

```

%{
-----
    Daily stock price changes in 2015 for S&P500 stocks
    Principal Component Analysis exercise
    Sam Qian
-----

I obtained daily percentage price changes over 236 trading days in 2015 for 497 stocks
in the S&P500 index. I created an 497 x 236 matrix and performed PCA to ask
if stocks cluster in their daily behavior over the year.

Stock data obtained with 'Multiple Stock Quote Downloader'
(http://investexcel.net/)

Other sources consulted:
http://matlabdatamining.blogspot.com/2010/02/principal-components-analysis.html

%}

clear all;
clc;

% Load data matrix, stock symbols, and sectors
load dailyChange.mat
load symList.mat
load sectors.mat
input = dailyChange';

% Descriptive stats
[n, m] = size(input);
stockMeans = mean(input);
stockStd = std(input);

% Standardize input data
input_standardized = zscore(input);

% PCA
[coeff, score, latent] = pca(input_standardized);
% coeff is the coefficients for the principal component.
% score contains the actual principal components (i.e. it is the coefficient
% times the standardized data.
% latent are the variances of the respective principal components

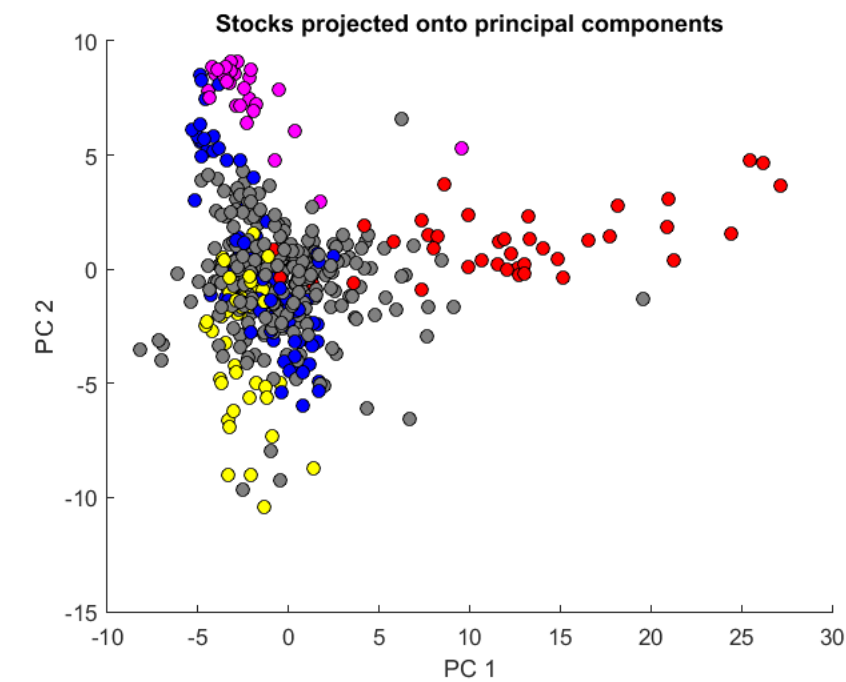
% to go back to the original data, multiply each observation by the sample
% standard deviation vector and add the mean vector
% ((input_standardized * coeff) * coeff') .* repmat(stockStd,[n 1]) + repmat(stockMeans,[n 1]);

% Plot stocks on PC1 and PC2, color-coded based on sector
figure;
hold on;
for i=1:size(input,1);
    if strcmp(symList{2,i},sectors{1})==1;
        plot(score(i,1),score(i,2),'o','MarkerFaceColor','[0.5 0.5 0.5]', 'MarkerEdgeColor','k');
    end
    if strcmp(symList{2,i},sectors{2})==1;
        plot(score(i,1),score(i,2),'o','MarkerFaceColor','[0.5 0.5 0.5]', 'MarkerEdgeColor','k');
    end
    if strcmp(symList{2,i},sectors{3})==1;
        plot(score(i,1),score(i,2),'o','MarkerFaceColor','r', 'MarkerEdgeColor','k');
    end
    if strcmp(symList{2,i},sectors{4})==1;
        plot(score(i,1),score(i,2),'o','MarkerFaceColor','b', 'MarkerEdgeColor','k');
    end
    if strcmp(symList{2,i},sectors{5})==1;
        plot(score(i,1),score(i,2),'o','MarkerFaceColor','y', 'MarkerEdgeColor','k');
    end
    if strcmp(symList{2,i},sectors{6})==1;
        plot(score(i,1),score(i,2),'o','MarkerFaceColor','[0.5 0.5 0.5]', 'MarkerEdgeColor','k');
    end
    if strcmp(symList{2,i},sectors{7})==1;
        plot(score(i,1),score(i,2),'o','MarkerFaceColor','[0.5 0.5 0.5]', 'MarkerEdgeColor','k');
    end
    if strcmp(symList{2,i},sectors{8})==1;
        plot(score(i,1),score(i,2),'o','MarkerFaceColor','[0.5 0.5 0.5]', 'MarkerEdgeColor','k');
    end
    if strcmp(symList{2,i},sectors{9})==1;
        plot(score(i,1),score(i,2),'o','MarkerFaceColor','[0.5 0.5 0.5]', 'MarkerEdgeColor','k');
    end
    if strcmp(symList{2,i},sectors{10})==1;
        plot(score(i,1),score(i,2),'o','MarkerFaceColor','m', 'MarkerEdgeColor','k');
    end
end

title('Stocks projected onto principal components');
xlabel('PC 1');
ylabel('PC 2');
disp('Sectors: Utilities (Magenta), Financials (Blue), Health Care (Yellow), Energy (Red), and all others (Gray)');

```

Sectors: Utilities (Magenta), Financials (Blue), Health Care (Yellow), Energy (Red), and all others (Gray)



```
% Interpretation
disp('Interpretations:');
disp('Stocks within the same sector behave similarly.');
```

Interpretations:
Stocks within the same sector behave similarly.
PCA identified a subset of stocks that separate out from the bulk of data points along PC 1. This subset mostly include stocks in the Energy sector (colored red), w