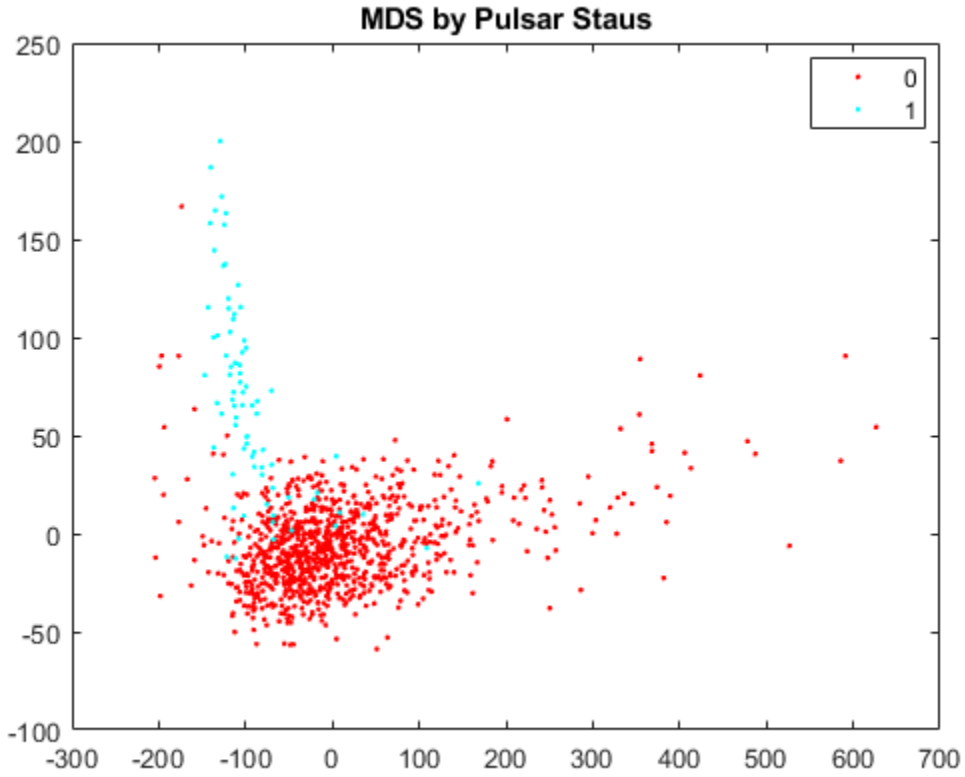
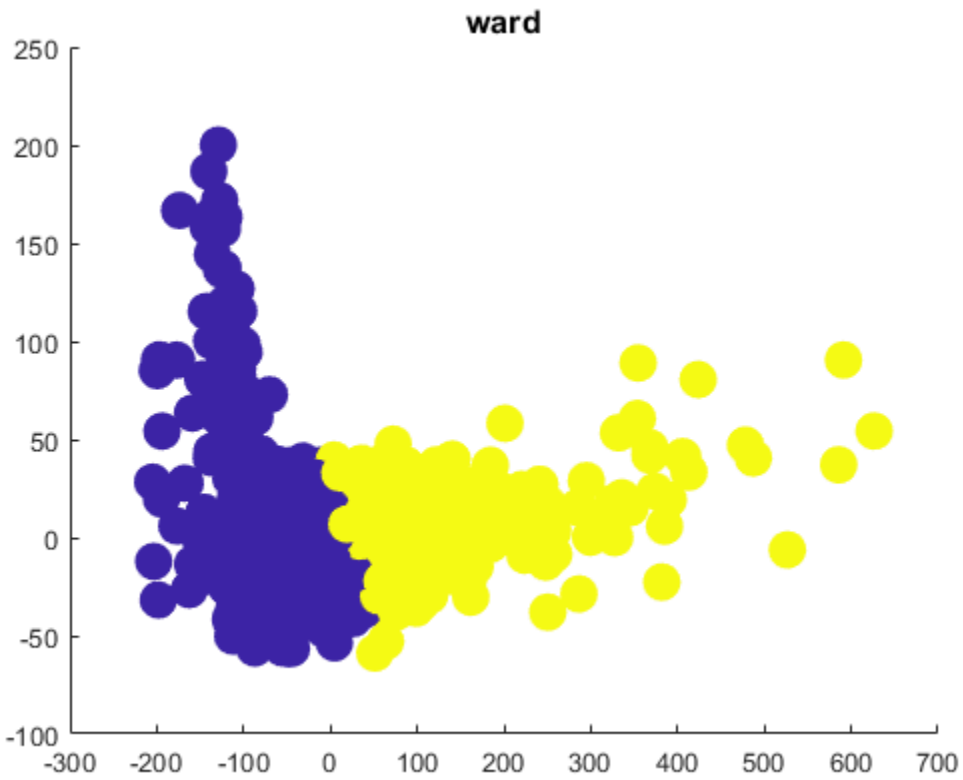
```
%Name: David George
%Student ID: 251004930

%A)
    T = readtable("pulsar.csv");
    newTable = T;
     D = table2array(rmmissing(T(:,[1:8])));
     ArrayD = D;
     D = squareform(pdist(D));

    %Classical MDS
    MDS = mdscale(D,2,'criterion','sammon');

    newTable.Var9 = num2str(newTable.Var9);

    figure
    gscatter(MDS(:,1), MDS(:,2),newTable.Var9);
    title("MDS by Pulsar Staus");
    %The pulsars are denoted by the value of 1 on the legend, 0 means
 its not a
    %pulsar
```




```
%B)
      mod = fitglm(table2array(rmmissing(T(:,[1:8])))   ,
table2array(rmmissing(T(:,9))),'Distribution', 'binomial');
```

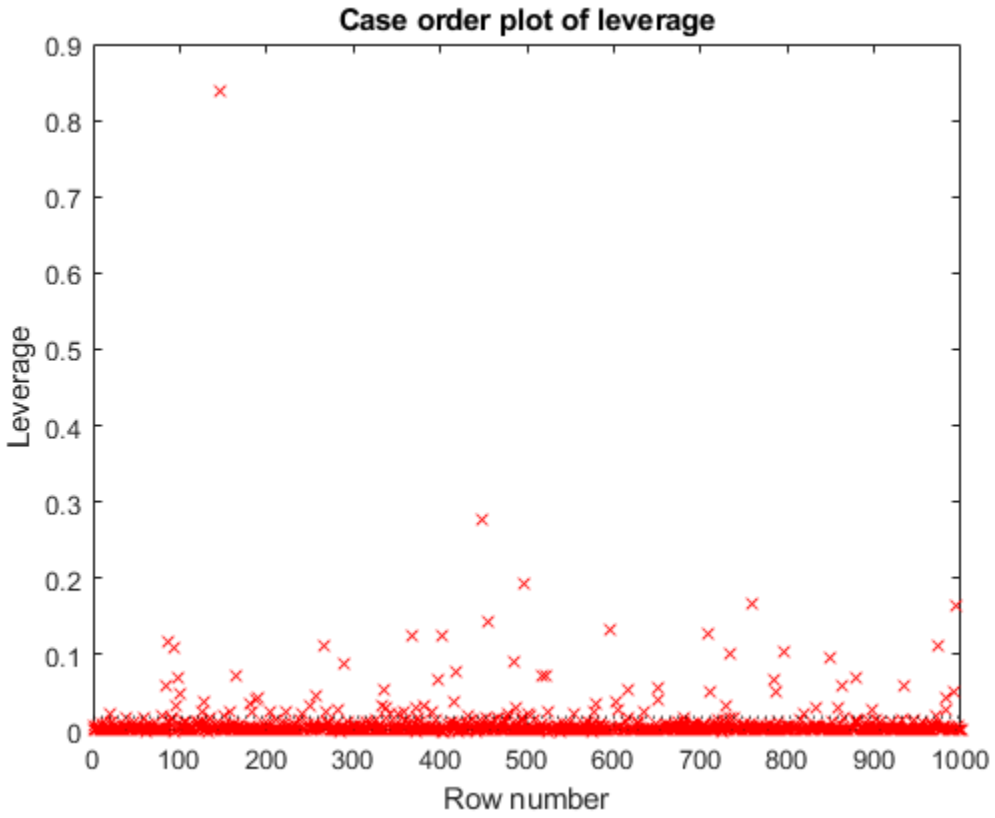
```
      p = mod.Fitted.Probability;
      [fpr,tpr,thresholds,AUC, opt] =
perfcurve(table2array(rmmissing(T(:,9))), p, 1);
     %This is an effective method, this is because the Area under the
Curve
     %(AUC) is 0.9763, which is incredbly close to the maximum value
of 1.0,which
     %means it is an effective method.

%C)

    methods = ["ward"];


   %WARD clustering is the closest and most accurate clustering
   %Comparing this to the mds, it is clear.
   figure


         m = methods(1);
         L = linkage(MDS,m);
         C = cluster(L,'Maxclust',2);
        scatter(MDS(:,1), MDS(:,2),200, C,'Filled');
        title(methods(1),'FontSize',12);
```




```
%D)
   figure
```

```
        T = table2array(T(:, :));

          mod =
fitglm(ArrayD,rmmissing(T(:,9)),'Distribution', 'binomial');
          p = mod.Fitted.Probability;
          coef = mod.plotDiagnostics;


          [fpr,tpr,thresholds,AUC, opt] = perfcurve( T(:,9), p, 1);

          %B)
              %We want a true postive rate of 90%,
              %Here I found the closest tpr with a range of 0.01
plus
              %minus, then I looked manuaally to find the smallest
FPR

              NinetyTPR = tpr(tpr < 0.90 +.01 & tpr >= 0.90 - 0.01);
              indexPos = find(tpr < 0.90 +.01 & tpr >= 0.90 - 0.01);
              NinetyFPR = fpr(indexPos);
              SmallestFpr= min(NinetyFPR);
              CorrespondingIndex = (find(fpr==SmallestFpr));
              ThreshHoldBest = thresholds(143);

              % The closest TRUE psotive rate is 0.09008 and the
smallest, correspondig FPR is
              % 0.3053.
              % I used the corresponding index to find the
threshhold,
              % which I found 0.0428


      mod = glmfit(ArrayD,rmmissing(T(:,9)), 'binomial');
      Table2 = readtable("pulsar2.csv");

      %Now all we have to is calcualte the probablilty, for each
row,
      %that it is a pulsar, then compare it to our thresh hold to
      %determine how it should be classfied. Increment a counter to
count
      %how many pulsars there are.
      count = 0;
      for idx = 1:length(Table2.Var1)

          if logistic_fct( table2array(Table2(idx, :)), mod) >
ThreshHoldBest
              count = count +1;
          end


      end

  %There are 48 pulsars given in this data
  disp("There are this many pulsars in this data");
```

```
disp(count);
```




```
function y = logistic_fct(x,b)

    %X is the value that is being inputed, b is the vlaue of logisitic
    %regression

    %Function used for logisitc regression

     y = 1/(1 +exp(-(b(1) + b(2)*x(1) + b(3)*x(2) + b(4)*x(3) +
 b(5)*x(4) + b(6)*x(5) + b(7)*x(6) + b(8)*x(7) + b(9)*x(8))));
end
```

```
There are this many pulsars in this data
    48
```

Published with MATLAB® R2019b