Lab01

The goal is to extract some data (anything) from the NDAWN website.

following web tutorial

https://www.youtube.com/watch?v=E5cSNSeBhjw

In [18]:
```
import requests
from bs4 import BeautifulSoup
import pandas as pd
import os
```

In [25]:
```
#directions to the NDAWN wesite via its URL
page = requests.get("https://ndawn.ndsu.nodak.edu/current.html")
soup = BeautifulSoup(page.content, "html.parser")
#print(soup)

#Goal, extract current weather data from the 'Current Weather Table" (has data for 168 stations throughout ND)
table = soup.find(id ='table')
```

In [26]:
```
#telling BeutifulSoup what webpage I want to grab data from
page = requests.get("https://ndawn.ndsu.nodak.edu/current.html")
soup = BeautifulSoup(page.content, "html.parser")

#locating Current Weather table and it values
table = soup.find(id ='table')

#extracting several variables and their values from the 'Current weather table'
stationlist = table.find_all(class_ = 'station')
lastupdate = soup.find_all(class_ = 'update')
airtemp = soup.find_all(class_='hasunits cur left')
winddir = soup.find_all(class_='winddir')
#this variable double counts, because 2 columns within the table use the 'class_='cur windspd''
#I need to seperate the values using the following for loop (order matters!)
windspeed = soup.find_all(class_='cur windspd')#double counts
c_windspeed=[]
p_windspeed=[]
number = 0
for item in windspeed:
    number+=1
    if number%2 ==0:
        c_windspeed.append(item)
    else:
        p_windspeed.append(item)
#last variable
relhum = soup.find_all(class_='cur hum')
```

In [27]:
```
#checking to make sure the extractions are the right length
#(there should be 168 oberservations per list because that is how many stations there are in ND)
list_of_variables=[stationlist,lastupdate,airtemp,winddir,c_windspeed,p_windspeed,relhum]
for variable in list_of_variables:
    print(len(variable))
```

```
168
168
168
168
168
168
168
```

In [28]:
```
#Using list comprehensions to isolate the text form the rest of the html within the lists (only need the text)
station_names = [text.get_text() for text in stationlist]
time_acquired = [text.get_text() for text in lastupdate]
Air_temp = [text.get_text() for text in airtemp]
Wind_dir = [text.get_text() for text in winddir]
Cur_wind_spd = [text.get_text() for text in c_windspeed]
Peak_wind_spd = [text.get_text() for text in p_windspeed]
Rel_hum = [text.get_text() for text in relhum]
```

In [29]:
```
# converting all these variables to pandas dataframe
weather_table = pd.DataFrame(
    {'Stations' : station_names,
    'Date of acquisistion': time_acquired,
    'Air temp': Air_temp,
    'Wind Direction': Wind_dir,
    'Current Wind Speed': Cur_wind_spd,
    'Peak Wind Gust': Peak_wind_spd,
    'Relative Humidity': Rel_hum})

weather_table
```

Out[29]:

	Stations	Date of acquisistion	Air temp	Wind Direction	Current Wind Speed	Peak Wind Gust	Relative Humidity
0	Ada 1N	02 Oct 11:05 CDT	59°	NNW	9 mph	4 mph	88 %
1	Adams 5N	02 Oct 11:05 CDT	62°	NNE	6 mph	4 mph	62 %
2	Alamo 2S	02 Oct 11:05 CDT	60°	NNW	11 mph	8 mph	50 %
3	Alexander 7SW	02 Oct 11:05 CDT	59°	NNW	10 mph	8 mph	49 %
4	Alvarado 4N	02 Oct 11:05 CDT	59°	NNW	10 mph	8 mph	86 %
...	...	...	...	...	...	...	...
163	Williston 5SW	02 Oct 11:05 CDT	62°	ESE	4 mph	2 mph	47 %
164	Wishek 5W	02 Oct 11:05 CDT	62°	NW	9 mph	7 mph	69 %
165	Wolford 4E	02 Oct 11:05 CDT	62°	WSW	5 mph	3 mph	73 %
166	Wolverton 2E	02 Oct 11:05 CDT	61°	N	12 mph	8 mph	87 %
167	Zeeland 7NE	02 Oct 11:05 CDT	63°	NNW	10 mph	7 mph	72 %

168 rows × 7 columns

In [30]:
```
#exporting data to a csv file in my 'NDAWN_data' folder
path = r'C:\Users\runac\Downloads\Fall_2021\ArcGIS1\Labs\Lab01\NDAWN_data'

weather_table.to_csv(os.path.join(path,'ND_weather_oct_2-1105.csv'))
```

In [ ]:
```

```