University of Dayton

Lab 11

CPS 563 – Data Visualization

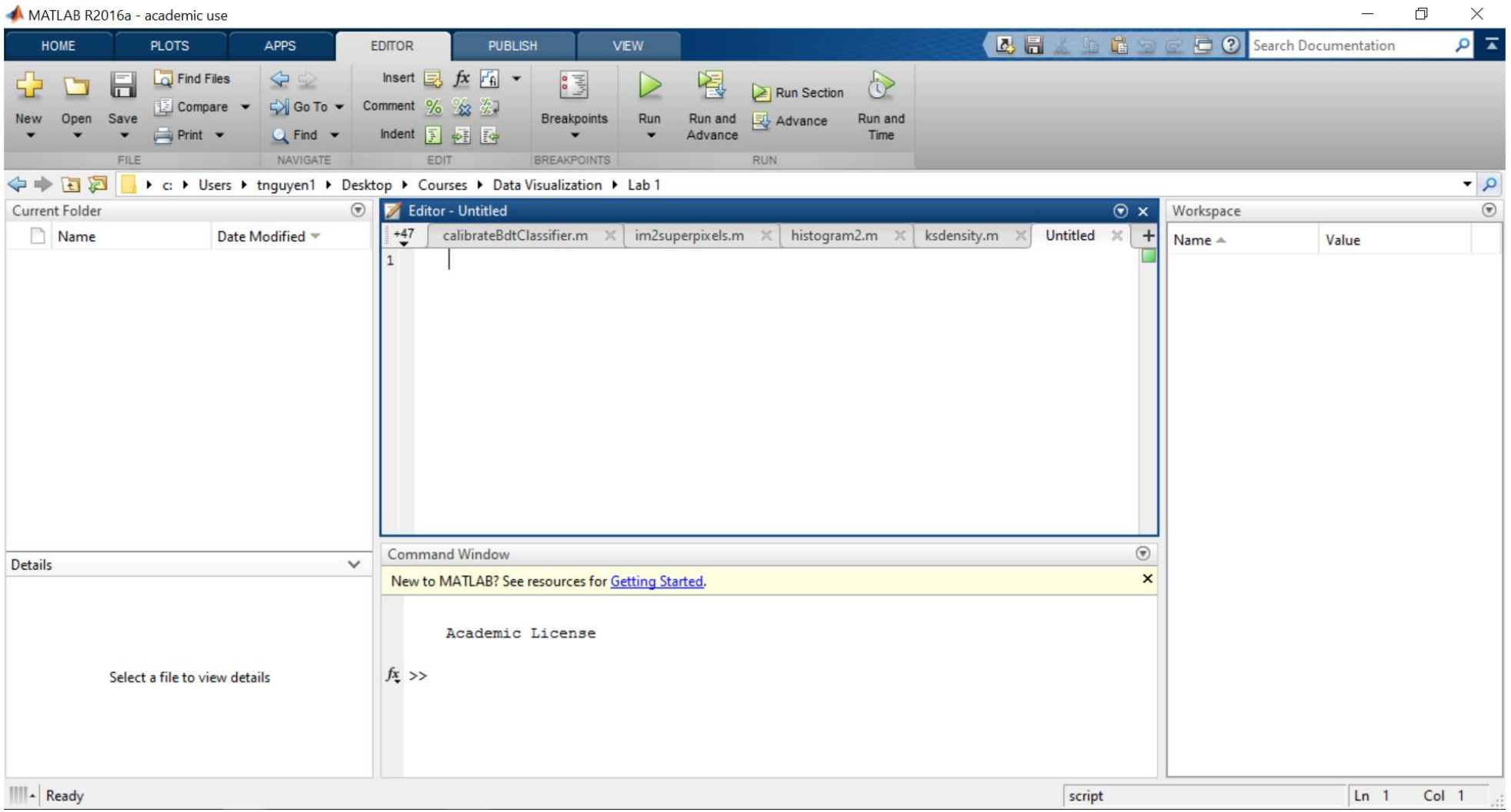
Dr. Tam Nguyen

tamnguyen@udayton.edu

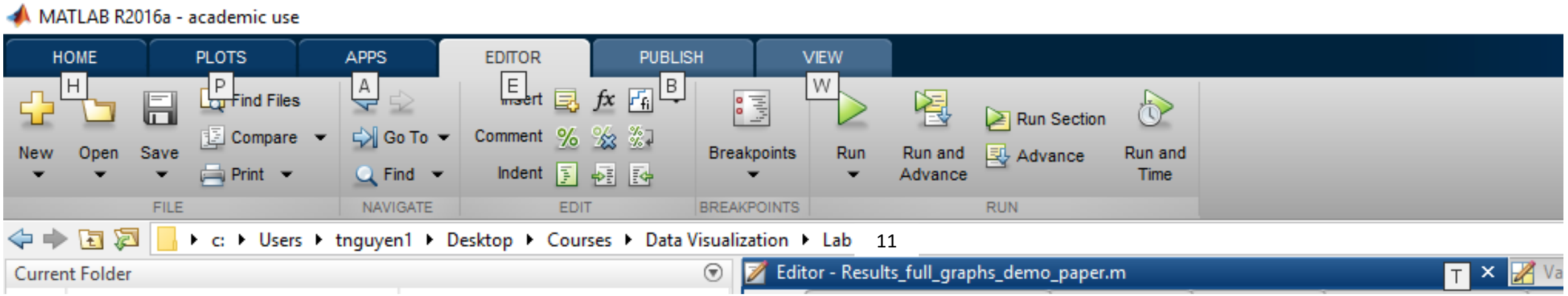
Outline

- Practice with feature selection

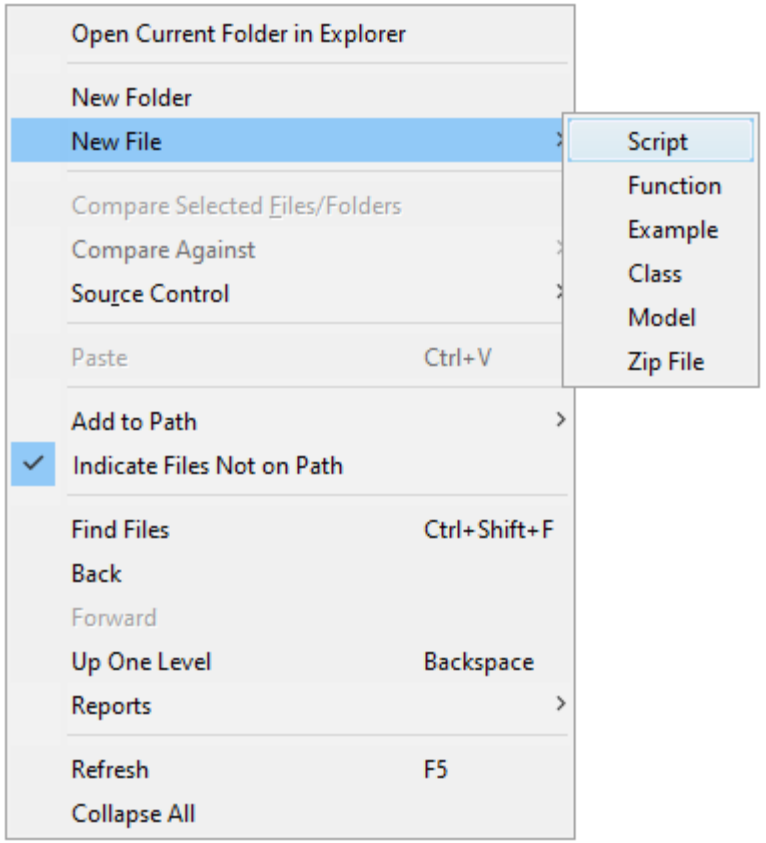
Start MATLAB



Create Lab 11 folder



Create new script file: Lab11.m



Lab11.m

```
close all;
clear all;
clc;
```

Reading the data File

```
%Reading the data File:
fileName = 'flag.data';  %Name of the file
[dat,limit]=importdata(fileName);  %Reading the data
```

Flag Dataset

- https://archive.ics.uci.edu/ml/datasets/Flags

Data Set Characteristics:	Multivariate	**Number of Instances:**	194
Attribute Characteristics:	Categorical, Integer	**Number of Attributes:**	30
Associated Tasks:	Classification	**Missing Values?**	No

Flags

Splitting the data

```
%Spliting the data
data = cell(size(dat,1),1);  %initialize the cell array
for ii = 1:size(dat,1)
    data = dat{ii,1};
    seperate = strsplit(data,limit);
    for k = 1:size(seperate,2)
        dat{ii,k} = seperate{1,k};
    end
end
```

Probability Calculations

```
%start the probability calculations (histograms)
prob = cell(1,size(dat,2));

ENTROPY = zeros(1,size(dat,2));
feature = ENTROPY;
```

Probability Calculations (cellhist function can be downloaded from isidore)

```
for ii = 1:size(dat,2)
    d = dat(:,ii);
    [n,o] = cellhist(d);
    point = n / sum(n);
```

Storing Probability Values

```
%saving probability values:
    for k = 1:size(n,1)
        o{k,2} = n(k,1);
        o{k,3} = point(k,1);
        prob{1,ii} = o;
    end
```

Entropy Calculations

```
    %Entropy calculations:
    E1 = -sum(point.*log2(point));
    ENTROPY(1,ii) = E1;
end %(the end of the for loop in Slide #12)
```

Joint Entropy Calculations

```
%start Joint Entropy Calculations:
JENTROPY = zeros(size(dat,2));
for ii = 1:size(dat,2)  %The first feature
    for k = 1:size(dat,2)  %the second feature
        sum = 0;  %reinitialize sum of joint entropy to zero
        for m = 1:size(prob{1,ii},1)
            for n = 1:size(prob{1,k},1)
                probofX = cell2mat(prob{1,ii}(m,3));
```

Joint Entropy Calculations (cont.)

```
            probofY = cell2mat(prob{1,k}(n,3));
            probofXY = probofX * probofY;
            sum = sum + probofXY * log2(probofXY);
        end
    end
    %assign the summation value to the joint entropy between these two
    %features currently under study:
    JENTROPY(ii,k) = -sum;
  end
end
```

Reshaping:

```
%reshaping the joint entropy features into an array:
JntENTROPYVECT = reshape(JENTROPY,[size(dat,2)*size(dat,2),1]);
[order,Index] = sort(JntENTROPYVECT,'descend');
```

Selecting the features

```
sel = zeros(1,1);%size(data_1,2));
countr = 1;  %counter
indCates_cnt = 1; %initialize counter for indicies and order
threshold = 10;  %thresholdvalue of entropy
```

Selecting Feature Values

```
while (1)
    ord = order(indCates_cnt,1);
    if ord <= threshold
        break;
    end
    ind = Index(indCates_cnt,1);
    [Ind,Jack] = ind2sub([size(dat,2),size(dat,2)],ind);
    Ind_chk = find(sel == Ind,1);
    if isempty(Ind_chk) == 1
        sel(1,countr) = Ind;
        countr = countr + 1;
    end
```

Selecting Feature Values

```
J_chk = find(sel == Jack,1);
   if isempty(J_chk) == 1
      sel(1,countr) = Jack;
      countr = countr + 1;
   end
   indCates_cnt = indCates_cnt + 1;  %increment counter
end
```

Display the Values:

```
%displaying the results of the feature selection algorithm:
disp('The list of useful features are: ');
disp(sel);
```

Displaying the Values:

```
list = 1:30;
red = list(1,~ismember(list,sel));  %the list of remaining features
disp('   ');
disp('While the list of redundant features are: ');
disp(red);
```

Output

Command Window

```
The list of useful features are:
     1     4     5     6     7     2    18    29


While the list of redundant features are:
     3     8     9    10    11    12    13    14    15    16    17    19    20    21    22    23    24    25    26    27    28    30

fx >>
```

Q&A