Data Manipulation

.csv file

Comma Separated Version (.csv)

- Can be opened with Excel or similar spreadsheet tool
- Created to hold large amounts of data

Demo

- import csv
- reader()
- writer()
- dictReader()
- dictWriter()
- writeheader()
- writerows()

pandas

- Data analysis library
- Pandas library

Demo

- pandas
- dataframe
- Pandas and csv files

Data

- Temperature change in Canada 1961-1990 (link)
- Temperature Change in Canada – Seasonal 1961-1990 (link)
- Precipitation change in Canada 1961-1990 (link)
- Precipitation Change in Canada – Seasonal 1961-1990 (link)
- Aleutian Low Pressure Index (ALPI) (link)

In-class assignment 8a

1. Create a python program that will add the following lines to the example 1 file:
   - `['George','September 22','EECS']` and `['Diana','December 13','STATS']`
   - Change Diana's department to 'ECON'
2. Create a python program that:
   - Download both the temperature and precipitation .csv files
   - Combines the two files into a pandas dataframe
   - Remove all the rows with NaN values from the dataframe
     - Pandas docs: 1, 2, 3

In-class assignment 8b

Using the database creator program we made last week, add 2 new functions

1. Add a function to read a dictionary from a file
2. Add a function to write a dictionary into a file

This will create a ‘database’ that can be modified.

Assignment 8

1. Write a Python program that takes in seasonal precipitation and temperature data from the provided .csv files.
    - Take these two sets of data and create one dataframe called ‘weather’ and remove all empty fields (NaN)
    - Calculate the total change of precipitation and saves it into a new column called ‘Annual (P)’.
    - Calculate the total change of temperature and saves it into a new column called ‘Annual (T)’.
    - Save into a new .csv file called ‘seasonal_weather.csv’

Links: 1

Assignment 8

2. Write a Python program that takes a snapshot of the data in ‘seasonal_weather.csv’ and saves it into a ‘weather_snapshot.csv’ file.

- The snapshot file should be the data from year 1990 to the latest possible year
- Add a new column called ‘Pressure’ and fill it with the ALPI data from only using the years 1990, make sure there are no NaN values