



Pandas Cheat Sheet

- Pandas is a package and is imported like this below:

```
import pandas as pd
```

- "pd" is an alias for the package so when you call it every time in your code, you don't have to type out "pandas" every time
- a series is a 1D dataframe which is list of numbers (similar to a 1D numpy array)
- a dataframe is tabular, meaning that it is a matrix of numbers (similar to a 2D numpy array)
- what's unique to dataframes is that there can be labels for rows and names for columns, making indexing through the dataframe a lot easier
- you can also have different types of data in a dataframe, meaning that you could have strings, floats, or ints all in one
- common attributes:
 - dim, shape, size
 - columns → gives us the names of the columns
 - index → gives us the index values
- indexing with dataframes:
 - this can either be done via location or via labels
 - location based:
 - use "iloc"
 - the same as index location ([start:stop:step])
 - use the numbers of the columns/rows

```
my_data_frame.iloc[row_start:row_stop:row_step, col_start:col_stop:col_step]
my_data_frame.iloc[0, 1:3]
```

- label based:

- grabs specific rows/columns by name!
- this is much easier
- use "loc" for row name(s)

```
my_data_frame.loc['row_name', :]
```

- to grab columns, you actually don't have to use loc (see below)

```
my_data_frame['column_name']
```

- common methods:

- head() → returns the first entries (5 if you don't specify)
- tail() → returns the last entries (5 if you don't specify)
- sort_value → ascending or descending
- groupby → let's you group according to column values
- median, mean, sum, max, min

- with dataframes, you can also change the index column to really anything you want!

- dates are common when working with large sets of data

- to read a csv file into a dataframe, you

```
filename = 'streamflow_week8.txt'
filepath = os.path.join('../data', filename)
print(os.getcwd())
print(filepath)

data = pd.read_table(filepath, sep='\t', skiprows=30,
                    names=['agency_cd', 'site_no', 'datetime', 'flow', 'code'],
                    parse_dates=['datetime']
                    )
```

- to make a new dataframe, simply use `pd.DataFrame()`
 - to add to the dataframe, make columns using brackets and enter values

```
df = pd.DataFrame()

df['column1'] = ['a', 'b', 'c']
df['column2'] = [1, 2, 3]
df['more_descriptive_name'] = ['hi', 'yo', 'this is an example']
```