Climate Change Explained through Earth Surface Temperature Data

By: Fauzan Amjad

Describing the Dataset

- https://kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data?utm_medium=social&utm_campaign=kaggle-dataset-share&utm_source=twitter
- A clean and data-saturated dataset with 13 different files from Berkeley Earth (Lawrence Berkeley National Laboratory)
  - 1.6 billion temperature reports
- Description: The dataset compiles temperature reports (high, low, and average temperatures) from 1750 (Land data) or 1850 (land and ocean data) to the present time, is adjusted for regional averages, corrected in the case of biases, and is sorted by location as indicated by the different files available.
- Attributes: Year, City, State, Country, High Temperature, Low Temperature, Average Temperature, Land/Ocean, and Uncertainty (Confidence Interval)

What is in this data set?

	Land + Ocean anomaly using air temperature above sea ice				Land + Ocean using water temperature below sea ice			
Year,	Annual Anomaly,	Annual Unc.,	Five-year Anomaly,	Five-year Unc.,	Annual Anomaly,	Annual Unc.,	Five-year Anomaly,	Five-year Unc.
1850	-0.463	0.177	NaN	NaN	-0.424	0.160	NaN	NaN
1851	-0.346	0.151	NaN	NaN	-0.321	0.136	NaN	NaN
1852	-0.340	0.154	-0.367	0.112	-0.306	0.140	-0.336	0.101
1853	-0.371	0.129	-0.334	0.100	-0.339	0.116	-0.303	0.091
1854	-0.317	0.142	-0.349	0.094	-0.290	0.128	-0.316	0.086
1855	-0.298	0.135	-0.393	0.092	-0.257	0.122	-0.357	0.084
1856	-0.420	0.120	-0.400	0.096	-0.388	0.110	-0.364	0.087

- Screenshot of one file within the dataset (http://berkeleyearth.lbl.gov/auto/Global/Land_and_Ocean_summary.txt)
- Years Range: 1850 – 2020 (Land and Ocean Data)
- For each year, this file of the data set shows the changes of Earth's global average surface temperature which is calculated through the combination of Lawrence Berkeley National Laboratory's land-surface data and HadSST's ocean temperature data.

What are other files in this data set?

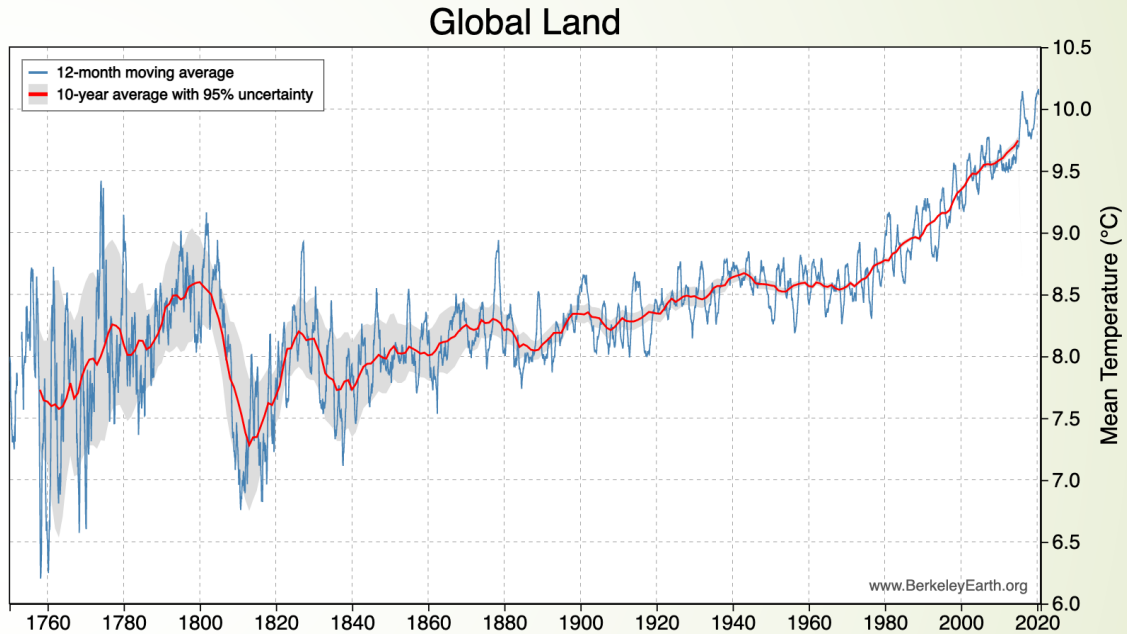
- Files (Name Implies Characterization):
  - GlobalLandTemperaturesByCountry.csv
  - GlobalLandTemperaturesByState.csv
  - GlobalLandTemperaturesByMajorCity.csv
  - GlobalLandTemperaturesByCity.csv
  - GlobalAverages.csv
    - Includes the global land temperatures averages in Celsius
    - Uses 95% confidence as a threshold

	A	B	C
1	dt	LandAverageTemperature	LandAverageTemperatureUncertainty
2	1750-01-01	3.034	3.574
3	1750-02-01	3.083	3.702
4	1750-03-01	5.626	3.076
5	1750-04-01	8.49	2.451
6	1750-05-01	11.573	2.072
7	1750-06-01	12.937	1.724
8	1750-07-01	15.868	1.911
9	1750-08-01	14.75	2.231
10	1750-09-01	11.413	2.637
11	1750-10-01	6.367	2.668
12	1750-11-01		
13	1750-12-01	2.772	2.97
14	1751-01-01	2.495	3.469
15	1751-02-01	0.963	3.827
16	1751-03-01	5.8	3.051
17	1751-04-01	7.67	2.368
18	1751-05-01		
19	1751-06-01	13.827	1.801
20	1751-07-01		
21	1751-08-01	14.405	2.296
22	1751-09-01	10.673	2.656

GlobalAverages.csv

Applications of the Dataset

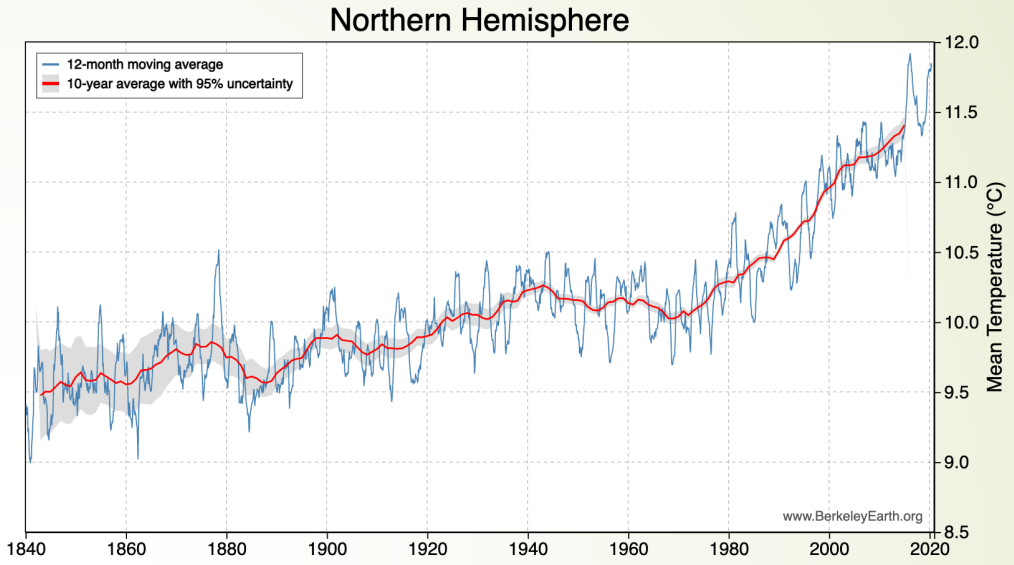
- What were the hottest monthly and annual averages on record?
  - The hottest February's were in the years 2016, 2020, 2017, 1998, 2002, 1999, 1995, 2015, 2004, and 2006 in that order.
  - The hottest annual temperatures were 2020, 2016, 2019, 2017, 2015, 2018, 2007, 2005 2010, and 2013 in that order.
    - 8/10 of the hottest years came in the last 10 years.



Source: Berkeley Earth (Dataset Provider)

Applications of the Dataset (cont.)

- What is the average rate of change of temperature record of the Northern Hemisphere in the last two centuries?
  - 1990: 3.56 ± 0.20
  - 1960: 2.56 ± 0.20
  - 1860: 0.96 ± 0.10



Mean Rate of Change (°C / Century)

Since:	1760	1810	1860	1910	1960	1990
Northern Hemisphere	-	0.79 ± 0.16	0.96 ± 0.10	1.19 ± 0.05	2.56 ± 0.20	3.56 ± 0.20
Global Land	0.43 ± 0.18	0.79 ± 0.13	0.91 ± 0.08	1.11 ± 0.03	2.16 ± 0.11	2.78 ± 0.13

Source: Berkeley Earth (Dataset Provider)

What are interesting questions that the data set can help answer?

- Although the dataset confirms the increase of average temperatures overtime, the change in average temperature will differ depending on the location of the world we are looking at. Which countries will be negatively impacted by increasing average temperatures?
  - Using this dataset, we can not only see which countries/regions have the highest change in average temperatures, but also how much they potentially contribute in change of global temperature.
  - This would be useful in pinpointing locations that require assistance in alleviating the negative effects of increasing average regional temperatures on the standard of living.
- Some claim that climate change is not a result of human activity, but rather seasonal changes that Earth simply exhibits, having high temperatures in some periods and low temperatures in others (like a sin(x) graph). Is this conjecture supported by this dataset?
  - In theory, the dataset of global average temperatures should have a point of inflection to establish that the rate of change changes from positive to negative or negative to positive if this conjecture was true.

What would I do with the data set?

- Within the United States, public views of climate change often differ: climate change is the biggest threat to humanity, climate change doesn't exist at all, or climate change exist but its exaggerated to some extent.
  - I would use this dataset along with datasets that include data on greenhouse gas emission and fossil fuel growth and regression model or correlation tests to see if there exists a correlation between the average increase in temperature and the average increase/decrease in greenhouse emission or fossil fuel use.
  - I would also use the data set to see if years that had statistically significant increases of average global/regional temperatures had a statistically significant increase in extreme natural events in those respective areas.
- I would also use the average ocean temperature data that is offered by the data set in conjunction with datasets that provide the acidity of surface ocean waters and run correlation test or regression models.
  - Both exhibit an increase over the past few decades; however, there is likely a confounding value – an increase in carbon dioxide.