**[Numpy](https://github.com/HAS-Tools-2023/Course-Materials23/blob/main/Assignments/Week9_Section1.7_CodeFormatting.md" \l "pandas)**

* The general purpose of the numpy library

**Numpy is used to calculate statistics and other mathematical functions.**

* What numpy array's are and what kinds of things they can (and can't contain)

**It is a multidimensional array containing only numerical values. It cannot contain letters or strings. There is also no row or column names.**

* Various approaches for how to slice a numpy array (e.g. grabbing out a row, column, range of values)

**Numpy is sliced by indexing.**

**1D Array 5th index: Array\_name[6]**

**Row: Array\_name[ row\_#, :]**

**Cell: Array\_name[ row\_#, column\_#]**

**Column: Array\_name[ :, column\_#]**

* How to create a numpy array from scratch (please show at least 2 options)

1. **sample\_array = np.array([[23, 46, 85],**

**[43, 56, 99],**

**[11, 34, 55]])**

1. **array=np.arange(0, 5000, 500)**

**array\_1=np.asarray(array)**

* List 5-6 helpful numpy functions and what they do

1. **Np.random() : generates a random array of certain define dimensions**
2. **np.reshape(array, dimensions) : takes an input array that may be like a 1D array and puts it into what ever dimension type defined**
3. **np.mean() : the average of the array values**
4. **np.ones() : an array of defined dimensions of only 1’s as the values**
5. **array\_name.flatten() : makes a multi dimensional array into a 1D array**
6. **np.nanmean() : skips over the nan values while calculating the mean value**

[**Pandas**](https://github.com/HAS-Tools-2023/Course-Materials23/blob/main/Assignments/Week9_Section1.7_CodeFormatting.md#pandas)

* The general purpose of the Pandas library

**Pandas handles different types of data (integers, strings, etc.) and can put them in a data frame like a matrix altogether. In this class, it makes columns of different things like dates, flows, months. It also is able to be plotted and read data in from a CSV file.**

* What makes a pandas data frame is different from a numpy array

**Pandas focuses on a 2D tabular data structures and locates items in the table by using the labels of pieces of data rather than its position. Numpy is numerical based. Numpy also has no row and column names but pandas does.**

* An explanation of what the 'index' of a dataframe and why its different from other columns

**The index is like the header of the row and is the x value. All other columns are based off the index. Columns are dependent on the index like time being the index to flow that is dependent on it.**

* How to setup a pandas dataframe by reading a file

**data\_frame = pd.DataFrame(data,**

**columns=['flow', 'year', 'month', 'day'])**

* How to set the index of a pandas dataframe

**df = df.set\_index('datetime')**

* How to slice a pandas dataframe:
  + using loc and iloc to get rows

**iloc: based on numerical location**

**->new\_df.iloc[0:8:2, 0:3:2] = 0**

**loc: based on column names**

* **data\_frame.loc[['a', 'e']] = 3**
  + grabbing out columns by name or number

**new\_df.loc[['a', 'c', 'e', 'g'], ['data1', 'data3']] = 0**

* 5-6 helpful pandas functions or methods that you can use to inspect your dataframe (list each and explain what it does)

1. **.describe : tells all of the statistical values for the chosen columns**
2. **.shape : number of rwos and columsn which is important when handling big data sets**
3. **.dropna() : one of my favorites because it drops the null values**
4. **.sort\_values : helps to put items in ascending order, especially when dealing numerically**
5. **.groupby : helped me make my October dataframe by grouping all common things and pulling it out. (group by operation using the job title column to get the mean salary corresponding to each job title.)**
6. **.head(): gives the first 5 rows in the dataframe**