Pandas Dataframe

import pandas as pd

Great website
https://www.danielmsullivan.com/pages/research.html

BIG HELP for basic data manipulation given Stata syntax is so simple
https://www.danielmsullivan.com/pages/tutorial_stata_to_python.html

Documentation online
https://pandas.pydata.org/docs/user_guide/10min.html

Sort within defined groups in a Dataframe
df.sort_values(['column'],ascending=False).groupby('another_column')

Show a list of all columns within a Dataframe
df.columns

Get unique values in a column and the frequency of those values
df['Name"].value_counts()

Create binary True/False Variable (or use np.where)
df['title_bout'] **=** df['Fight_type'].apply(**lambda** X: **True if** 'Title Bout' **in** X **else False**)

- Create binary/TrueFalse variable as new variable

df.loc[df[‘col1’] *condition*, ‘col2’]

- Find the value of another column based on some condition for filtering
- Ex: get the names of all people who are 20yrs old or less

Criteria when dealing with multiple columns
Use .loc or np.where or something else here

Drop Random duplicate observations

df = df.sample(frac=1).drop_duplicates()

Here, we are taking a sample equal to the full size of the dataframe, without replacement. This effectively shuffles the position of all rows, allowing us to drop duplicates and keeping the first row, previously randomized.

Str Time documentation pandas
https://docs.python.org/3/library/datetime.html#strftime-and-strptime-behavior

- Convert various formats to a pandas date that can be sorted, etc.

Charting - Matplotlib

Creating charts for data

Numpy

Get index of a numpy array
arr[row:col]

Beautiful Soup (bs4)

See Page source HTML

Press F12

Get HTML from page to parse

page = requests.get(url)
Soup = BeautifulSoup(page.text, 'html.parser')