Project 1. Exploring Weather Trends

Completed by Kirill Ryzhov on 25 December 2017

1. Summary

The project explores global and local ambient temperatures observed over the period of 250+ years. The project further analyses temperature time series and temperature trends to conclude on similarities and differences between those sets.

2. Methodology

The analysis utilises annual average temperatures and simple arithmetic moving averages method on two period terms 10-year and 100-year. The analysis also utilises additional statistics analysis, such as correlation, and basic regression with normalised residuals. Each data is plotted along the entire period of temperature observations for visual analysis on similarities and differences.

3. Data extraction and preparation

3.1. Extracting Global Data:

SELECT *
FROM global_data;
The data is gathered between 1750 and 2015.

3.2. Selecting the list of cities in UK:

SELECT *
FROM city_list
WHERE country LIKE '%King%' OR country LIKE '%UK%'

3.3. Fetching temperature observations filtering by country, city and year:

SELECT *
FROM city_data
WHERE (country LIKE 'United Kingdom') AND (city LIKE 'London') AND (year>=1750);

3.1. Alternative method using merging two tables together (trial tests crashed the online workspace data. The original tables are not available)

SELECT global_data.year, global_data.avg_temp, city_data.year, city_data.city
FROM global_data INNER JOIN city_data USING (year)
WHERE (city LIKE 'London') AND (year>=1750);

3.2. Excel

Combining data from three data-sets and setting 10-year and 100-year moving average.

	A	B	C	D	E	F	G	H	I	J	K	L
1		=AVERAGE(B5:B268)	=AVERAGE(C5:C268)		=SUMIF(D51:D							
2		avg_temp	avg_temp	avg_temp	=B4	=C4	=E2	=F2	=D4	=G2	=H2	=I2&", "&D3
3		Global	United Kingdom	Sri Lanka								
4	year	Global	London	Colombo	10year MA	10year MA	100year MA	100year MA	100year MA	100year MA_norm	100year MA-Norm	100year MA_norm
5	1750	8.72	=VLOOKUP(A5&$C$4&$C$3,city_temp!C2:E71312,3,FALSE)	=VLOOKUP(A5								
6	1751	7.98	=VLOOKUP(A6&$C$4&$C$3,city_temp!C3:E71313,3,FALSE)	=VLOOKUP(A6								
7	1752	5.78	=VLOOKUP(A7&$C$4&$C$3,city_temp!C4:E71314,3,FALSE)	=VLOOKUP(A7								
8	1753	8.39	=VLOOKUP(A8&$C$4&$C$3,city_temp!C5:E71315,3,FALSE)	=VLOOKUP(A8								
9	1754	8.47	=VLOOKUP(A9&$C$4&$C$3,city_temp!C6:E71316,3,FALSE)	=VLOOKUP(A9								
10	1755	8.36	=VLOOKUP(A10&$C$4&$C$3,city_temp!C7:E71317,3,FALSE)	=VLOOKUP(A1								
11	1756	8.85	=VLOOKUP(A11&$C$4&$C$3,city_temp!C8:E71318,3,FALSE)	=VLOOKUP(A1								
12	1757	9.02	=VLOOKUP(A12&$C$4&$C$3,city_temp!C9:E71319,3,FALSE)	=VLOOKUP(A1								
13	1758	6.74	=VLOOKUP(A13&$C$4&$C$3,city_temp!C10:E71320,3,FALSE)	=VLOOKUP(A1								
14	1759	7.99	=VLOOKUP(A14&$C$4&$C$3,city_temp!C11:E71321,3,FALSE)	=VLOOKUP(A1								
15	1760	7.19	=VLOOKUP(A15&$C$4&$C$3,city_temp!C12:E71322,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B5:B15)	=AVERAGE(C5:C						
16	1761	8.77	=VLOOKUP(A16&$C$4&$C$3,city_temp!C13:E71323,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B6:B16)	=AVERAGE(C6:C						
17	1762	8.61	=VLOOKUP(A17&$C$4&$C$3,city_temp!C14:E71324,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B7:B17)	=AVERAGE(C7:C						
18	1763	7.5	=VLOOKUP(A18&$C$4&$C$3,city_temp!C15:E71325,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B8:B18)	=AVERAGE(C8:C						
19	1764	8.4	=VLOOKUP(A19&$C$4&$C$3,city_temp!C16:E71326,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B9:B19)	=AVERAGE(C9:C						
20	1765	8.25	=VLOOKUP(A20&$C$4&$C$3,city_temp!C17:E71327,3,FALSE)	=VLOOKUP(A2	=AVERAGE(B10:B20)	=AVERAGE(C10						
21	1766	8.41	=VLOOKUP(A21&$C$4&$C$3,city_temp!C18:E71328,3,FALSE)	=VLOOKUP(A2	=AVERAGE(B11:B21)	=AVERAGE(C11						
22	1767	8.22	=VLOOKUP(A22&$C$4&$C$3,city_temp!C19:E71329,3,FALSE)	=VLOOKUP(A2	=AVERAGE(B12:B22)	=AVERAGE(C12						
23	1768	6.78	=VLOOKUP(A23&$C$4&$C$3,city_temp!C20:E71330,3,FALSE)	=VLOOKUP(A2	=AVERAGE(B13:B23)	=AVERAGE(C13						
24	1769	7.69	=VLOOKUP(A24&$C$4&$C$3,city_temp!C21:E71331,3,FALSE)	=VLOOKUP(A2	=AVERAGE(B14:B24)	=AVERAGE(C14						
25	1770	7.69	=VLOOKUP(A25&$C$4&$C$3,city_temp!C22:E71332,3,FALSE)	=VLOOKUP(A2	=AVERAGE(B15:B25)	=AVERAGE(C15						
26	1771	7.85	=VLOOKUP(A26&$C$4&$C$3,city_temp!C23:E71333,3,FALSE)	=VLOOKUP(A2	=AVERAGE(B16:B26)	=AVERAGE(C16						
27	1772	8.19	=VLOOKUP(A27&$C$4&$C$3,city_temp!C24:E71334,3,FALSE)	=VLOOKUP(A2	=AVERAGE(B17:B27)	=AVERAGE(C17						
28	1773	8.22	=VLOOKUP(A28&$C$4&$C$3,city_temp!C25:E71335,3,FALSE)	=VLOOKUP(A2	=AVERAGE(B18:B28)	=AVERAGE(C18						
29	1774	8.77	=VLOOKUP(A29&$C$4&$C$3,city_temp!C26:E71336,3,FALSE)	=VLOOKUP(A3	=AVERAGE(B19:B29)	=AVERAGE(C19						
30	1775	9.18	=VLOOKUP(A30&$C$4&$C$3,city_temp!C27:E71337,3,FALSE)	=VLOOKUP(A3	=AVERAGE(B20:B30)	=AVERAGE(C20						
103	1848	7.98	=VLOOKUP(A103&$C$4&$C$3,city_temp!C100:E71410,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B93:B10	=AVERAGE(C93						
104	1849	7.98	=VLOOKUP(A104&$C$4&$C$3,city_temp!C101:E71411,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B94:B10	=AVERAGE(C94						
105	1850	7.9	=VLOOKUP(A105&$C$4&$C$3,city_temp!C102:E71412,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B95:B10	=AVERAGE(C95	=AVERAGE(B4:B					
106	1851	8.18	=VLOOKUP(A106&$C$4&$C$3,city_temp!C103:E71413,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B96:B10	=AVERAGE(C96	=AVERAGE(B5:B	=AVERAGE(C5:C105)				
107	1852	8.1	=VLOOKUP(A107&$C$4&$C$3,city_temp!C104:E71414,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B97:B10	=AVERAGE(C97	=AVERAGE(B6:B	=AVERAGE(C6:C106)				
151	1896	8.21	=VLOOKUP(A151&$C$4&$C$3,city_temp!C148:E71458,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B141:B1!	=AVERAGE(C14	=AVERAGE(B50:	=AVERAGE(C50:C150)				
152	1897	8.29	=VLOOKUP(A152&$C$4&$C$3,city_temp!C149:E71459,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B142:B1!	=AVERAGE(C14	=AVERAGE(B51:	=AVERAGE(C51:C151)	=SUMIF(D51:D151,"<>0")/COUNTIF(D51:D	=G152/$B$1	=H152/$C$1	=I152/$D$1
153	1898	8.18	=VLOOKUP(A153&$C$4&$C$3,city_temp!C150:E71460,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B143:B1!	=AVERAGE(C14	=AVERAGE(B52:	=AVERAGE(C52:C152)	=SUMIF(D52:D152,"<>0")/COUNTIF(D52:D	=G153/$B$1	=H153/$C$1	=I153/$D$1
154	1899	8.4	=VLOOKUP(A154&$C$4&$C$3,city_temp!C151:E71461,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B144:B1!	=AVERAGE(C14	=AVERAGE(B53:	=AVERAGE(C53:C153)	=SUMIF(D53:D153,"<>0")/COUNTIF(D53:D	=G154/$B$1	=H154/$C$1	=I154/$D$1
155	1900	8.5	=VLOOKUP(A155&$C$4&$C$3,city_temp!C152:E71462,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B145:B1!	=AVERAGE(C14	=AVERAGE(B54:	=AVERAGE(C54:C154)	=SUMIF(D54:D154,"<>0")/COUNTIF(D54:D	=G155/$B$1	=H155/$C$1	=I155/$D$1
156	1901	8.54	=VLOOKUP(A156&$C$4&$C$3,city_temp!C153:E71463,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B146:B1!	=AVERAGE(C14	=AVERAGE(B55:	=AVERAGE(C55:C155)	=SUMIF(D55:D155,"<>0")/COUNTIF(D55:D	=G156/$B$1	=H156/$C$1	=I156/$D$1
157	1902	8.3	=VLOOKUP(A157&$C$4&$C$3,city_temp!C154:E71464,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B147:B1!	=AVERAGE(C14	=AVERAGE(B56:	=AVERAGE(C56:C156)	=SUMIF(D56:D156,"<>0")/COUNTIF(D56:D	=G157/$B$1	=H157/$C$1	=I157/$D$1
158	1903	8.22	=VLOOKUP(A158&$C$4&$C$3,city_temp!C155:E71465,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B148:B1!	=AVERAGE(C14	=AVERAGE(B57:	=AVERAGE(C57:C157)	=SUMIF(D57:D157,"<>0")/COUNTIF(D57:D	=G158/$B$1	=H158/$C$1	=I158/$D$1
159	1904	8.09	=VLOOKUP(A159&$C$4&$C$3,city_temp!C156:E71466,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B149:B1!	=AVERAGE(C14	=AVERAGE(B58:	=AVERAGE(C58:C158)	=SUMIF(D58:D158,"<>0")/COUNTIF(D58:D	=G159/$B$1	=H159/$C$1	=I159/$D$1
160	1905	8.23	=VLOOKUP(A160&$C$4&$C$3,city_temp!C157:E71467,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B150:B1(	=AVERAGE(C15	=AVERAGE(B59:	=AVERAGE(C59:C159)	=SUMIF(D59:D159,"<>0")/COUNTIF(D59:D	=G160/$B$1	=H160/$C$1	=I160/$D$1
161	1906	8.38	=VLOOKUP(A161&$C$4&$C$3,city_temp!C158:E71468,3,FALSE)	=VLOOKUP(A1	=AVERAGE(B151:B1(	=AVERAGE(C15	=AVERAGE(B60:	=AVERAGE(C60:C160)	=SUMIF(D60:D160,"<>0")/COUNTIF(D60:D	=G161/$B$1	=H161/$C$1	=I161/$D$1

To support the conclusions, a third city – Colombo, Sri-Lanka – was selected. The city characterises different geographical and industrial region, which subsequently aims to increase representativeness of the analysis and conclusions. Adding the third city to the analysis requires some data preparation, namely:
- normalisation of data based on some average for each set;
- omission of years with no observed temperatures;
- shifting the period of observation.

3.3. Excel Data Analysis: Correlation and Regression

Correlation on all data sets using standard CORREL function, respectively on annual average, 10- and 100- years moving averages. Excel was also used to regress London temperatures on Global temperatures using 10-year moving average data set:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.883
R Square	0.780
Adjusted R Square	0.779
Standard Error	0.170
Observations	253

ANOVA

|	df	SS	MS	F	Significance F
Regression	1	25.74	25.74	887.69	0.00
Residual	251	7.28	0.03		
Total	252	33.02			

4. Discussion:

Graph1 shows that temperature observations for London had more anomalies, outliers and are more volatile. It also indicates a more refined and less volatile structure of the global temperatures observations, especially in the last 50 years period.

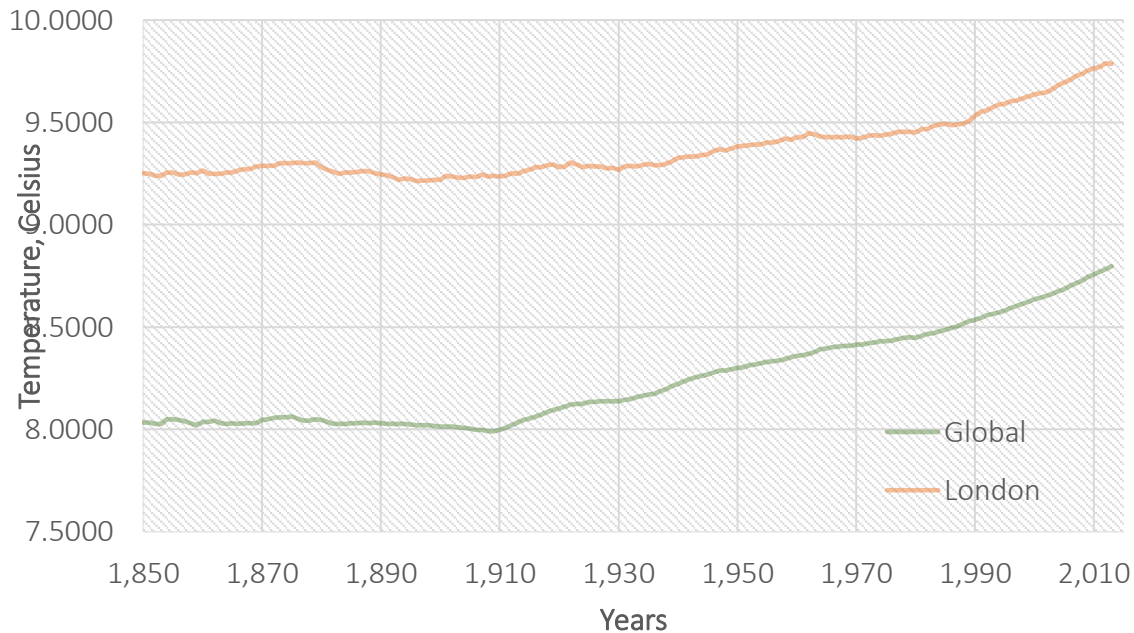Graph1 Average Annual Ambient Temperature Time-Series
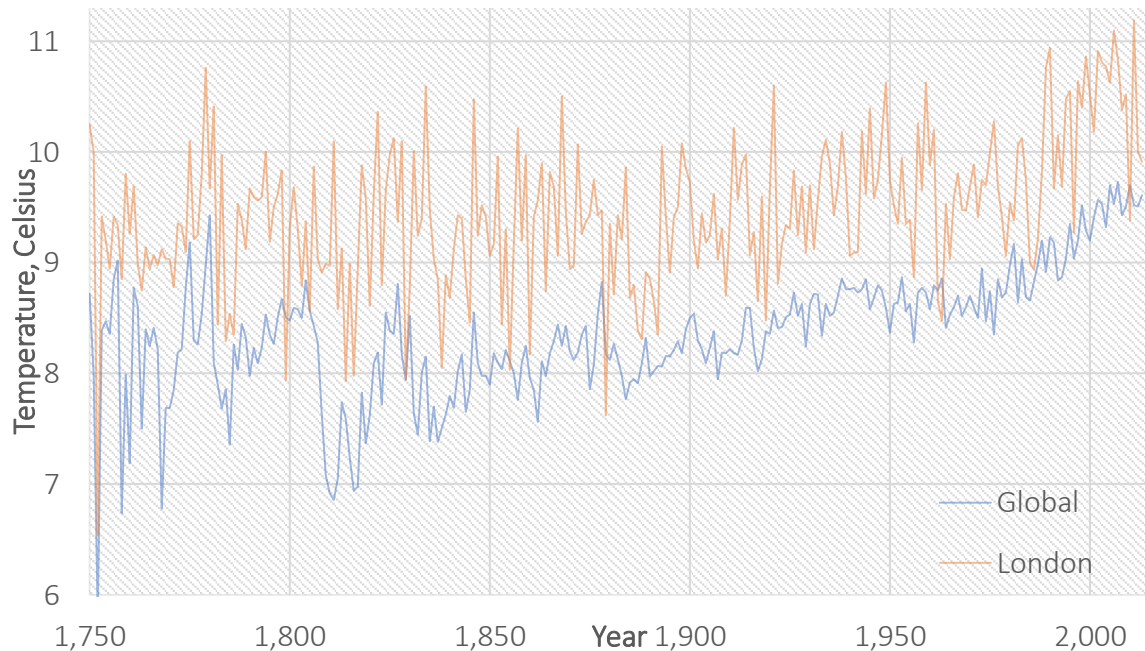
Graph 2 shows 10 year moving averages overlaps original temperature sets. The upward trend is uncertain.

Graph 2. Average Annual Ambient Temperature Time-Series with 10-year Moving Average
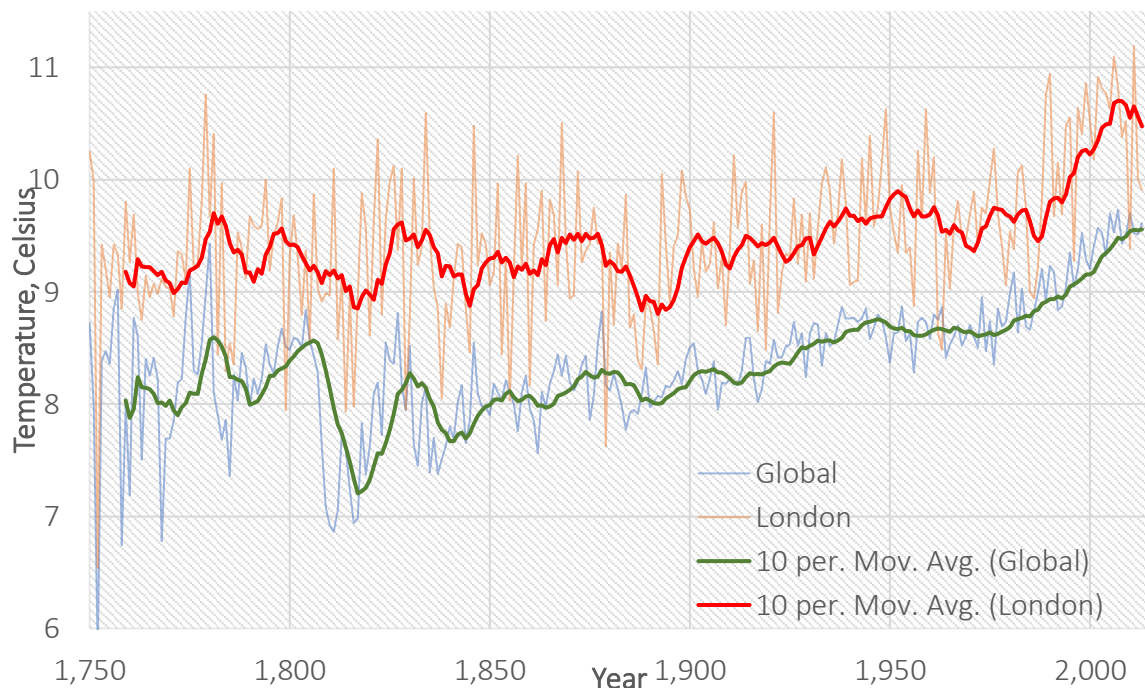
100 year moving average graph shows smoother behaviour. It is obvious that the 100-year average temperature has been on the rise since 1920's, which may be related to historical trends and macrofactors.

Graph 3. Average Annual Ambient Temperature: 100-year Moving Average

To support the conclusions another city was added to the set. A normalised graph is shown below.

Graph 4. Average Annual Ambient Temperatures: normalised 100-year Moving Average
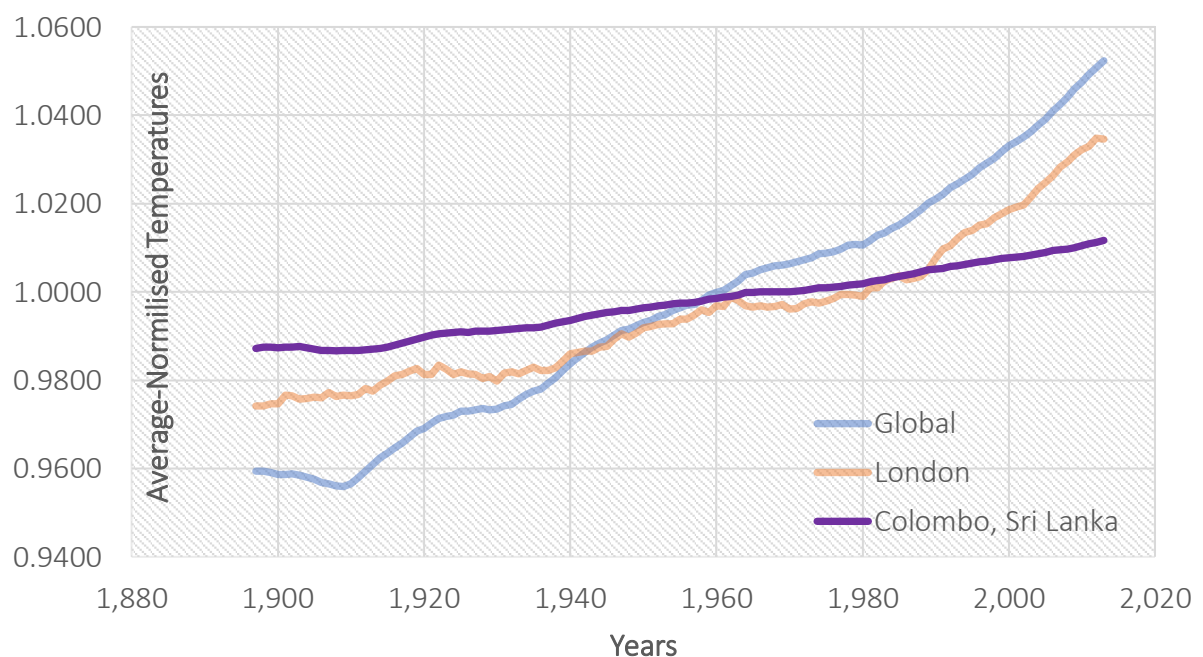
Correlation between annual average temperatures for London and World is 0.68, which is rather significant. It is ever so obvious with 10-year moving average and 100-year moving average data sets:

Table 1. Correlation of global temperature data sets with selected cities sets

Data sets used	London	Colombo
Avg.temperature	0.5631	0.8482
10year MA	0.8832	0.9567
100year MA	0.9755	0.9980

Regression of London 10-year moving average on Global 10-Year moving average results in high level explanation (R=0.78), which is expected due to correlation structure above. The regression F-test and P-value indicates that global temperatures and local London sets are highly interconnected and provide a good fit than the linear model
The regression residuals however are non-normal and show structuring over several periods. This may suggest on another undefined variable in global-local temperature dynamics.

Graph 5. London on Global temperatures regression residuals.
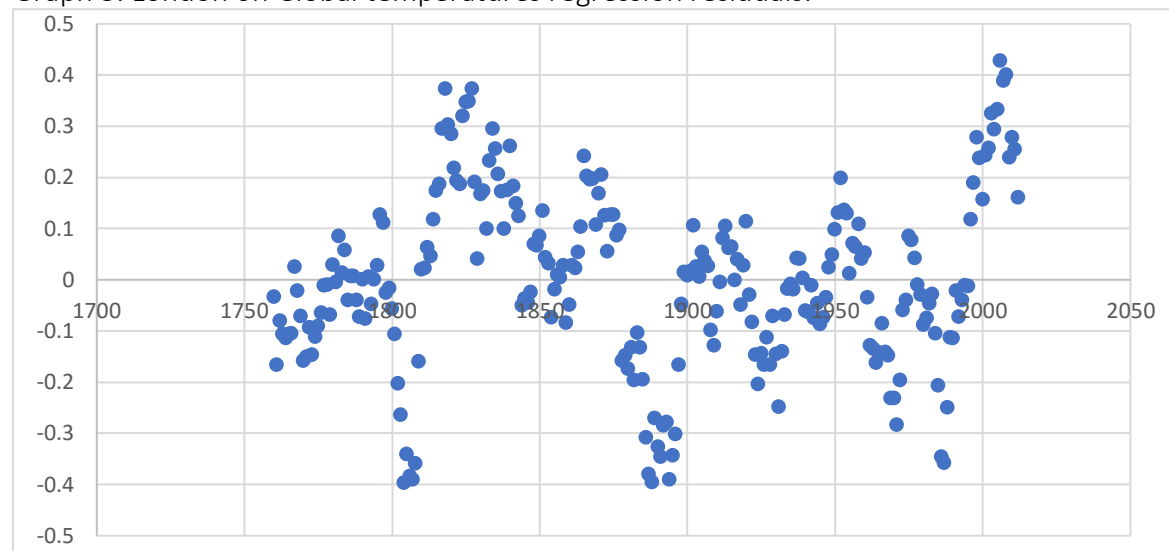
5. Conclusions

1. Selected city has higher temperature, observations are more volatile with more outliers and anomalies present. Global temperature observations are less volatile over the last 50 years period.
2. It is evident that local and global temperatures are on the upward trends (Graph 3) since 1920's.
3. Global temperature trend rises at a higher rate (Graph 4) than both selected local temperature trends.
4. Nonetheless, the correlation structure regressed on temperatures and moving averages suggest that those upward trends are positively and highly correlated, which may suggest a single uniform cause for temperature trend dynamic.
5. A regression suggests high interconnection between two sets. However, it is obvious that the global-local temperature dynamic is lacking an explanatory variable.