

% Exploratory Data Analysis for MIT-BIH Supraventricular Arrhythmia Dataset

```
T = readtable ("MIT-BIH Supraventricular Arrhythmia Database.csv");
```

Warning: Column headers from the file were modified to make them valid MATLAB identifiers before creating variable names for the table. The original column headers are saved in the VariableDescriptions property. Set 'VariableNamingRule' to 'preserve' to use the original column headers as table variable names.

```
T = rmmissing(T);
if any(strcmp('0_rPeak', T.Properties.VariableNames))
    T = T(T.('0_rPeak') > -3000 & T.('0_rPeak') < 3000, :);
end

%(z-score normalization)
numericVars = varfun(@isnumeric, T, 'OutputFormat', 'uniform');
T_norm = T;
T_norm(:, numericVars) = varfun(@(x) (x - mean(x)) / std(x), T(:, numericVars));

disp('First 5 Rows:');
```

First 5 Rows:

```
disp(head(T, 5));
```

record	type	x0_pre_RR	x0_post_RR	x0_pPeak	x0_tPeak	x0_rPeak	x0_sPeak	x0_qPeak
800	{'N'}	168	167	-0.00065862	-0.076691	1.08	-0.95884	-0.0098488
800	{'N'}	167	169	-0.0072373	-0.080968	1.1175	-0.99395	-0.037873
800	{'N'}	169	170	-0.015322	-0.065713	1.0912	-0.9848	-0.047307
800	{'N'}	170	166	-0.011573	-0.067132	1.0755	-0.91229	-0.018548
800	{'N'}	166	169	-0.0042693	-0.081436	1.1633	-0.99168	-0.013624

```
%% Unique Types and Records
disp('Unique Beat Types:');
```

Unique Beat Types:

```
disp(unique(T.type));
```

```
{'F' }
{'N' }
{'Q' }
{'SVEB'}
{'VEB' }
```

```
disp('Unique Records:');
```

Unique Records:

```
disp(unique(T.record));
```

```
800
801
802
803
```

804
805
806
807
808
809
810
811
812
820
821
822
823
824
825
826
827
828
829
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884

885
886
887
888
889
890
891
892
893
894

```
% Summary Statistics for Numeric Features
```

```
numericData = T(:, varfun(@isnumeric, T, 'OutputFormat', 'uniform'));  
disp('Mean Values:');
```

Mean Values:

```
disp(varfun(@mean, numericData));
```

<u>mean_record</u>	<u>mean_x0_pre_RR</u>	<u>mean_x0_post_RR</u>	<u>mean_x0_pPeak</u>	<u>mean_x0_tPeak</u>	<u>mean_x0_rPeak</u>	<u>mean_x0_sPeak</u>
851.93	96.533	96.514	0.078229	0.4692	0.5332	-0.801

```
disp('Standard Deviation:');
```

Standard Deviation:

```
disp(varfun(@std, numericData));
```

<u>std_record</u>	<u>std_x0_pre_RR</u>	<u>std_x0_post_RR</u>	<u>std_x0_pPeak</u>	<u>std_x0_tPeak</u>	<u>std_x0_rPeak</u>	<u>std_x0_sPeak</u>
27.523	27.184	27.239	0.25367	0.71165	1.0154	0.67936

```
% Histograms of Selected Features
```

```
selectedFeatures = {'x0_rPeak', 'x0_pPeak', 'x0_tPeak'};
```

```
figure;
```

```
for i = 1:length(selectedFeatures)
```

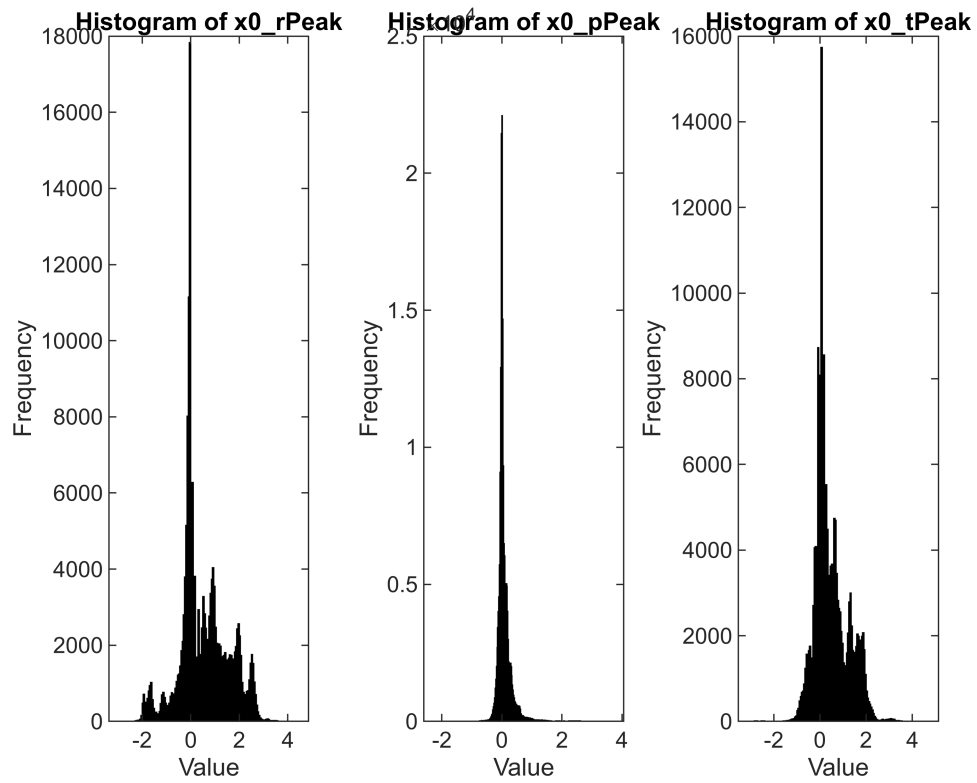
```
    subplot(1, length(selectedFeatures), i);
```

```
    histogram(T.(selectedFeatures{i}));
```

```
    title(['Histogram of ', strrep(selectedFeatures{i}, '_', '\_')]);
```

```
    xlabel('Value'); ylabel('Frequency');
```

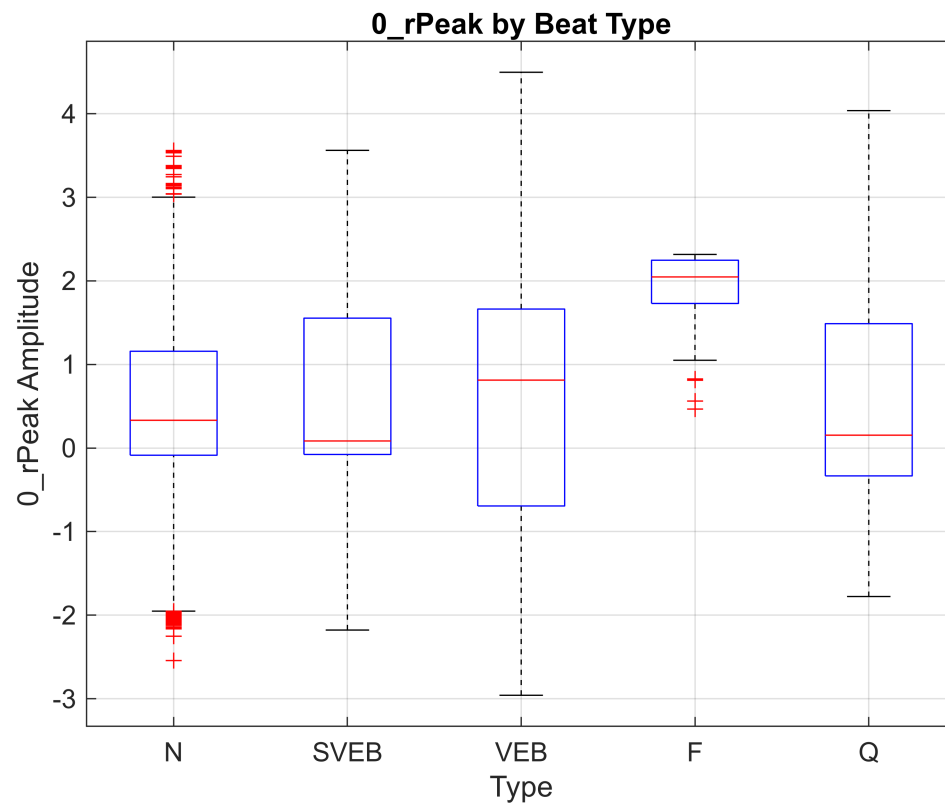
```
end
```



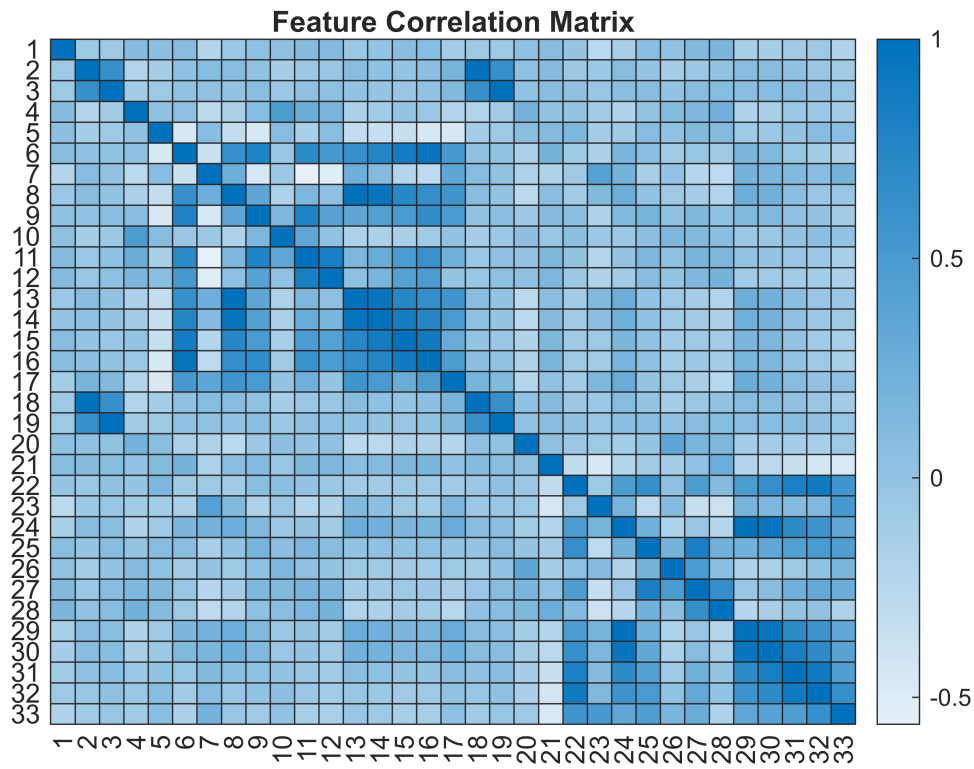
```

%% Boxplot of  $\theta_r$ Peak by Beat Type
figure;
boxplot(T.('x0_rPeak'), T.type);
title('\theta\_rPeak by Beat Type');
xlabel('Type'); ylabel('\theta\_rPeak Amplitude');
grid on;

```



```
%% Correlation Matrix
numericMatrix = table2array(numericData);
corrMatrix = corr(numericMatrix, 'Rows', 'complete');
figure;
heatmap(corrMatrix);
title('Feature Correlation Matrix');
```



```
% Beat Type Distribution
typeCounts = groupcounts(T.type);
figure;
bar(typeCounts);
xticks(1:length(typeCounts));
xticklabels(unique(T.type));
title('Class Distribution (Beat Types)');
xlabel('Beat Type'); ylabel('Count');
grid on;
```

