

From Raw Data to Shareable Insights

- Via a ~30 minute **Introduction to MATLAB** session

Session Overview:

1) Access and Review the Data	1
2) Import the Data.....	1
3) Visualize and Explore the Data.....	1
4) Determine Amount of Data to be Used for Analysis	4
5) Implement Analysis of Data	7
Working with Data from Multiple Files.....	7
6) View a Report of the Analysis	10
7) Next Steps - Available Resources.....	10

1) Access and Review the Data

The storm event data files are freely available from National Oceanic and Atmospheric Administration (NOAA):

<ftp://ftp.ncdc.noaa.gov/pub/data/swdi/stormevents/csvfiles/>

For example: StormEvents_details-ftp_v1.0_d2020_c20200516.csv

2) Import the Data

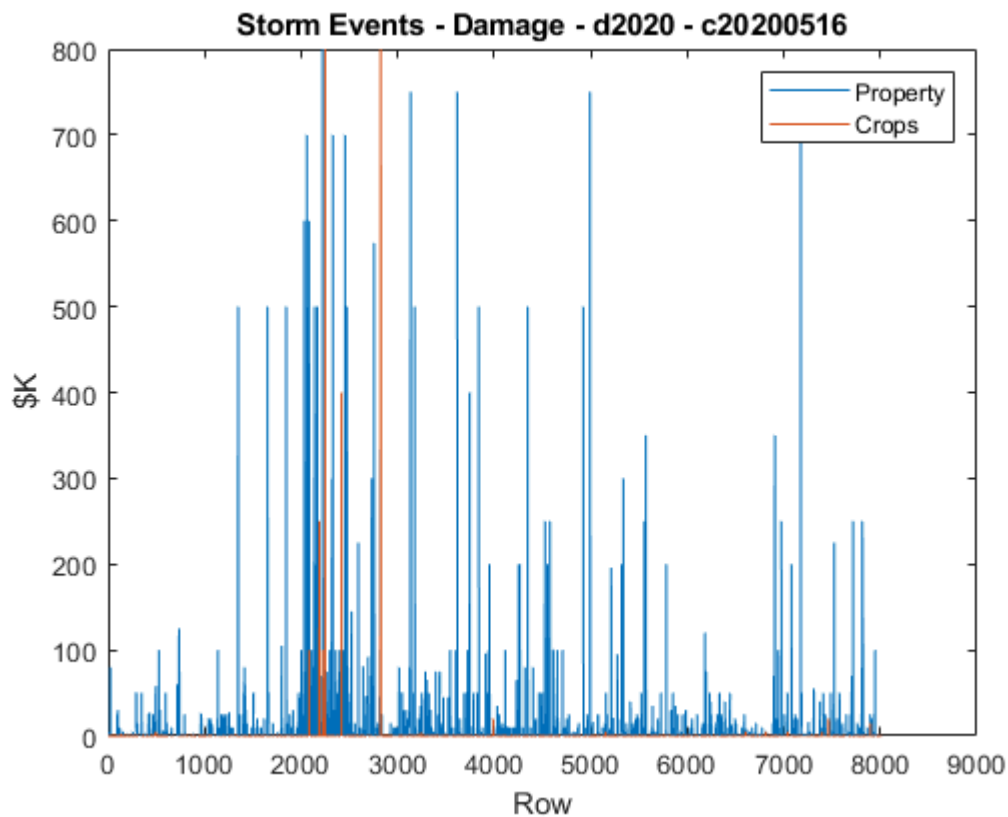
https://www.mathworks.com/help/matlab/import_export/import-data-interactively.html

```
% uiimport('StormEvents_details-ftp_v1.0_d2020_c20200516.csv')
StormEventsdetailsftpv1 = importfile("StormEvents_details-ftp_v1.0_d2020_c20200516.csv");
% load("initialDataSet.mat")
```

3) Visualize and Explore the Data

What Weather Related Financial Costs are Reported?

```
plot(StormEventsdetailsftpv1.DAMAGE_PROPERTY, 'DisplayName', 'Property');hold on;plot(StormEvents
legend('show');
ylabel('$K');
xlabel('Row');
title('Storm Events - Damage - d2020 - c20200516');
```



Observations:

- More records reporting significant property damage than crop damage
- Fewer records have property damages > \$500K

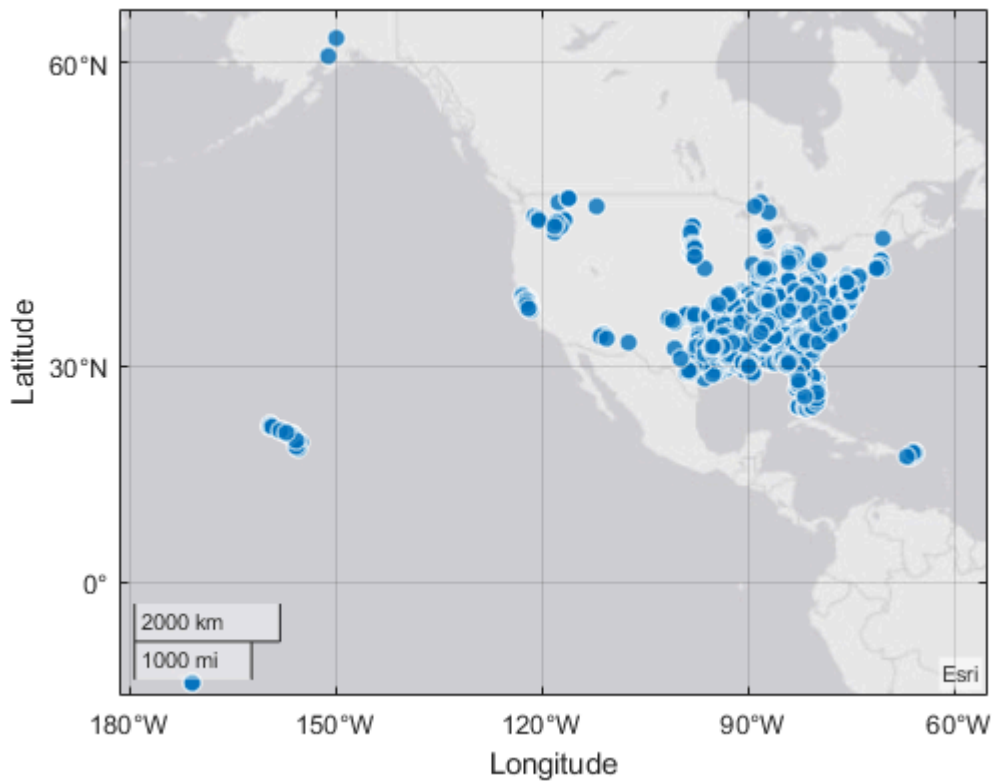
What are the Most Frequent Weather Event Types?

```
wordcloud(StormEventsdetailsftp1.EVENT_TYPE);
```



What Locations Reported Weather Events?

```
geobubble(StormEventsdetailsftpv1.BEGIN_LAT,StormEventsdetailsftpv1.BEGIN_LON);
```



4) Determine Amount of Data to be Used for Analysis

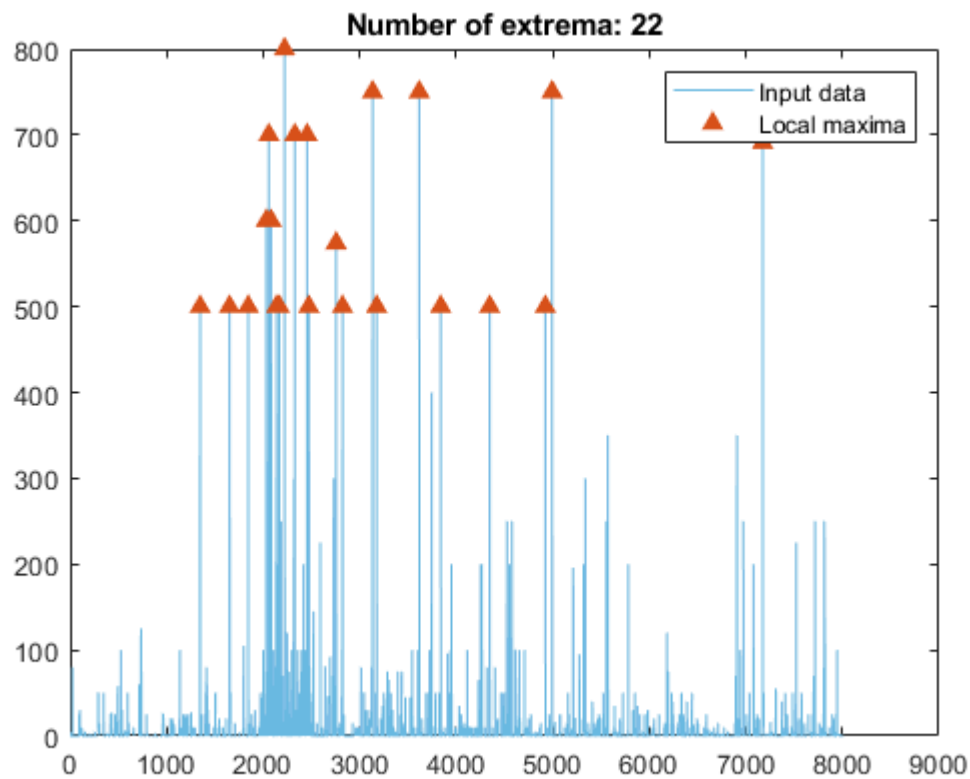
Goal: Identify Locations Reporting the Higher Property Damages

<https://www.mathworks.com/help/matlab/ref/findlocalextrema.html>

```
% Find local maxima
maxIndices = islocalmax(StormEventsdetailsftpv1.DAMAGE_PROPERTY,...
    'MinProminence',500);

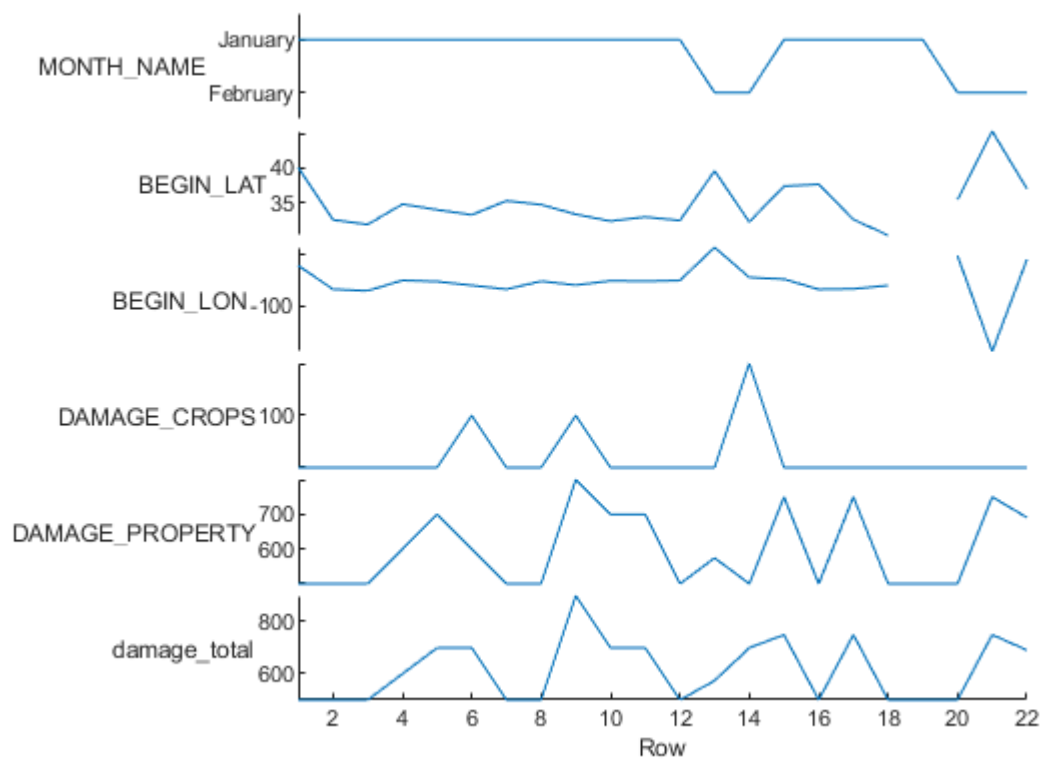
% Display results
clf
plot(StormEventsdetailsftpv1.DAMAGE_PROPERTY,'Color',[109 185 226]/255,...
    'DisplayName','Input data')
hold on

% Plot local maxima
plot(find(maxIndices),StormEventsdetailsftpv1.DAMAGE_PROPERTY(maxIndices),'^',...
    'Color',[217 83 25]/255,'MarkerFaceColor',[217 83 25]/255,...
    'DisplayName','Local maxima')
title(['Number of extrema: ' num2str(nnz(maxIndices))])
hold off
legend
```

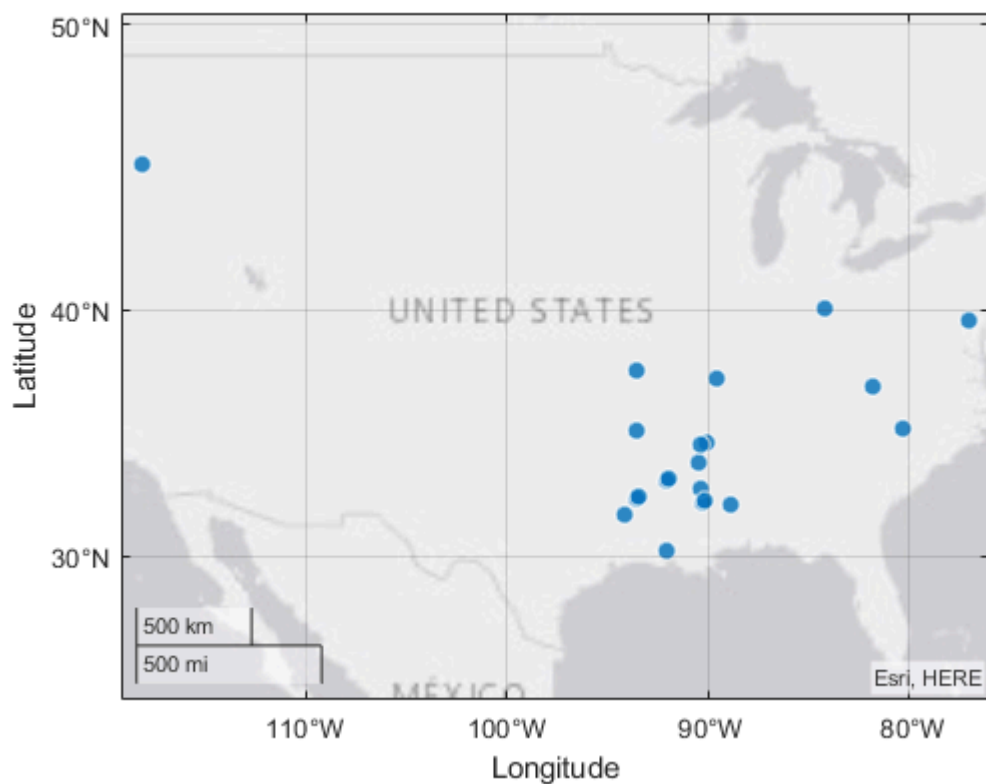


Work with a Subset of New Data to Report Maximum Damage Areas

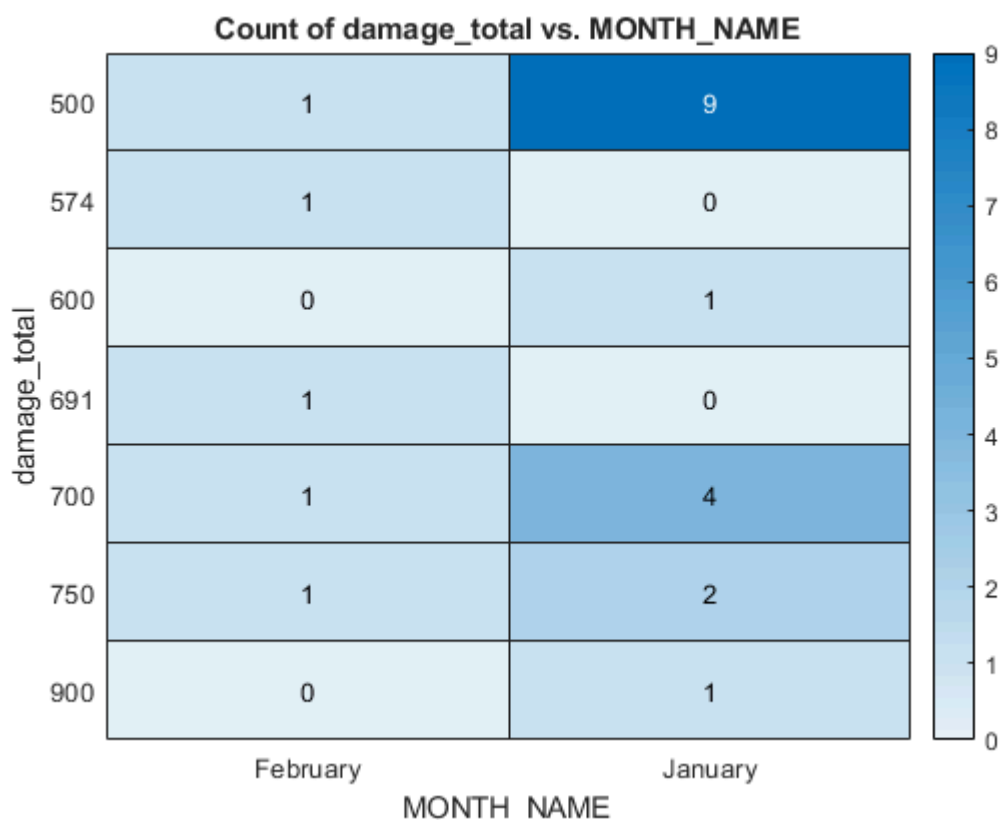
```
costlyStormEvents = StormEventsdetailsftpv1(maxIndices,["MONTH_NAME","BEGIN_LAT","BEGIN_LON","D
costlyStormEvents.damage_total = costlyStormEvents.DAMAGE_CROPS + costlyStormEvents.DAMAGE_PROPR
stackedplot(costlyStormEvents);
```



```
geobubble(costlyStormEvents.BEGIN_LAT,costlyStormEvents.BEGIN_LON);
```



```
heatmap(costlyStormEvents, "MONTH_NAME", "damage_total");
```



5) Implement Analysis of Data

Working with Data from Multiple Files

```
clear
cd NextSteps\
dsAllStormEvents = datastore("data1999_2020");
processedStormEvents = preprocessLargerDataset(dsAllStormEvents);
```

Starting parallel pool (parpool) using the 'local' profile ...
Connected to the parallel pool (number of workers: 4).

```
summary(processedStormEvents)
```

Variables:

Time: 1236327×1 datetime

Values:

Min 01-Jan-99 00:00:00
Median 30-Apr-10 13:25:00
Max 29-Feb-20 19:00:00

event_type: 1236327×1 categorical

Values:

Astronomical Low Tide 544
Avalanche 639
Blizzard 11569
Coastal Flood 2722
Cold/Wind Chill 9656
Debris Flow 1128
Dense Fog 12598

Dense Smoke	84
Drought	52826
Dust Devil	204
Dust Storm	1107
Excessive Heat	8547
Extreme Cold/Wind Chill	12415
Flash Flood	75526
Flood	49731
Freezing Fog	392
Frost/Freeze	11863
Funnel Cloud	7508
Hail	2.5496e+05
Heat	18593
Heavy Rain	23107
Heavy Snow	53448
Heavy Wind	4
High Snow	1
High Surf	8721
High Wind	62524
Hurricane	178
Hurricane (Typhoon)	1347
Ice Storm	8653
Lake-Effect Snow	2272
Lakeshore Flood	249
Landslide	372
Lightning	14025
Marine Dense Fog	8
Marine Hail	709
Marine High Wind	506
Marine Hurricane/Typhoon	50
Marine Lightning	1
Marine Strong Wind	141
Marine Thunderstorm Wind	28132
Marine Tropical Depression	11
Marine Tropical Storm	170
Northern Lights	8
OTHER	1
Rip Current	1255
Seiche	57
Sleet	658
Sneakerwave	18
Storm Surge/Tide	1225
Strong Wind	20159
Thunderstorm Wind	3.0296e+05
Tornado	29492
Tropical Depression	409
Tropical Storm	4873
Tsunami	33
Volcanic Ash	68
Volcanic Ashfall	70
Waterspout	4543
Wildfire	6933
Winter Storm	66591
Winter Weather	59734

damage_property: 1236327x1 double

Values:

Min	0
Median	0
Max	9.1e+08
NumMissing	423466

damage_crops: 1236327x1 double

Values:

Min	0
Median	0
Max	7e+08


```

        NumMissing      507478
damage_total: 1236327x1 double
Values:
    Min      0
    Median    0
    Max      9.4e+08
    NumMissing 532209

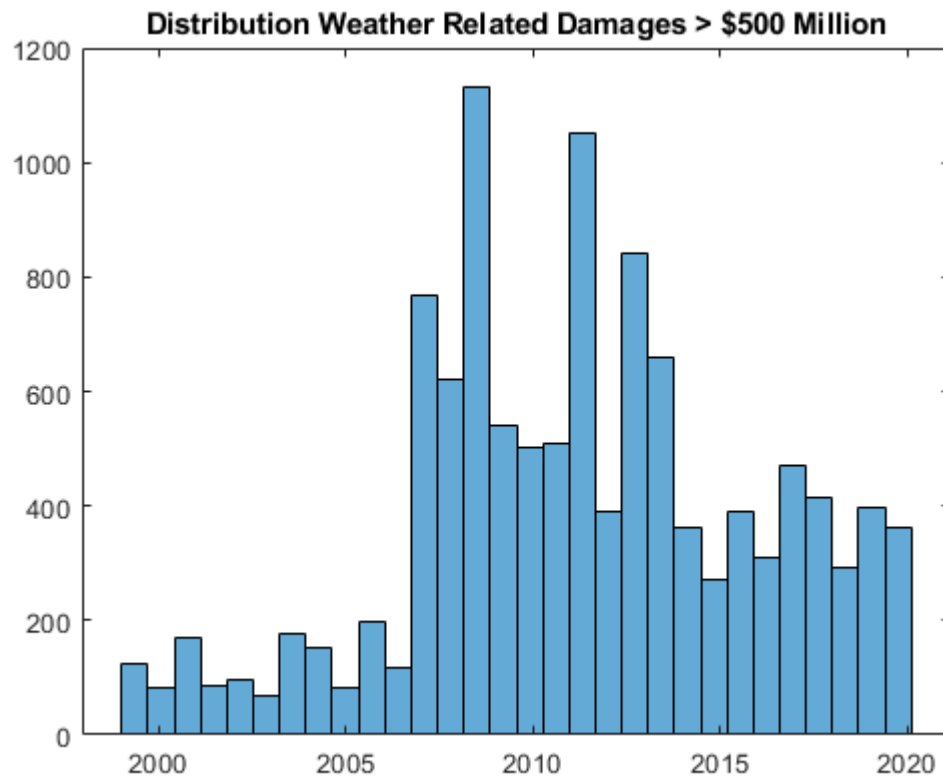
```

```

maxMindices = (processedStormEvents.damage_total > 500000);
selectedData = processedStormEvents(maxMindices,:);
histogram(selectedData.Time);

title('Distribution Weather Related Damages > $500 Million')

```



Could Machine Learning algorithms be used to develop a predictive model from this example dataset?

<https://www.mathworks.com/help/stats/classificationlearner-app.html>

```
load("stormEvents2train.mat");
```

```
classificationLearner
```

Model 1 (Bagged Trees)

True Class	Hail	2901	10	761
	Heavy Rain	45	246	75
	Thunderstorm Wind	940	57	2716
		Hail	Heavy Rain	Thunderstorm Wind
		Predicted Class		

6) View a Report of the Analysis

7) Next Steps - Available Resources

```
web('www.mathworks.com');
```