```matlab
%Name: David George
%StudentID: 251004930



T = input_data();

%Part A)
    ans = var_prob(T, 'npreg', 2);
    disp(ans);

%Part B)
    compareFigure(T);

%Part C)

        D = table2array(rmmissing(T(:,[1:6])));
        D = squareform(pdist(D));

        %Classical MDS
        MDS  = cmdscale(D, 2);


        figure

        newTable = readtable("diabetes-pima.csv");

        newTable.npreg = str2double(newTable.npreg);
        newTable.bp = str2double(newTable.bp);
        newTable.glu = str2double(newTable.glu);
        newTable.skin = str2double(newTable.skin);
        newTable.bmi = str2double(newTable.bmi);
        newTable.age = str2double(newTable.age);

        %Changing the NA to be defaulted to 0
            for idx = 1:length(newTable.diabetic)

                if newTable.diabetic{idx} == "NA"
                    newTable.diabetic{idx} = 'No';
                end
            end


        newTable=newTable(~any(ismissing(newTable),2),:);
         gscatter(MDS(:,1), MDS(:,2),newTable.diabetic);
          title("MDS by Diabetic Status");


    %The MDS figure is not homogenous between the diabetic and non
 diabetic
    % indivduals. There is a clear split, the non diabetic are closer
 to
```

```matlab
    % the left, while the diabetic individuals tend to be on the
right.
    %The fact that they are split shows there is hetrogenity and there
are
    %other factors, the other varibles, that somehow impact the
occurance
    %of diabetes amogst indivuals.




%A) Funciton that calcualtes the probaility for any five biological
%varibles
    function prob = var_prob(T, varType, value)
    X = table2array( T(:, {varType}) );

    % Isolate the response variable:
    Y = T.diabetic;


    % Fit the bivariate logistic regression:
    model = glmfit(X, Y,'binomial');


    % The variable `b` stores the regression coefficients

    % the probability given the values
    prob = logistic_fct(value, model);



    end

function y = logistic_fct(x1,b)

    %Function used for logisitc regression
    y = 1 / (1 + exp(-(b(1) + b(2) * x1)));
end

%Funciton for part B)
    function compareFigure(T)


     XnPreg = linspace(0,20,21);
     YxnPreg = [];

     XbP = linspace(50,100,101);
     YbP = [];

     XGlu = linspace(70,200,131);
     YGlu = [];

     XSkin = linspace(10,100,91);
     YSkin = [];
```

```matlab
    Xbmi = linspace(15,50,100);
    Ybmi = [];

    Xage = linspace(20,80,61);
    Yage = [];

 %All the forloops are used to create corresponding y values for
each
 %variable type
    for idx = 1:length(XnPreg)
        YxnPreg(idx) = var_prob(T, 'npreg', idx);
    end

    for idx = 1:length(XbP)
        YbP(idx) = var_prob(T, 'bp', idx);
    end

    for idx = 1:length(XGlu)
        YGlu(idx) = var_prob(T, 'glu', idx);
    end

    for idx = 1:length(XSkin)
        YSkin(idx) = var_prob(T, 'skin', idx);
    end

    for idx = 1:length(Xbmi)
        Ybmi(idx) = var_prob(T, 'bmi', idx);
    end

    for idx = 1:length(Xage)
        Yage(idx) = var_prob(T, 'age', idx);
    end

 %The following section is used to plot the 6 charts

    figure
    subplot(2, 3, 1);
    plot(XnPreg ,YxnPreg);
    title("Probability of Diabetes Vs. Number of Pregnancies");
    ylabel("Probability");
    xlabel("Number of Pregenancies");




    subplot(2, 3, 2);
    plot(XGlu ,YGlu);
    title("Probability of Diabetes Vs. Plasma Glucose Concentration
(mg/dL) ");
    ylabel("Probability");
    xlabel("plasma glucose concentration (mg/dL)");
```

```matlab
    subplot(2, 3, 3);
    plot(XbP ,YbP);
    title("Probability of Diabetes Vs. Blood Pressure");
    ylabel("Probability");
    xlabel("Blood Pressure (mm Hg)");


     subplot(2, 3, 4);
    plot(XSkin ,YSkin);
    title("Probability of Diabetes Vs. Tricep Skin Fold Thickness
 (mm)");
    ylabel("Probability");
    xlabel("Tricep Skin Fold Thickness (mm)");

     subplot(2, 3, 5);
    plot(Xbmi ,Ybmi);
    title("Probability of Diabetes Vs. BMI (kg m^-2)");
    ylabel("Probability");
    xlabel("BMI (kg m^-2)");


     subplot(2, 3, 6);
    plot(Xage ,Yage);
    title("Probability of Diabetes Vs. Age (years) ");
    ylabel("Probability");
    xlabel("Age (years)");

    end


function T = input_data()


%Funciton to read in values and cast them accordingy
T = readtable("diabetes-pima.csv");
T.npreg = str2double(T.npreg);
T.bp = str2double(T.bp);
T.glu = str2double(T.glu);
T.skin = str2double(T.skin);
T.bmi = str2double(T.bmi);
T.age = str2double(T.age);


%Going over the length of T.diabetic and and casting and making it
 boolean
%based
for idx = 1:length(T.diabetic)

    if T.diabetic{idx} == "NA"
        T.diabetic{idx} = '0';
    end

    if T.diabetic{idx} == "Yes"
        T.diabetic{idx} = '1';
```

```
        end

    if T.diabetic{idx} == "No"
        T.diabetic{idx} = '0';
    end

end
  T.diabetic = str2double(T.diabetic);


end

    0.276386489000900
```
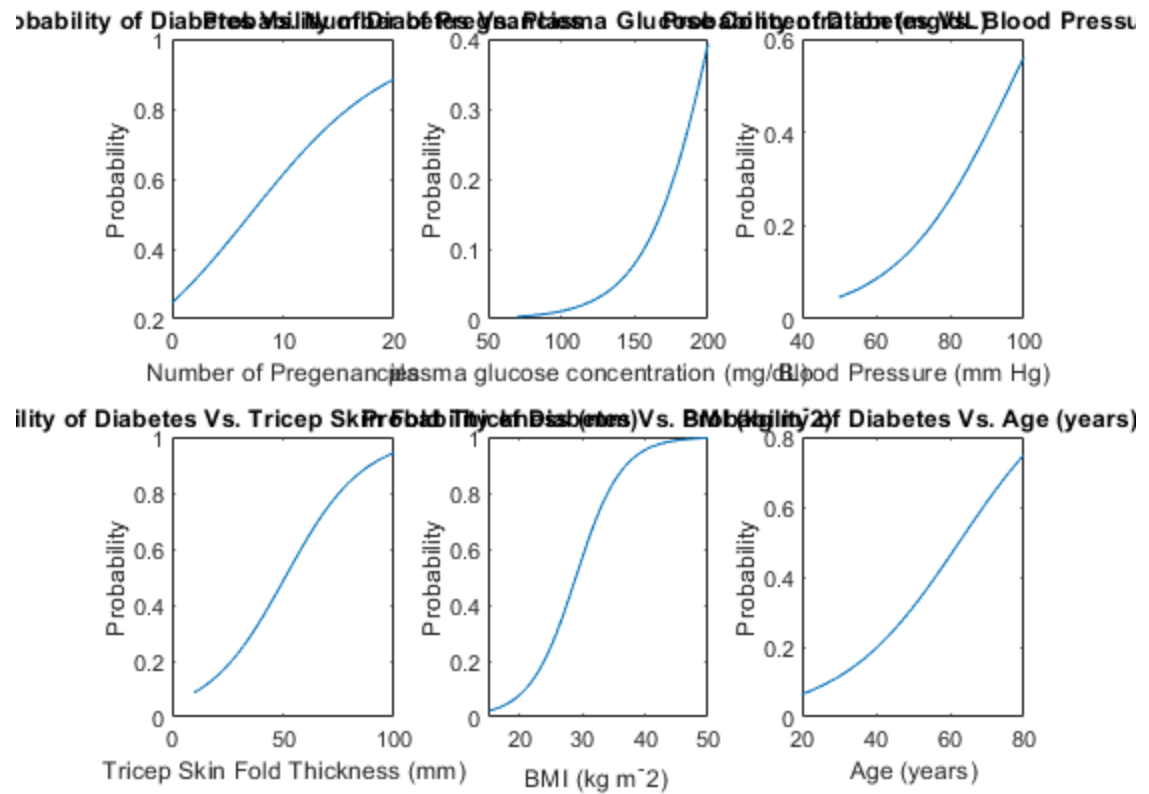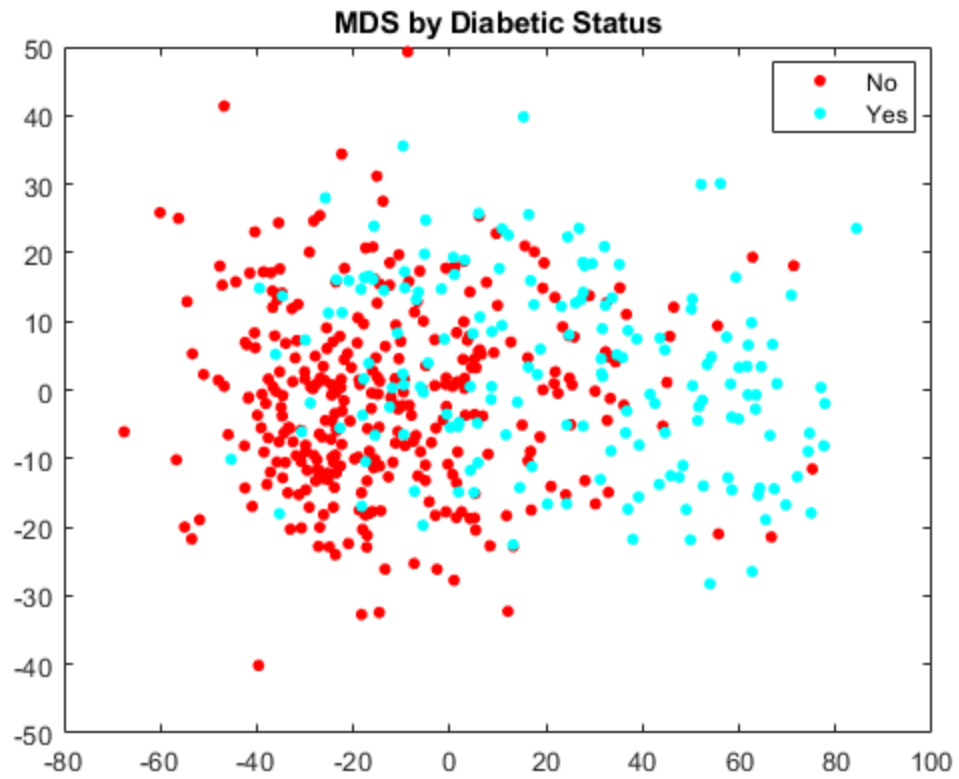
MDS by Diabetic Status

Published with MATLAB® R2019b