

LABO

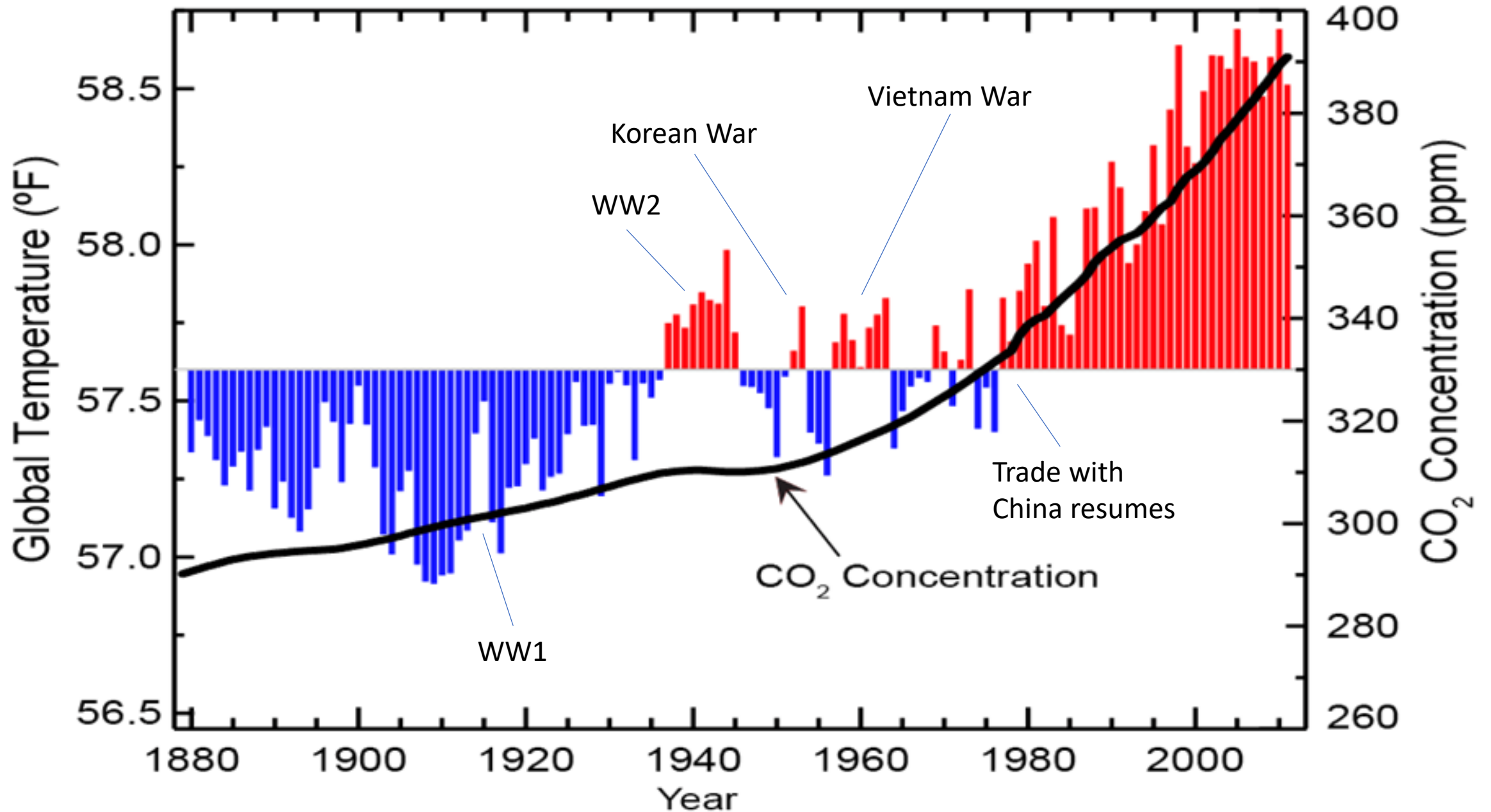
DATA MINING



Specification

- Write a Python application that displays the Earth's temperature as related to the Carbon emissions.
- Display the yearly carbon emission and global annual temperature for the years where the data overlaps.
- Output the slope of the Linear Regression line.

Global Temperature and Carbon Dioxide



Data Files

- Co2.html: Monthly CO2 carbon emissions from 1959 to 2018.
- Temperature.html: Annual global Temperatures from 1858 to 2018.

	A	B	C	D	E	F	G	H
1	CO2							
2	Year	Month	Decimal	Average	Interpolat	trend	#days	
3	1959	1	1959.042	315.62	315.62	315.71	-1	
4	1959	2	1959.208	316.71	316.71	315.62	-1	
5	...							
6	Temperature							
7	Year	Median	Upper	Lower				
8	1850	-0.373	-0.339	-0.425				
9	1851	-0.218	-0.184	-0.274				
10								

Pre-Processing Data

- The Temperature data is based on monthly values and the CO2 data is annual data.
- Refactor the Temperature data into annual values so the Temperature and CO2 can be compared in the same year units.
- Average the monthly temperature data into yearly data.
- The temperature is expressed in terms of degrees from the average temperature from 1960-1990.
- A positive temperature indicates degrees above the average.
- A negative temperature indicates degrees below the average.

Storing the Data

- Store the CO2 and Temperature data in separate dictionaries.
- Store CO2 and Temperature dictionary in a DataBase container. Use a primitive container like a List, Tuple or Dictionary.

- For example:

```
co2 = { key:value, ...
```

```
temp = { key:value, ...
```

```
database = [ co2, temp ]
```

Reading the Data

- Write a utility class to do file operations like opening a file, reading a file, writing a file and closing a file.
- The read data should be stored in a RAW DATA container, consisting of a [List] of str(strings).
- Return the RAW DATA container from the Read function.

Processing the Data

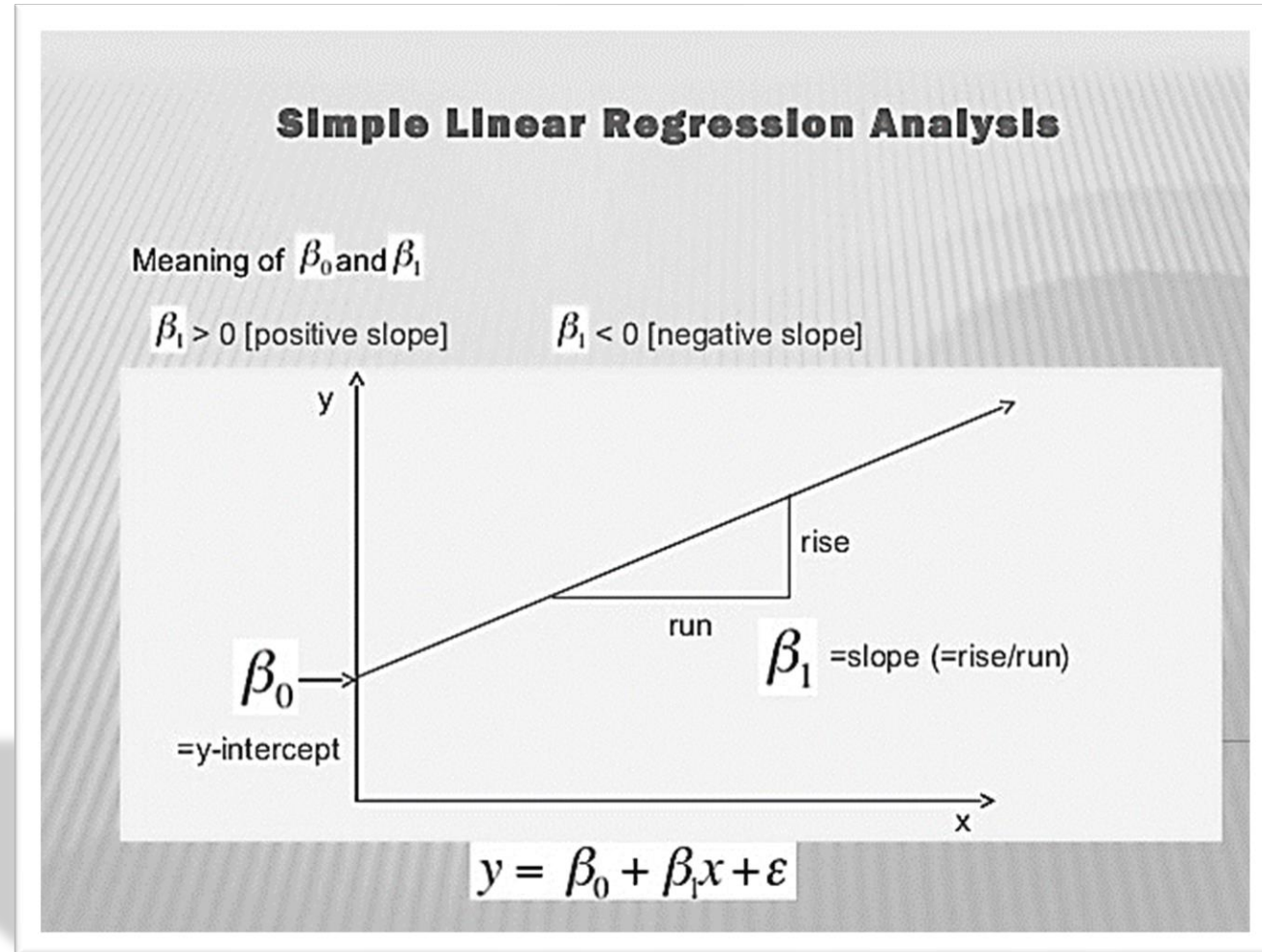
- Use Regular Expressions to parse the RAW input HTML data files.
- Don't use HTML parsers, like BeautifulSoup, to parse the data.
- Create a class to handle HTML processing using Regular Expressions.

Output

- Process the data to determine the Slope of the Linear Regression line for the data, where the X axis is the CO2 level and the Temperature, Y axis, is the dependent variable.
- Note: Using the module SKLearn, it's possible to find the Slope function for Linear Regressions. Write your own version of the Slope function and then use the SKLearn version to verify your results are the same.
- See this website for information about linear regressions:

<https://www.wallstreetmojo.com/regression-formula/>

Regression Slope



Lab0 Object Factory Sequence

- Input: CO2 and TEMP data
- Process: Parse the data
- Store: Store the data in application database
- Run: The Process
- Output: Slope of line

