```
% BE 700 A1 Fall 2024
% Final Project, Data Exploration
% Cal Parise, 11/29/2024

clear vars, close all

diab_geneexp = readtable("Insulin Resistance Gene Expression.csv");
diab_pheno = readtable("Insulin Resistance Phenotypes.csv");

genes_ordered = diab_geneexp.Var1(2:end);
si = diab_pheno.si_ch1;
diab_geneexp_clean = table2array(removevars(diab_geneexp,"Var1"));
diab_geneexp_clean = diab_geneexp_clean';
[patients,genes] = size(diab_geneexp_clean);

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

diab_principal_components = diab_geneexp_clean * coeff;

fam_hist_categories = zeros(50,1);

for i = 1:patients

    switch string(diab_pheno.familyHistory_ch1{i})

        case 'Family history negative'
            fam_hist_categories(i) = 1;
        case 'Family history positive - 1 parent'
            fam_hist_categories(i) = 2;
        case 'Family history positive - 2 parents'
            fam_hist_categories(i) = 3;
        case 'DM'
            fam_hist_categories(i) = 4;
    end

end

geneexp_labeled = [diab_geneexp_clean fam_hist_categories];
fam_hist_negative = diab_principal_components((geneexp_labeled(:,end) ==
 1),1:2);
fam_hist_one = diab_principal_components((geneexp_labeled(:,end) == 2),1:2);
fam_hist_two = diab_principal_components((geneexp_labeled(:,end) == 3),1:2);
diabetics = diab_principal_components((geneexp_labeled(:,end) == 4),1:2);

figure(1)

subplot(1,2,1)
hold on

scatter(diab_principal_components(:,1),diab_principal_components(:,2),si*10,"k");

xlabel("PC1, "+num2str(explained(1))+"%","FontWeight","bold")
```

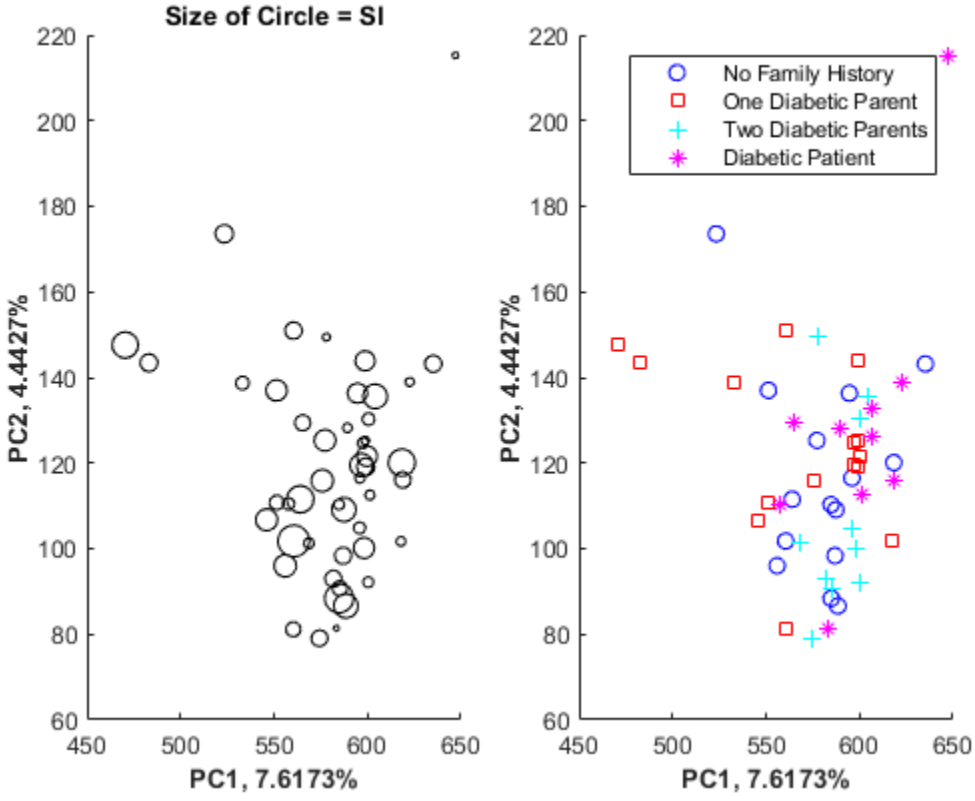
```
ylabel("PC2, "+num2str(explained(2))+"%","FontWeight","bold")
title("Size of Circle = SI")
hold off

subplot(1,2,2)
hold on

plot(fam_hist_negative(:,1),fam_hist_negative(:,2),"bo");
plot(fam_hist_one(:,1),fam_hist_one(:,2),"rs");
plot(fam_hist_two(:,1),fam_hist_two(:,2),"c+");
plot(diabetics(:,1),diabetics(:,2),"m*");

xlabel("PC1, "+num2str(explained(1))+"%","FontWeight","bold")
ylabel("PC2, "+num2str(explained(2))+"%","FontWeight","bold")
legend("No Family History","One Diabetic Parent","Two Diabetic
 Parents","Diabetic Patient");
hold off
```

Warning: Column headers from the file were modified to make them valid MATLAB identifiers before creating variable names for the table. The original column headers are saved in the VariableDescriptions property.
Set 'VariableNamingRule' to 'preserve' to use the original column headers as table variable names.



Published with MATLAB® R2023a