

Analyzing Global Temperatures

By Aletia Trepte

- [→ Purpose](#)
- [→ Introduction](#)
- [→ General Summary](#)
- [→ Process](#)
- [→ Temperatures by Year](#)
- [→ 10-Year Moving Average Temperatures by Years](#)
- [→ Observations](#)
- [→ References](#)

Purpose & Introduction

This is the first project for Udacity Data Analyst Nanodegree Course, Explore Weather Trends, that analyzes local temperature trends to overall global temperature trends.

Data was generated from the SQL workspace in Udacity course pages and was filtered by city and average temperatures. The two files queried were `city_data.csv` and `global_data.csv` by way of the following SQL queries:

SQL Query 1:

```
SELECT year, avg_temp FROM city_data
```

```
WHERE city = 'San Diego';
```

SQL Query 2:

```
SELECT * FROM global_data;
```

General Data Summary

city_data Information

RangeIndex: 165 entries, 0 to 164

Data columns (total 2 columns):

#	Column	Non-Null Count	Dtype
0	year	165 non-null	int64
1	avg_temp	165 non-null	float64

global_data Information

RangeIndex: 266 entries, 0 to 265

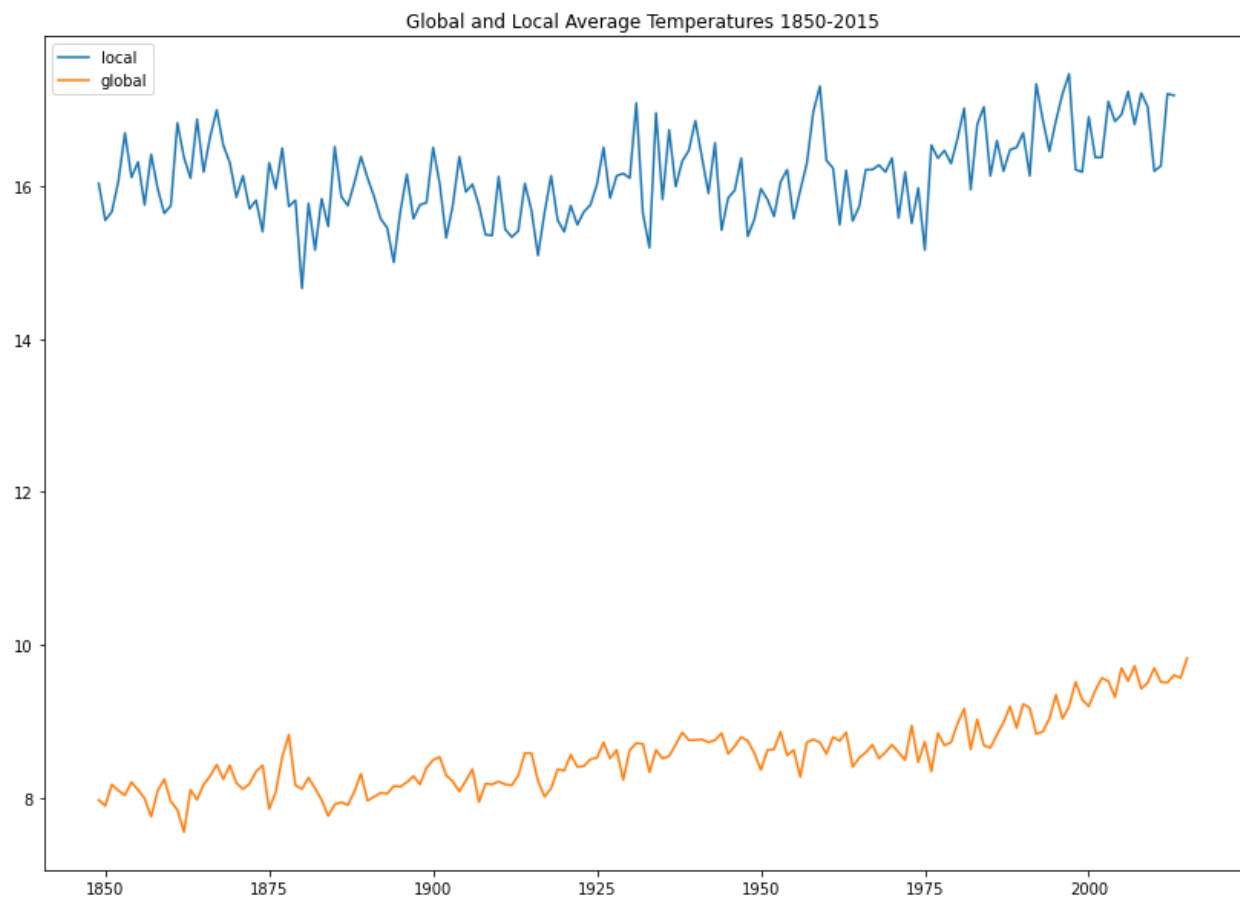
Data columns (total 2 columns):

#	Column	Non-Null Count	Dtype
0	year	266 non-null	int64
1	avg_temp	266 non-null	float64

Process

- Tools used:
 - SQL
 - Python, Jupyter Notebook
 - Pandas, NumPy, matplotlib, cProfile
 - Version Control: Git
 - Github: <https://github.com/parcheesime/Temperature-Trends>
- Dates were not matched between datasets
 - Global data years used 1849-2013
 - `gdf.loc[gdf['year']>1848],`
 - `gdf.loc[gdf['year']<2014]`
 - City data years used 1849-2013
- Moving average:
 - MA calculated using `pandas.DataFrame.rolling()`
 - `ldf['10 year MA']=ldf.avg_temp.rolling(10).mean()`
 - `gdf['10 year MA']=gdf.avg_temp.rolling(10).mean()`
 - Used 10-year average
- Key considerations when visualizing the trends:
 - Global & Local Temperatures in Celcius
 - Actual Temperatures
 - As opposed to temperature anomalies
 - City selected, San Diego
 - Showed overall warmer temperatures
 - San Diego's climate is considered, a 'Mediterranean'
 - General temperature trends similar to global trends
 - Both global and city temperatures rise over time

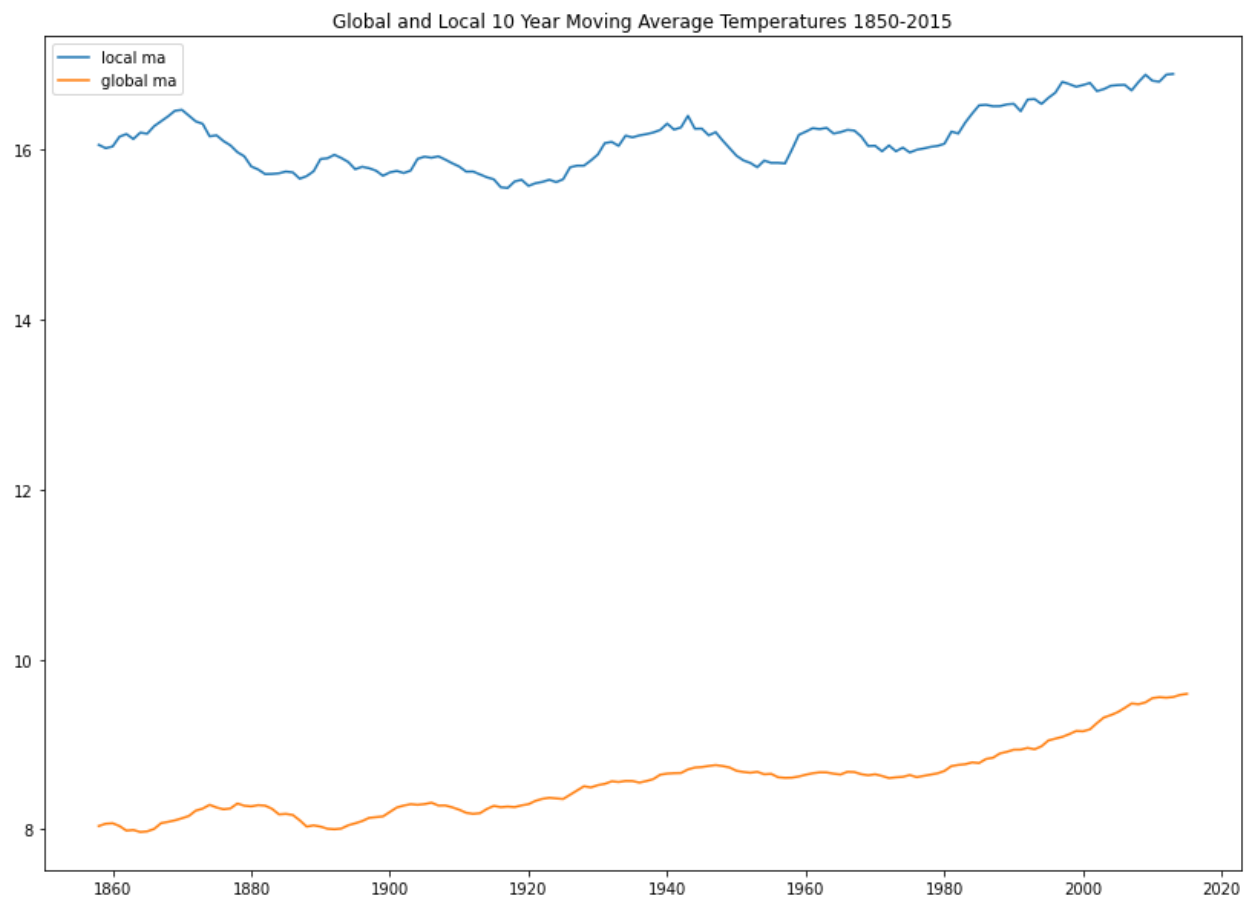
Temperatures by Year



Python code:

```
plt.figure(figsize=(14,10))  
plt.title('Global and Local Average Temperatures 1850-2015')  
plt.plot(lidf['year'], lidf['avg_temp'], label='San Diego Temperatures')  
plt.plot(gdf['year'], gdf['avg_temp'], label='Global Temperatures')  
plt.legend(('local', 'global'), loc='upper left')
```

10-Year Moving Average



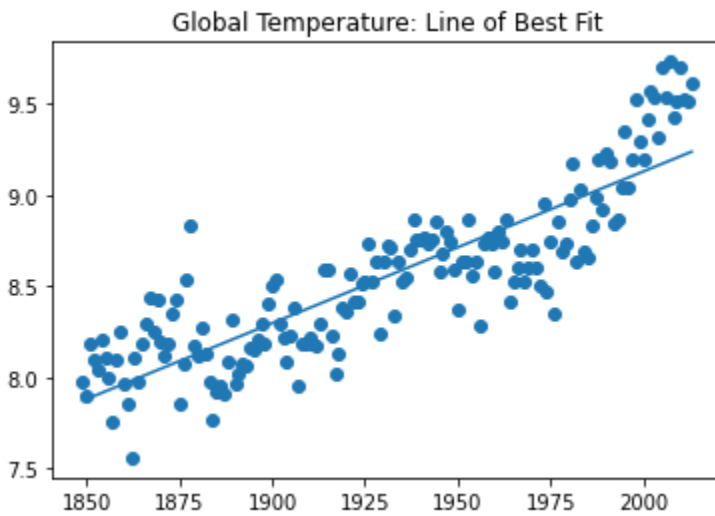
Python code:

```
plt.figure(figsize=(14,10))  
plt.title('Global and Local 10 Year Moving Average Temperatures 1850-2015')  
plt.plot(lidf['year'], lidf['10 year MA'], label='San Diego MA Temperatures')  
plt.plot(gdf['year'], gdf['10 year MA'], label='Global MA Temperatures')  
plt.legend(('local ma', 'global ma'), loc='upper left')
```

Line of Best Fit

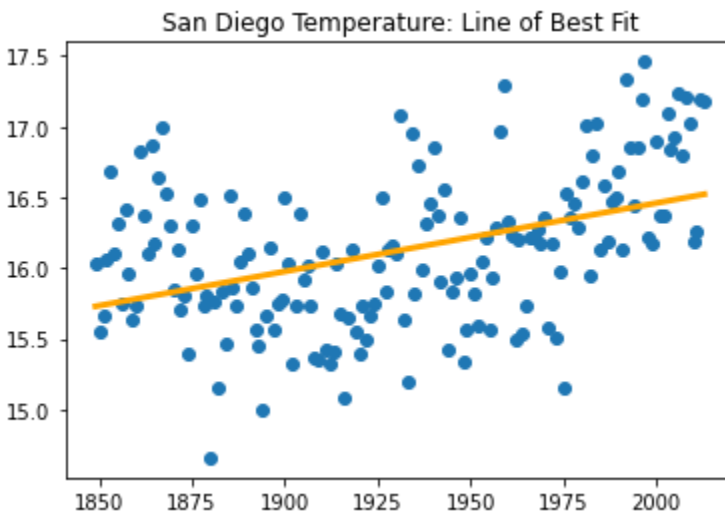
Global Temperature: line of best fit

Slope: 0.0083



San Diego Temperature: line of best fit

Slope: 0.0048



Observations

- Observation 1- general temperature
 - Data is actual temperature, not anomaly values.
 - San Diego is quite a bit warmer on average compared to the global average.
- Observation 2 - Comparing Temperature Changes
 - Though Global Temperatures are consistently lower than San Diego, San Diego Temperatures are rising slower than Global Average Temperatures.
- Observation 3 - the overall trend
 - The world getting hotter over time.
 - The trend of temperature rising is more easily observable in the average global temperatures.
- Observation 4 - the spread of data
 - There is more variance in the local temperature data from San Diego than there is in the Global Temperature data.
 - Sand Diego std: 0.546219
 - Global std: 0.460165
- Observation 5 - correlation
 - There is a small positive linear relationship between the variables.
 - $r = 0.1283$

References

Healthcare IT Leaders, "Why is it always sunny in San Diego?", Grimes, June 2015,
<https://www.healthcareitleaders.com/blog/why-is-it-always-sunny-in-san-diego/>

Statology, "How to calculate a rolling mean in Pandas", Zach, December 2020,
<https://www.statology.org/rolling-mean-pandas/>

Statology, "How to plot the line of best fit in Python", Zach, October 2021,
<https://www.statology.org/line-of-best-fit-python/>

National Centers for Environmental Information, "Anomalies vs. Temperature", August 2022,
<https://www.ncei.noaa.gov/access/monitoring/dyk/anomalies-vs-temperature>