- What are they and how are they different from the other object types we have worked with so far
    - Pandas series (1-D) & dataframes (2-D) are objects that can store data in table form, and can store multiple data types (unlike numpy arrays)
- How to make a pandas dataframe from scratch & by reading in a csv
    - See pictures

```
8   # Create a basic dataframe from scratch
9   dataframe_1 = pd.DataFrame(columns=["Precip_mm", "ET_mm"],
10                             data=[
11                                  [0.3, 0.7],
12                                  [0.2, 0.5],
13                                  [0.2, 0.9]
14                                  ])
```

```
17  # Create a dataframe by reading in a file
18  filename = 'streamflow_week8.txt'
19  filepath = os.path.join('../../working_drafts/data', filename)
20  print(os.getcwd())
21  print(filepath)
22
23  # Read the data into a pandas dataframe
24  data = pd.read_table(filepath, sep='\t', skiprows=30,
25                       names=['agency_cd', 'site_no', 'datetime',
26                              'flow', 'code'],
27                       parse_dates=['datetime'], index_col=['datetime']
28                       )
29
```

Create a cheat sheet on Pandas basics that covers the following:

- Summarize Pandas
    - How to slice pandas dataframes -- both using loc and iloc
        - .iloc = uses the index location within a dataframe to select data
            - Have to give row selection and then column selection
                - Ex: dataframe_1.iloc[0:2, 0:2] gives the first 2 rows in the first two columns of the dataframe
        - .loc = uses the “name” of an index label to select data, whatever the label of the index values happen to be
    - What is the index of a pandas dataframe -- why is it different than other columns and how can you work with it?
        - Index won’t show up as a column if trying to call it as a column
            - Can use dataframe_1[[‘column1’]] to pull just the data from column 1 into a new dataframe
        - Can use .iloc or .loc to access the index
    - Key methods associated with pandas dataframes
        - .set_index() = sets a designated column as an index
        - .head() = default gives the first 5 rows of data in the dataframe
        - .tail() = default gives the last 5 rows of data in the dataframe
        - .describe() = gives summary statistics of the dataframe
    - Key attributes associated with pandas dataframes
        - .index = lists all the index values that are set in a dataframe
        - .columns = lists all the column names that are in the dataframe

Pandas Cheat Sheet