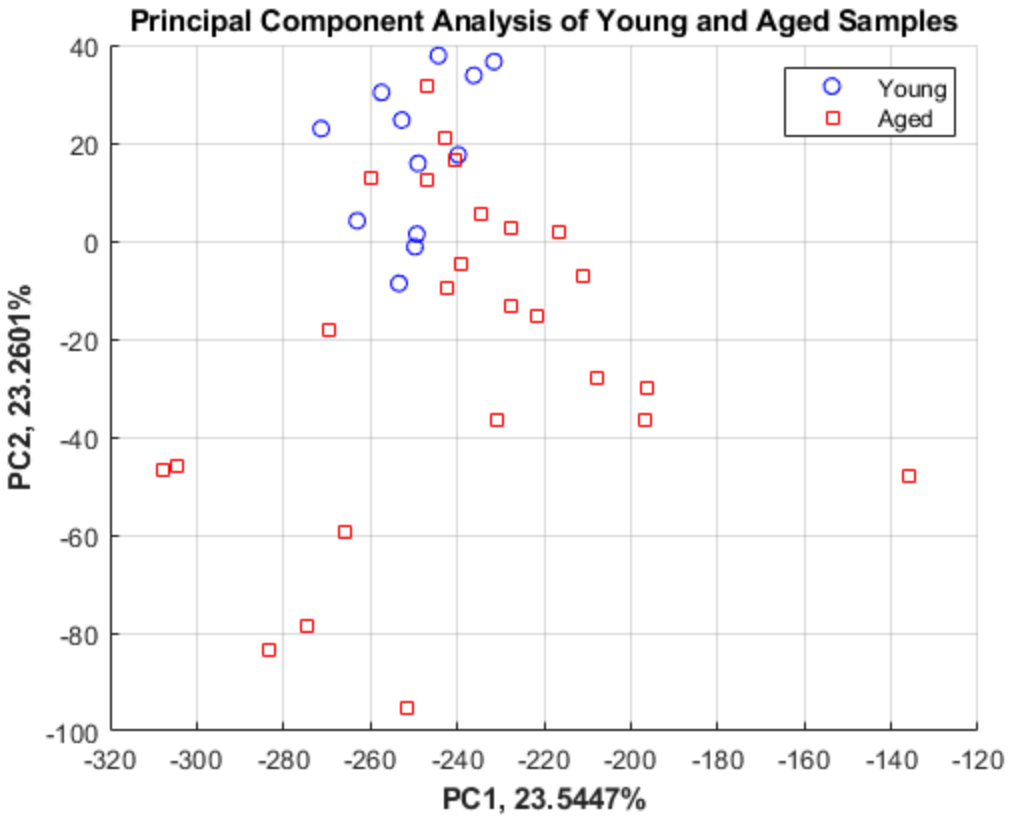
```
% BE 700 A1 Fall 2024
% Final Project, Data Exploration
% Cal Parise, 11/30/2024

brain_data = readtable("Human Brain Data Clean.csv");

gene_id = brain_data.IDENTIFIER;

brain_data_clean = table2array(removevars(brain_data,[1 2]))';
[patients,genes] = size(brain_data_clean);

young = brain_data_clean(1:12,:); % first 12 patients are young adults (<40yo)
aged = brain_data_clean(13:37,:); % next are aged as defined in paper
old = brain_data_clean(38:end,:); % final 4 old aged
categories = [];

[hvals,pvals] = ttest2(young,aged);
zscores = (-1 * log10(pvals))';

genes_sorted = table(gene_id, pvals', zscores);
genes_sorted = sortrows(genes_sorted,2); % array of gene (probe) names sorted
 by p-value and z-score

brain_d_z = [brain_data_clean(1:37,:)' zscores];
brain_d_z = flipud(sortrows(brain_d_z,patients-3));

new_brain = brain_d_z(1:1000,1:patients-4)';

brain_n_centered = zeros(37,1000);
for i = 1:1000
    brain_n_centered(:,i) = ( new_brain(:,i) - mean(new_brain(:,i)) );
end

h = HeatMap(brain_n_centered','Colormap","redgreencmap");
h.addTitle("Relative expression of top 1000 differentially expressed
 genes","FontWeight","bold");
h.addXLabel("Samples","FontWeight","bold");
h.addYLabel("Genes","FontWeight","bold");
c =
 clustergram(brain_n_centered','Cluster","column","Colormap","redgreencmap");
c.addTitle("Hierarchical Clustering of top 1000 differentially expressed
 genes","FontWeight","bold");
c.addXLabel("Samples","FontWeight","bold");
c.addYLabel("Genes","FontWeight","bold");

[coeff,~,~,~,explained] = pca(brain_data_clean);

brain_projected = brain_data_clean * coeff;

young_proj = brain_projected(1:12,:);
aged_proj = brain_projected(13:37,:);
```

```
figure(2)
hold on
scatter(young_proj(:,1),young_proj(:,2),"bo");
scatter(aged_proj(:,1),aged_proj(:,2),"rs");

title("Principal Component Analysis of Young and Aged
 Samples","FontWeight","bold");
xlabel("PC1, "+num2str(explained(1))+"%","FontWeight","bold")
ylabel("PC2, "+num2str(explained(2))+"%","FontWeight","bold")
legend("Young","Aged");
grid on
hold off
```


Clustergram 2
File Tools Desktop Window Help
lierarchical Clustering of top 1000 differentially expressed gene
Genes
Samples



Published with MATLAB® R2023a