What are they and how are they different from the other object types we have worked with so far

- pandas: open source data analysis and manipulation tool. It can create dataframes and series using any type of object (integers, strings, etc.).
    - In comparison to numpy, numpy only works with numerical data.

How to make a pandas dataframe from scratch:

- df = pd.DataFrame(np.random)
- df = pd.Dataframe(

    [[1,2,3],

    [4,5,6],

    [7,8,9]]

    index = [1,2,3]

    column = [‘a’, ‘b’, ‘c’])
    - Output:

	a	b	c
1	1	2	3
2	4	5	6
3	7	8	9

How to read-in different data as a pandas dataframe (they are almost the same format):

- CSV:
    - pd.read_csv(filename)
- Table:

        - pd.read_table(filename)
    - Excel File:
        - pd.read_excel(filename)
    - JSON:
        - pd.read_json(json_string)

How to slice pandas dataframes -- both using loc and iloc

- Pandas Indexing and how it differs from other columns:
    - The index for the pandas dataframe is more specific than other indexes. The column and index values can be specified to contain specific variables (i.e. the actual names of the month instead of 1-12). Numpy indexes by the rows and columns (#, #) while Pandas allows you to choose the different items within the dataframe based on the given columns.
- How do we index (and slice) using loc and iloc?
    - loc vs. iloc
        - loc is label based while iloc is integer based.
    - Select a certain position in a series:
        - s.iloc[#]
    - Select a certain index in a series:
        - s.loc[#]
    - Select a row:
        - df.iloc[0,:]
    - Select an element:
        - df.iloc[#,#]
        - s.loc[#:#]

- Functions:
    - head()
        - This gives the first 5 (or more/less if specified) rows of data.
    - describe()
        - This is used to find the min, max, mean, and percentiles.
    - astype()
        - This lets you view the different data types in order to change all the necessary values to the same data type.
    - groupby()
        - This groups things by a specific column to do math on. It can be just one single column or multiple and be multiplied by 8, for example.
    - sort_values()
        - This sorts the columns in the dataframe