```
%Name: David
%StudentNumber: 251004930

dat = readtable("diamonds.csv");

% a)

summary(dat)

%They (the numberical values) are not comparable units due to the
 varying
%ranges between the varibles.

% b)
ss =  grpFix(datasample(readtable("diamonds.csv"),2000));


% c)
[coeff,numscore,latent,tsquare,resultant] =
 pca(table2array(ss),'VariableWeights','variance');
% d)

% normalize the cofeffecitns of the vector
coef_norm = inv(diag(std(v))) * coeff ;

pc1 = numscore(:,1);
pc2 = numscore(:,2);
pc3 = numscore(:,3);

number = 7;


[max2 maxtwo] = maxk(pc2,number);
pc1_2max = pc1(maxtwo);

[min2 mintwo] = mink(pc2,number);

[max1 max] = maxk(pc1,number);
pc2_1max = pc2(max);

[min1 minone] = mink(pc1,number);
pc2_1min = pc2(minone);

pc1_2min = pc1(mintwo);

[max3 maxthree] = maxk(pc3,number);
pc1_3max = pc2(maxthree);

[min3 ] = mink(pc3,number);
pc1_3min = pc2(idxmin3);

% e) Plot the PCA on the first 2 principal component
```

```
figure
scatter(pc1, pc2,200, 'MarkerFaceColor','red');
alpha(0.1);
grid();
set(gca,'FontSize',20);

figure
scatter3(pc1,pc2,pc3,200, 'MarkerFaceColor','red');
alpha(0.1);
set(gca,'FontSize',20);

figure();
pareto(resultant);

fprintf("The first three components are responsible for  %0.2f %
 varience of the dataset given \n",sum(resultant(1:3)));

%this is a lot as it covers more than the majority of the varience, in
 just
%three of the prinicpal componens

function data =  grpFix(ss)
ss.cut = grp2idx(ss.cut);
ss.clarity = grp2idx(ss.clarity);
ss.color =  grp2idx(ss.color);
data = ss;
end
```

```
Variables:

    carat: 53940×1 double

        Values:

            Min         0.2
            Median      0.7
            Max        5.01

    cut: 53940×1 cell array of character vectors

    color: 53940×1 cell array of character vectors

    clarity: 53940×1 cell array of character vectors

    depth: 53940×1 double

        Values:

            Min          43
            Median     61.8
            Max          79
    table: 53940×1 double

        Values:

            Min         43
            Median      57
            Max         95

    price: 53940×1 double

        Values:

            Min          326
            Median      2401
            Max        18823

The first three components are responsible for  67.49
```

Published with MATLAB® R2019b