Exploring Weather Trends – Udacity Project 1

-Tarun Lingam

Extract the Data

Write a SQL query to extract the city level data. Extract to .CSV.

1) Query to view details of city of residence – Hyderabad, India:

```
SELECT *
FROM city_list
WHERE country LIKE ‘India’ AND city LIKE ‘Hyd%’;
```

Output was positive, hence city decided as Hyderabad, India.

2) Query to extract city level data:

```
SELECT *
FROM city_data
WHERE country = ‘India’ AND city = ‘Hyderabad’;
```

Output was positive, returning 218 results -
The temperature data for Hyderabad, India from year 1796 to 2013.
Extracted to .CSV file.

Write a SQL query to extract the global data. Extract to .CSV.

1) Query to extract global data:

```
SELECT *
FROM global_data;
```

Output was positive, returning 266 results –
The global temperature data from 1750 to 2015.

2) Temperature data required for the timeframe, 1796 to 2013. Query to extract required global data:

```
SELECT *
FROM global_data
WHERE year >= 1796 AND year <=2013;
```

Output was positive, returning 218 results –
The global temperature data from 1796 to 2013.
Extract to .CSV.

3

Calculate Moving Averages

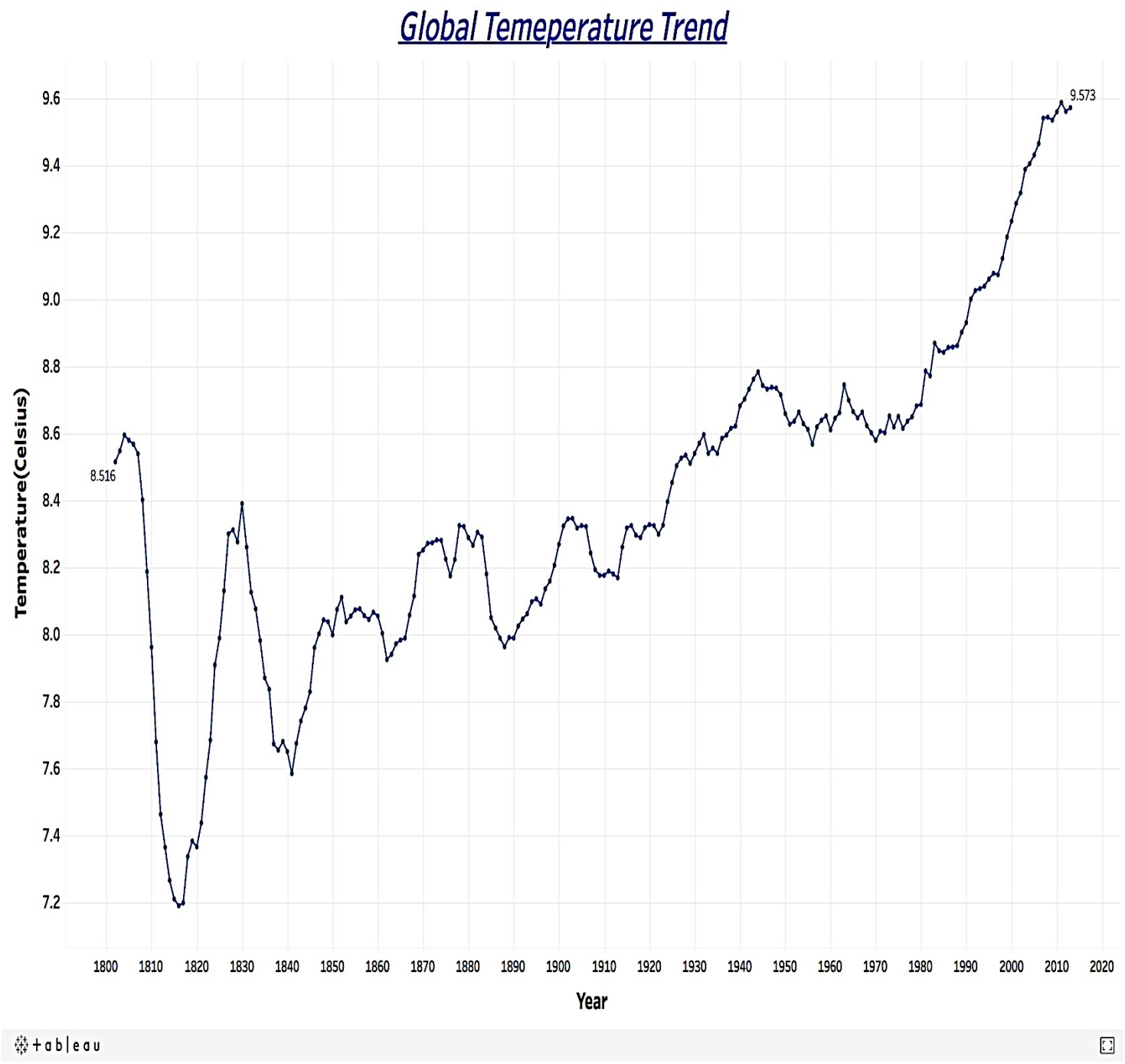
1) Open the .CSV files containing the global data and the city data in Microsoft Excel.
2) Calculate the moving average for 7 years using formula – =AVERAGE(Cell2:Cell8).
3) The values were stored in a new column named 'seven_year_MA' in both the CSV files.

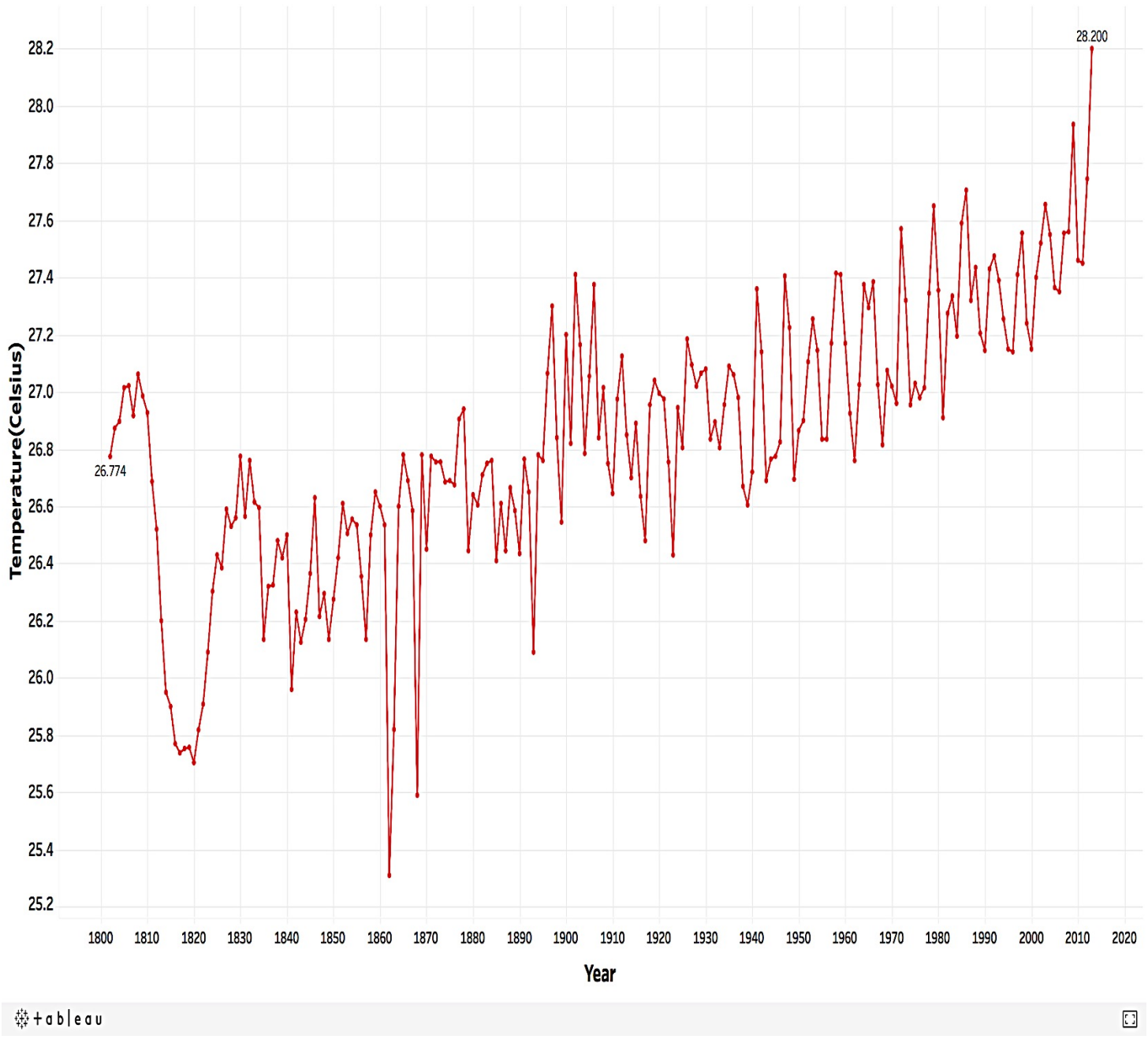
Create visualizations

Tableau was used to create the line charts.

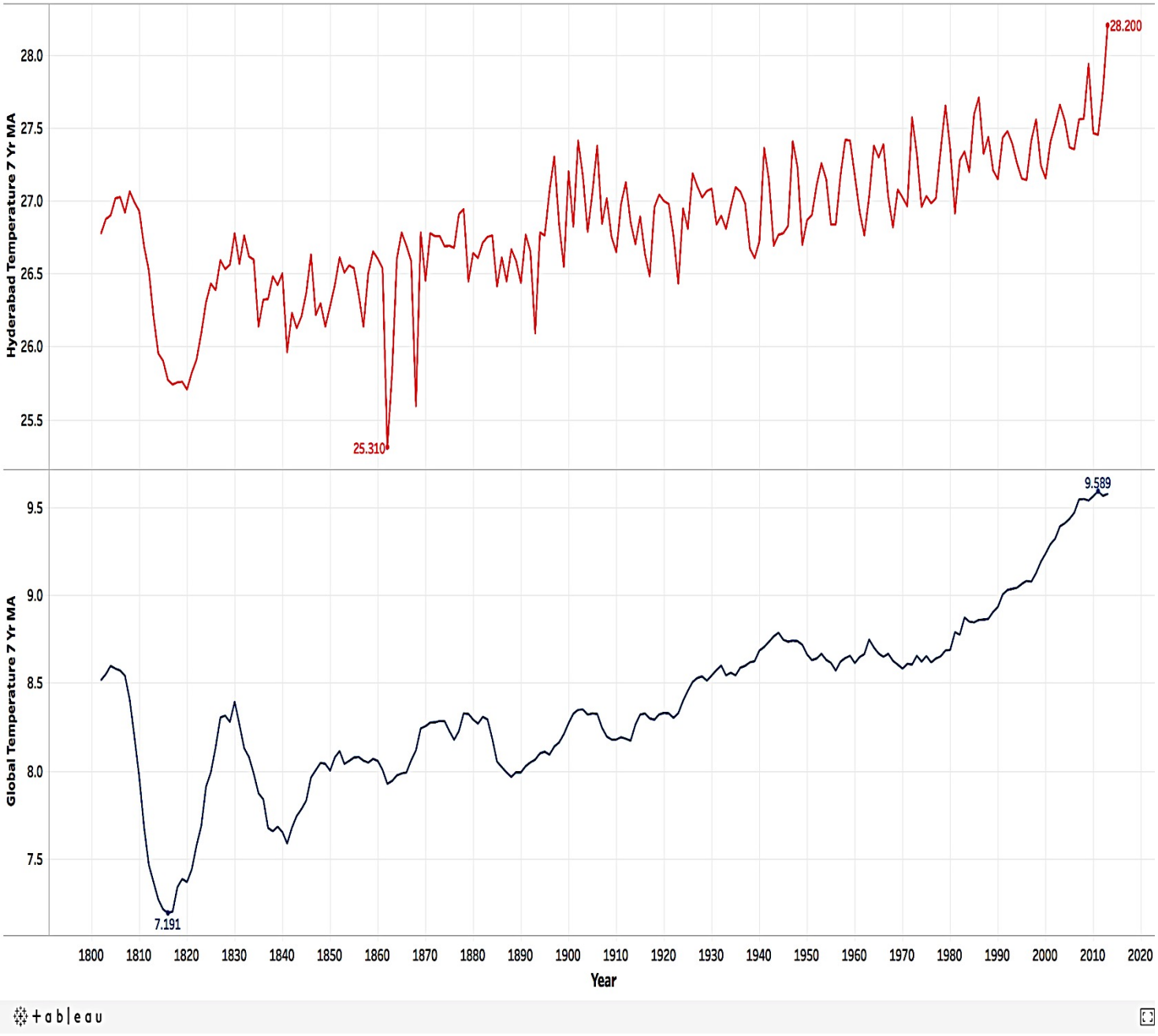
- In Tableau → Data source → Inner join of Global_results.csv and Hyderabad_results.csv is created.
- The null values are filtered and removed.
- In the worksheet, line charts are created:

1) Global Temperature Trend – A line chart plotting global temperature 7 year moving average vs time in years.
2) Hyderabad Temperature Trend – A line chart plotting Hyderabad temperature 7 year moving average vs time in years.
3) Temperature Trend Comparison – A line chart plotting global temperature 7 year moving average and Hyderabad temperature 7 year moving average vs time in years.
4) A dashboard is created with all the above line charts.


Global Temeperature Trend
9.573
8.516
9.6
9.4
9.2
9.0
8.8
8.6
8.4
8.2
8.0
7.8
7.6
7.4
7.2
Temperature(Celsius)
1800
1810
1820
1830
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Year
+ableau

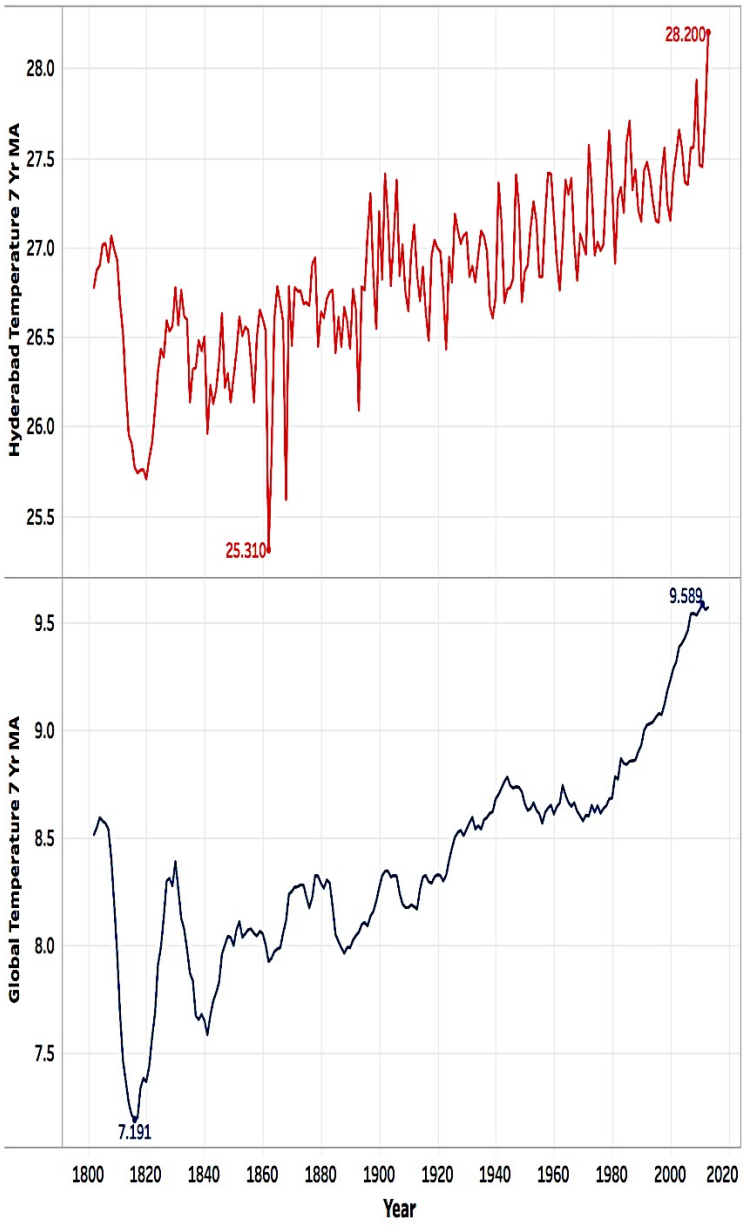

Hyderabad Temperature Trend
28.200
26.774
Temperature(Celsius)
28.2
28.0
27.8
27.6
27.4
27.2
27.0
26.8
26.6
26.4
26.2
26.0
25.8
25.6
25.4
25.2
1800
1810
1820
1830
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Year
+ableau

Temperature Trend Comparison

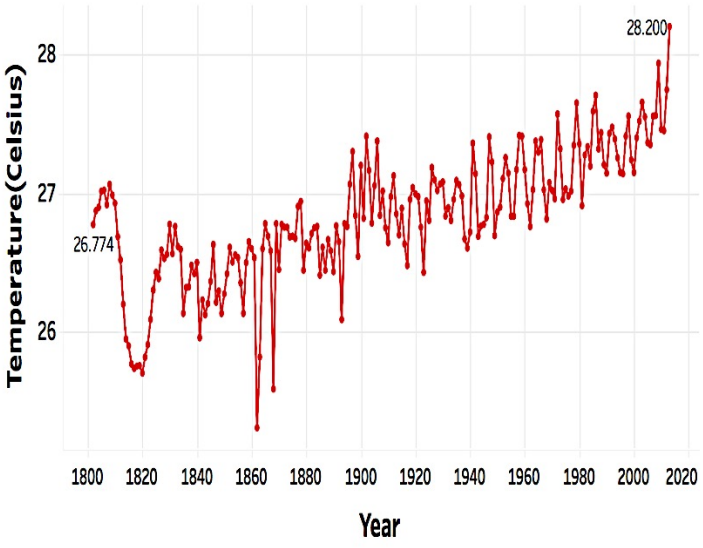


Exploring Weather Trends

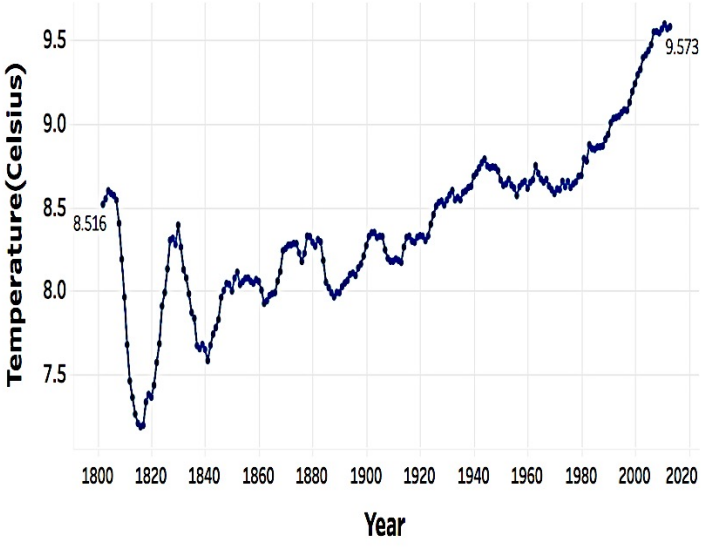
Temperature Trend Comparison



Hyderabad Temperature Trend



Global Temeperature Trend



+ableau

Observations

1. Hyderabad city being closer to the equator of the planet is significantly warmer than the global average temperature. This is seen in the temperature trend comparison chart.
2. The difference in the average temperature per year and the moving averages between Hyderabad city and the global temperatures is consistently between 18^0C and 19^0C.
3. Changes in city temperature in comparison with global temperatures look related. In a year by year comparison of the moving average, the highs and lows seem to be moving in a similar pattern with a few exceptions. This is seen in the temperature trends comparison chart.
4. Hyderabad city seems to have a wider variance of average temperature and moving average compared to the global values. This is due to the presence of outlier values which have a significant effect on the averages.
5. Both Hyderabad city and global temperatures show an upward trend. While highs and lows are clearly visible, this upward trend has been consistent since the mid 1800's and is very significant since the mid 1970's.
6. Global warming is a significant concern all over the world and the Temperature trend charts show this clearly.

Additional Insights

Correlation Coefficient

Correlation coefficient is used to find how strong a relation is between data.

The results vary between -1 and 1.

- -1 signifies strong negative relationship.
- 1 signifies a strong positive relationship.
- 0 signifies no relationship.

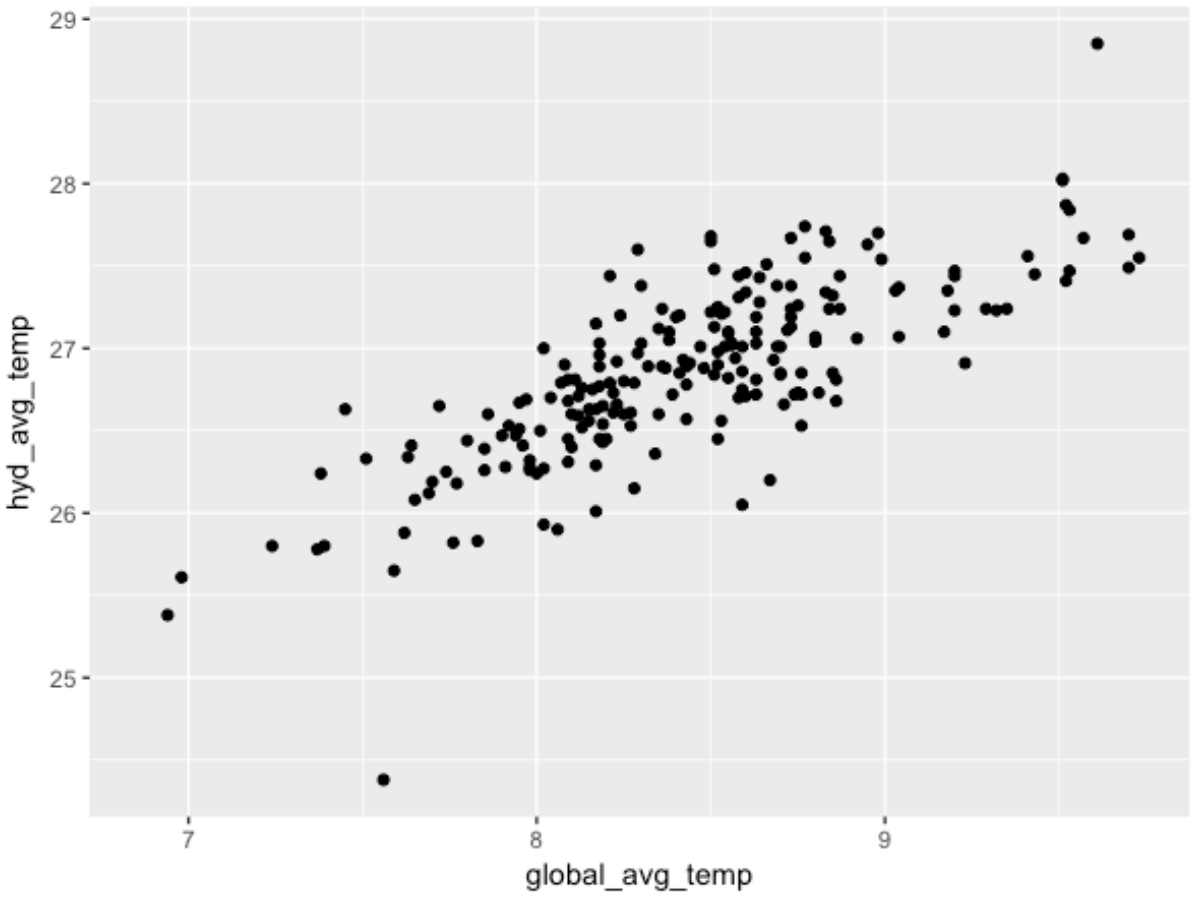
- The correlation coefficient for global and Hyderabad city average temperatures is –

i. 0.782033 using Pearson's method
ii. 0.5888735 using Kendall's method
iii. 0.7747645 using Spearman's method.

- The correlation coefficient for global and Hyderabad city 7-year temperature moving averages is –

i. 0.8368766 using Pearson's method
ii. 0.6507159 using Kendall's method
iii. 0.8401776 using Spearman's method

Therefore, there seems to be definite positive correlation between Hyderabad city and global temperatures.



Calculating Correlation Coefficient

In order to calculate correlation coefficient, the data from the global and Hyderabad city datasets were merged into a new .CSV file and matching was done along Year column. The resulting dataset contained the average temperatures and the 7-year temperature moving averages for Global and Hyderabad city for years 1796 to 2013.

Correlation coefficient was calculated using R. The code is as follows –

```
#Load Package
library(ggplot2)

#Set working directory
setwd('//Users/Magnil/OneDrive/Udacity/Projects/Explore Weather Trends')
getwd()

#Read the dataset and store it into a dataframe
trends <- read.csv('Weather_Trends.csv')
```

```
View(trends)

# Add average temperatures into variables
global_avg_temp <- trends$global_avg_temp
hyd_avg_temp <- trends$hyd_avg_temp
global_MA <- trends$Global_seven_year_MA
hyd_MA <- trends$Hyderabad_seven_year_MA

# Plot to visually view if relation exists between global and local
temperatures
ggplot(trends,aes(global_avg_temp, hyd_avg_temp))+ geom_point()

# Computing correlation coefficient for average temperatures
cor.test(global_avg_temp, hyd_avg_temp) #default is Pearson’s
cor.test(global_avg_temp, hyd_avg_temp, method = "kendall")
cor.test(global_avg_temp, hyd_avg_temp, method = "spearman")

# Computing correlation coefficient for 7 year moving averages
cor.test(global_MA, hyd_MA) #default is Pearson’s
cor.test(global_MA, hyd_MA, method = "kendall")
cor.test(global_MA, hyd_MA, method = "spearman")
```