

## Step 2: Reconstruct Lake Michigan (Zone 0) from Linear to 2D

In this section, we will convert the data in linear form to 2-Dimensional arrays. This will help us to visualize and perform statistical analysis.

```
In [1]: import os
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from scipy.stats import skew
from tqdm import tqdm
```

### TO-DO:

Make sure to change the directory to the outputs generated from  
`01_extract_Lake_Michigan_imagery_data.ipynb`.

```
In [2]: os.getcwd()

## TO-DO: Change the directory here.
os.chdir("GOES_Hourly_Statistics/zone_0_2010Fall_2011Spring")
os.getcwd()
```

```
Out[2]: '/srv/scratch/NOAA/GOES_Hourly_Statistics/zone_0_2010Fall_2011Spring'
```

```
In [3]: len(os.listdir())
```

```
Out[3]: 15050
```

```
In [4]: filename = os.listdir()
len(filename)
```

```
Out[4]: 15050
```

```
In [5]: dataset = pd.read_csv('goes11.2010.10.01.1800.v01.nc-var1-t0.csv')

dataset.head()
```

Out [5]:

	corresponding row	value	datetime	latitude	longitude	partition
0	7942	0.0975	2010-10-01 18:00:00	41.78	-87.54	0
1	7943	0.1025	2010-10-01 18:00:00	41.78	-87.50	0
2	7944	0.1000	2010-10-01 18:00:00	41.78	-87.46	0
3	7945	0.0925	2010-10-01 18:00:00	41.78	-87.42	0
4	7946	0.0825	2010-10-01 18:00:00	41.78	-87.38	0

```
In [6]: dataset = pd.read_csv('goes11.2010.10.01.1800.v01.nc-var1-t0.csv')
lat_u = sorted(dataset['latitude'].unique())
lat_u
```

```
Out[6]: [41.78,  
         41.82,  
         41.86,  
         41.9,  
         41.94,  
         41.98,  
         42.02,  
         42.06,  
         42.1,  
         42.14,  
         42.18,  
         42.22,  
         42.26,  
         42.3,  
         42.34,  
         42.38,  
         42.42,  
         42.46,  
         42.5,  
         42.54,  
         42.58,  
         42.62,  
         42.66,  
         42.7,  
         42.74,  
         42.78,  
         42.82,  
         42.86,  
         42.9,  
         42.94,  
         42.98,  
         43.02,  
         43.06,  
         43.1,  
         43.14,  
         43.18,  
         43.22,  
         43.26,  
         43.3,  
         43.34,  
         43.38,  
         43.42,  
         43.46,  
         43.5,  
         43.54,  
         43.58,  
         43.62,  
         43.66,  
         43.7,  
         43.74,  
         43.78,  
         43.82,  
         43.86,  
         43.9,  
         43.94,  
         43.98,
```

44.02,  
44.06,  
44.1,  
44.14,  
44.18,  
44.22,  
44.26,  
44.3,  
44.34,  
44.38,  
44.42,  
44.46,  
44.5,  
44.54,  
44.58,  
44.62,  
44.66,  
44.7,  
44.74,  
44.78,  
44.82,  
44.86,  
44.9,  
44.94,  
44.98,  
45.02,  
45.06,  
45.1,  
45.14,  
45.18,  
45.22,  
45.26,  
45.3,  
45.34,  
45.38,  
45.42,  
45.46,  
45.5,  
45.54,  
45.58,  
45.62,  
45.66,  
45.7,  
45.74,  
45.78,  
45.82,  
45.86,  
45.9,  
45.94,  
45.98]

```
In [7]: print(len(lat_u), min(lat_u), max(lat_u))
```

106 41.78 45.98

```
In [8]: dataset = pd.read_csv('goes11.2010.10.01.1800.v01.nc-var1-t0.csv')
lon_u = sorted(dataset['longitude'].unique())
lon_u
```

```
Out[8]: [-87.9,  
-87.86,  
-87.82,  
-87.78,  
-87.74,  
-87.7,  
-87.66,  
-87.62,  
-87.58,  
-87.54,  
-87.5,  
-87.46,  
-87.42,  
-87.38,  
-87.34,  
-87.3,  
-87.26,  
-87.22,  
-87.18,  
-87.14,  
-87.1,  
-87.06,  
-87.02,  
-86.98,  
-86.94,  
-86.9,  
-86.86,  
-86.82,  
-86.78,  
-86.74,  
-86.7,  
-86.66,  
-86.62,  
-86.58,  
-86.54,  
-86.5,  
-86.46,  
-86.42,  
-86.38,  
-86.34,  
-86.3,  
-86.26,  
-86.22,  
-86.18,  
-86.14,  
-86.1,  
-86.06,  
-86.02,  
-85.98,  
-85.94,  
-85.9,  
-85.86,  
-85.82,  
-85.78,  
-85.74,  
-85.7,
```

```
-85.66,  
-85.62,  
-85.58,  
-85.54,  
-85.5,  
-85.46,  
-85.42,  
-85.38,  
-85.34,  
-85.3,  
-85.26,  
-85.22,  
-85.18,  
-85.14,  
-85.1,  
-85.06,  
-85.02,  
-84.98,  
-84.94,  
-84.9,  
-84.86,  
-84.82,  
-84.78]
```

```
In [9]: filename[0:10]
```

```
Out[9]: ['goes11.2010.12.10.0300.v01.nc-var1-t0.csv',  
        'goes11.2010.12.12.1615.v01.nc-var1-t0.csv',  
        'goes11.2010.12.11.0115.v01.nc-var1-t0.csv',  
        'goes13.2011.03.08.2145.v01.nc-var1-t0.csv',  
        'goes13.2011.02.18.1430.v01.nc-var1-t0.csv',  
        'goes13.2011.01.07.1300.v01.nc-var1-t0.csv',  
        'goes13.2011.03.21.0630.v01.nc-var1-t0.csv',  
        'goes11.2010.12.18.0645.v01.nc-var1-t0.csv',  
        'goes13.2011.02.04.1415.v01.nc-var1-t0.csv',  
        'goes11.2010.10.22.2030.v01.nc-var1-t0.csv']
```

**As shown above, the list of filenames is out of order. Therefore, a sort is needed.**

```
In [10]: filename.sort()
```

```
In [11]: # Just a quick inspection.  
filename[0:10]
```

```
Out[11]: ['goes11.2010.10.01.0000.v01.nc-var1-t0.csv',  
        'goes11.2010.10.01.0030.v01.nc-var1-t0.csv',  
        'goes11.2010.10.01.0045.v01.nc-var1-t0.csv',  
        'goes11.2010.10.01.0100.v01.nc-var1-t0.csv',  
        'goes11.2010.10.01.0115.v01.nc-var1-t0.csv',  
        'goes11.2010.10.01.0130.v01.nc-var1-t0.csv',  
        'goes11.2010.10.01.0145.v01.nc-var1-t0.csv',  
        'goes11.2010.10.01.0200.v01.nc-var1-t0.csv',  
        'goes11.2010.10.01.0215.v01.nc-var1-t0.csv',  
        'goes11.2010.10.01.0230.v01.nc-var1-t0.csv']
```

## TO-DO:

Make sure to change the output directory for the results. The best practice is to add another directory called `zone_0_T`.

```
In [12]: #!/mkdir /srv/scratch/NOAA/GOES_Hourly_Statistics/zone_0_T_2006Fall_2007Spring
```

```
In [13]: for fn in tqdm(filename[:1000]):
    #print('{value}'.format(value=fn[18:22]))
    dataset = pd.read_csv(fn)
    lat_u = lat_u
    lon_u = lon_u
    #lat_list = []
    lon_list = []
    table_dic = {}
    for lat in lat_u:
        #lat_list.append(str(lat))
        lon_ind_list = []
        for lon in lon_u:
            value_ind = dataset[(dataset['latitude'] == lat) & (dataset['longitude'] == lon)]
            if len(value_ind) == 0:
                lon_ind_list.append(None)
            else:
                lon_ind_list.append(str(value_ind[0]))
        table_dic[lat] = lon_ind_list
        lon_list.append(lon_ind_list)
    transform_df = pd.DataFrame(data = table_dic, index = lon_u)
    transform_df = transform_df.T
    df_1617fall = pd.DataFrame(data = transform_df)

    ## Change the directory here.
    df_1617fall.to_csv('/srv/scratch/NOAA/GOES_Hourly_Statistics/zone_0_T_2006Fall_2007Spring')
    # print(fn)
```

```
100%|██████████| 1000/1000 [1:23:11<00:00, 4.99s/it]
```

```
In [14]: # transform_df = pd.DataFrame(data = table_dic, index = lon_u.tolist())
# transform_df = transform_df.T
#transform_df
```

```
In [15]: # df_1416fall = pd.DataFrame(data = transform_df)
# df_1416fall.to_csv('E://sensing//2014Fall-2016Spring_CSV//2014Fall-2016Spring')
```

```
In [16]: df_1617fall
```



Out[16]:

	-87.90	-87.86	-87.82	-87.78	-87.74	-87.70	-87.66	-87.62	-87.58	-87.54	...	.
41.78	None	None	None	None	None	None	None	None	None	0.0025	...	
41.82	None	None	None	None	None	None	None	None	None	0.0025	...	
41.86	None	None	None	None	None	None	None	None	0.0025	None	...	
41.90	None	None	None	None	None	None	None	None	0.005	None	...	
41.94	None	None	None	None	None	None	None	0.0025	0.01	None	...	
...	...	...	...	...	...	...	...	...	...	...	...	
45.82	None	None	None	None	None	None	None	None	None	None	...	
45.86	None	None	None	None	None	None	None	None	None	None	...	
45.90	None	None	None	None	None	None	None	None	None	None	...	
45.94	None	None	None	None	None	None	None	None	None	None	...	
45.98	None	None	None	None	None	None	None	None	None	None	...	

106 rows × 79 columns

In [ ]:

In [ ]: