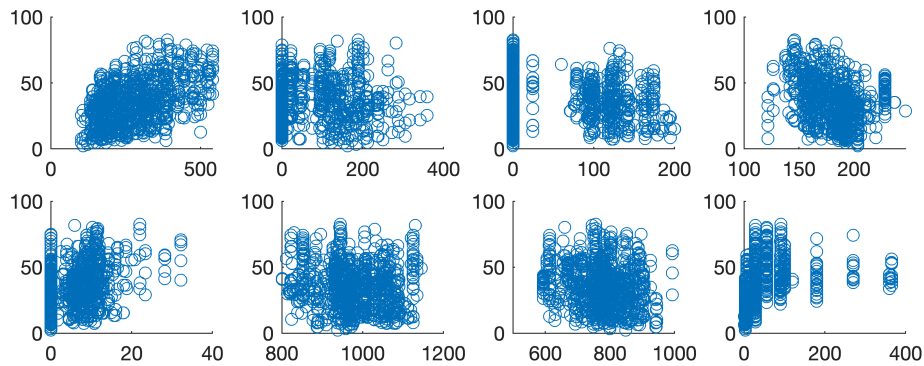# Data Analysis

Load Data

```
traindata = importdata('traindata.txt');
X = traindata(:, 1:8);
Y = traindata(:, 9);
```

Features Correlation plot

```
for k = 1:8
    data = rand(1,10);
    subplot(4, 4, k)
    scatter(X(:, k), Y);
end
```
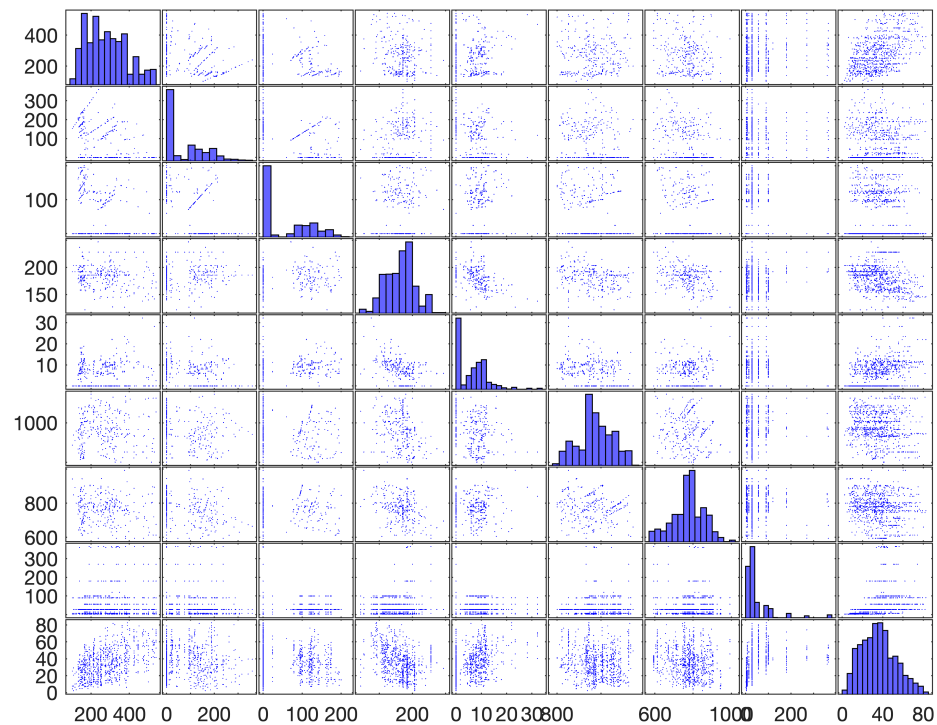

Features Correlation values

```
corre = corrcoef(traindata);
disp(corre(9, :));
```
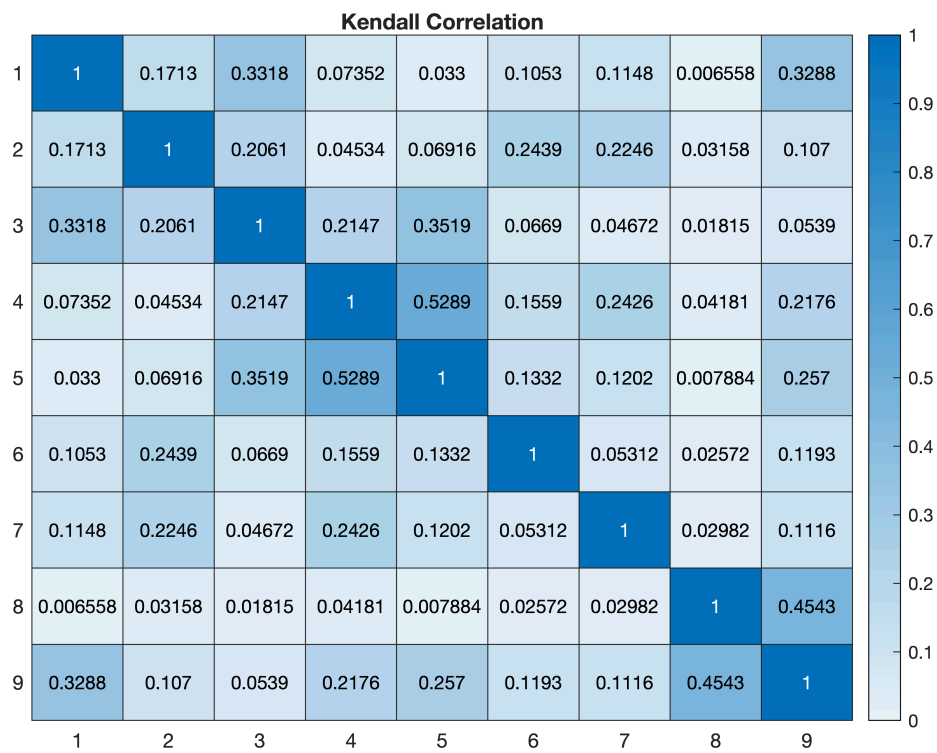
```
    0.4980    0.1148   -0.0983   -0.3015    0.3776   -0.1598   -0.1516    0.3364    1.0000
```

```
gplotmatrix(traindata);
```
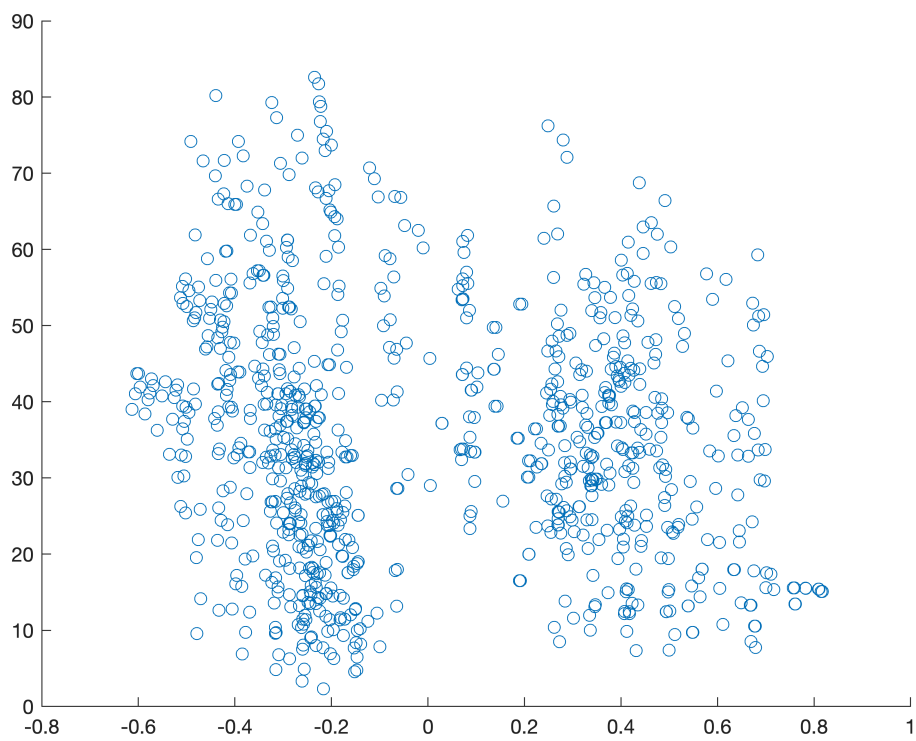
Features Correlation values plot

```matlab
type = "Kendall";
C = corr(traindata,'type',type);
heatmap(abs(C), 'ColorLimits',[0 1],'Title', type + " Correlation");
```

Kendall Correlation

	1	2	3	4	5	6	7	8	9
1	1	0.1713	0.3318	0.07352	0.033	0.1053	0.1148	0.006558	0.3288
2	0.1713	1	0.2061	0.04534	0.06916	0.2439	0.2246	0.03158	0.107
3	0.3318	0.2061	1	0.2147	0.3519	0.0669	0.04672	0.01815	0.0539
4	0.07352	0.04534	0.2147	1	0.5289	0.1559	0.2426	0.04181	0.2176
5	0.033	0.06916	0.3519	0.5289	1	0.1332	0.1202	0.007884	0.257
6	0.1053	0.2439	0.0669	0.1559	0.1332	1	0.05312	0.02572	0.1193
7	0.1148	0.2246	0.04672	0.2426	0.1202	0.05312	1	0.02982	0.1116
8	0.006558	0.03158	0.01815	0.04181	0.007884	0.02572	0.02982	1	0.4543
9	0.3288	0.107	0.0539	0.2176	0.257	0.1193	0.1116	0.4543	1

PCA for visualization

```
mu = mean(X);
r = range(X);
X = (X - mu) ./ r;
[P, S, V] = pca(X);

scatter(S(:, 1), Y);
```

```
scatter3(S(:,1), S(:,2), Y);
```