Green spaces against heatwaves

Klára Ambrožová

Content

1. Introduction

The Mayor of London has noticed the number of heatwaves that were seen in Europe last summer. Some of them affected the United Kingdom substantially, as the highest air temperature from July 2019 (38.1°C) was only 0.4°C below the overall record for the United Kingdom. The Mayor of London and the London Assembly realized how important is the mitigation of climate change, especially for a city like London where the air temperature is even higher due to Urban Heat Island.

They have heard about the benefits of having green spaces in cities, as they are not only important to the physical and mental health of the citizens, but they also help fight pollution and modify the local climate. However, before they start promoting the construction of more green spaces in London, they would like to quantify the effect of them on the local climate of London boroughs.

Therefore, the task is to find and if possible quantify the effect of green spaces on the air temperature in London.

Fig. 1 People relaxing in the Japanese Garden, Holland Park, London

2. Data

In order to estimate the influence of green spaces on local climate, climatic data were obtained from seven stations in London: Greenwich Park, Hampstead, Hampton W Wks, Heathrow, Kew Gardens, Northolt and Northwood. The mean values for the normal period 1981–2010 are available from this website: https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-climate-averages/gcpv7fnqu. Since the heatwaves are the most common in summer, we are the most interested in summer minimum and maximum temperatures. It is also possible to get latitude, longitude and altitude of the stations on the website. The information will be obtained with the use of web scraping.

Nearest climate station:

Hampstead

Location: 51.56, -0.178 **Altitude:** 137 m above mean sea level **Station type:** Observing Site

Average tables | Average graphs | Location comparison | Average maps

Station: Hampstead, 1981-2010

Month	Maximum temperature (°C)	Minimum temperature (°C)	Days of air frost (days)	Sunshine (hours)	Rainfall (mm)	Days of rainfall ≥1 mm (days)	Monthly mean wind speed at 10 m (knots)
January	7.1	2	8.6	57.5	64.7	12	–
February	7.4	1.7	9.5	76.4	46.6	9.7	–
March	10.5	3.5	4	107.1	48.9	10.2	–
April	13.3	5	1.5	151.6	51.5	9.9	–
May	16.8	8	0.1	192.2	58	9.5	–
June	19.9	10.9	0	191	54.2	9	–
July	22.4	13.2	0	199.9	50.4	8.5	–

Fig. 2 Example of the climatic data to be web-scraped and used for regression analysis. Source: https://www.metoffice.gov.uk/

The data on green spaces come from Foursquare location data. Specifically, we will get the number of green spaces, namely parks and gardens, in a certain radius around each of the climatic station. The idea is to try to extract green space counts for 1km, 5 km and 10 km radiuses to see which of them would have the largest impact on the air temperature of the stations.

3. Methodology

In order to carry out this analysis, we are using mainly the regression analysis.

At first, we preprocessed the data, which mainly included web-scraping them from the www.metoffice.gov.uk website. Then unnecessary variables had to be dropped and minimum and maximum temperature for summer were calculated from monthly values for each station.

The numbers of parks and gardens within 1-km, 5-km and 10-km radius around each station were derived from Foursquare. It was actually necessary to define the category as "garden" or "park" during the search because a lot of shopping centres have the word "park" in their name. The category IDs of "park" and "garden" were gained from the website https://developer.foursquare.com/docs/resources/categories. Furthermore, we have decided to derive a new variable for each radius called "total_gs" which represented the sum of parks and gardens within the specified radius. The reason was the hypothesis that green spaces can influence the air temperature no matter if it is a garden or a park.

Then, correlation coefficients were found for both minimum and maximum temperature to indicate which variables might have a strong relationship with the air temperature. Next, scatter plots for selected variables with minimum and maximum temperature were examined to find whether linear regression is sufficient or it might be better to use polynomial regression.

Finally, simple and multiple linear regressions were applied with minimum or maximum temperature as dependent variables. We specifically examined the slopes of independent variables to find out the impact on the air temperature if the value of the independent variable changed by one (garden, park, degree of latitude etc.). We also checked R^2 to see how much variability of the air temperature can be explained by the relationships.

The main questions we were trying to solve in this study were:

1) What is the impact of the number of green spaces on the air temperature in London?
2) Is the influence larger for minimum or maximum temperature?
3) Is topography a better predictor for the air temperature?
4) How can the analysis be improved?

4. Results

In Tab. 1, we can see the dataset to be used in the analysis. It is obvious that the number of parks within a radius of 10 km (parks_10km) is the same for all the climatic stations. Therefore, it cannot be a good predictor for the air temperature at the stations and it will be removed before the analysis.

Tab.1 Dataset for analysis of the influence of green spaces on the air temperature in London.

	Min_T	Max_T	Latitude	Longitude	Altitude	Stations	parks_1km	parks_5km	parks_10km	gardens_1km	gardens_5km	gardens_10km	total_gs_1km	total_gs_5km	total_gs_10km
0	13.100000	22.500000	51.47700	0.0040	47.0	Greenwich Park	3	50	50	8	46	50	11	96	100
1	12.400000	21.433333	51.56000	-0.1780	137.0	Hampstead	3	50	50	3	50	50	6	100	100
2	13.066667	21.933333	51.41194	-0.3781	12.0	Hampton	1	45	50	2	24	50	3	69	100
3	13.100000	22.566667	51.47900	-0.4490	25.0	Heathrow	1	34	50	0	1	50	1	35	100
4	12.166667	22.566667	51.48200	-0.2900	6.0	Kew Gardens	4	50	50	6	50	50	10	100	100
5	12.333333	22.300000	51.54800	-0.4150	40.0	Northolt	0	50	50	1	5	40	1	55	90
6	12.266667	21.600000	51.61690	-0.4000	113.0	Northwood	0	27	50	2	8	23	2	35	73

Min_T: Minimum summer temperature, Max_T: Maximum summer temperature, parks_1km: the number of parks within 1-km radius of each climatic station, parks_5km: the number of parks within 5-km radius of each climatic station, parks_10km: the number of parks within 10-km radius of each climatic station; gardens_1km, gardens_5km and gardens_10km: same as for parks, but with gardens; total_gs_1km, total_gs_5km, total_gs_20km: same as for parks, but for the sum of parks and gardens.

4.1 Correlation analysis

4.1.1 Strongest relationships

We can see that the strongest relationships for minimum temperature were with latitude with a correlation coefficient of -0.69 (Tab. 2), with the number of gardens and green spaces within a 10-km radius (correlation coefficients: 0.48) and with altitude (correlation coefficient: -0.37). The relationship with a number of green spaces is actually a result of collinearity with the number of gardens, so it will not be discussed further.

Similarly, for maximum temperature, the strongest relationships were found with altitude (correlation coefficient: -0.82), latitude (correlation coefficient: -0.53) and the number of gardens and green spaces within a 10-km radius (correlation coefficients: 0.43). Again, the number of green spaces within a 10-km radius will not be analysed in the rest of the report.

4.2.2 Green spaces influence

There have been found relationships between both minimum and maximum temperature and the number of gardens within a 10-km radius (Tab. 2); however, they are of the opposite sign than expected. In other words, the more gardens there were, the higher the air temperature at the climatic stations. The correlation coefficient should actually have been negative if it should have made the climate more pleasant.

Some negative relationships with any green spaces were found between minimum temperature and the number of gardens within 5 km with a correlation coefficient of -0.12, resp. the number of green spaces within a 5-km radius (correlation coefficient -0.11). The reasons for the lack of a strong negative relationship between green spaces and the air temperature will be discussed in the Discussion section.

Tab. 2 Extract from the correlation matrix showing the correlation coefficients between minimum temperature, maximum temperature and the rest of the variables.

	Min_T	Max_T	Latitude	Longitude	Altitude	parks_1km	parks_5km	gardens_1km
Min_T	1.00	0.32	-0.69	0.18	-0.37	-0.06	-0.05	0.01
Max_T	0.32	1.00	-0.53	0.03	-0.82	0.22	0.23	0.28

	gardens_5km	gardens_10km	total_gs_1km	total_gs_5km	total_gs_10km
Min_T	-0.12	0.48	-0.02	-0.11	0.48
Max_T	-0.05	0.43	0.27	0.03	0.43

Explanation of variables the same as in Tab. 1.

4.2. Scatter plots: Types of relationships

In this section, the scatter plots of several relationships were examined in order to distinguish whether linear or polynomial regression should be used (Fig. 3 and Fig. 4).

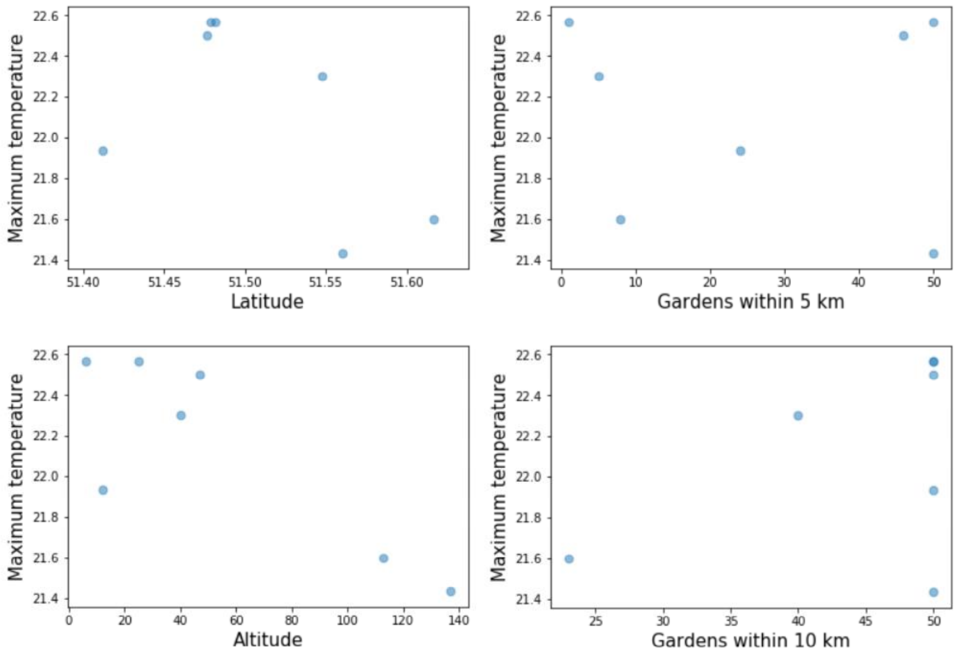

Fig. 3 The relationship between latitude (°), altitude (m a. s. l.), number of gardens within 5 km and number of gardens within 10 km with maximum temperature (°C)

Taking into account that the correlation coefficients were generally quite low, it is not surprising that it is very difficult to see any relationship in for either minimum temperature (Fig. 3) or maximum temperature (Fig. 4). Often we can see different temperatures for the same number of gardens.

Therefore, we only attempted linear regression, as polynomial relationships would not make much sense from the point of physics.

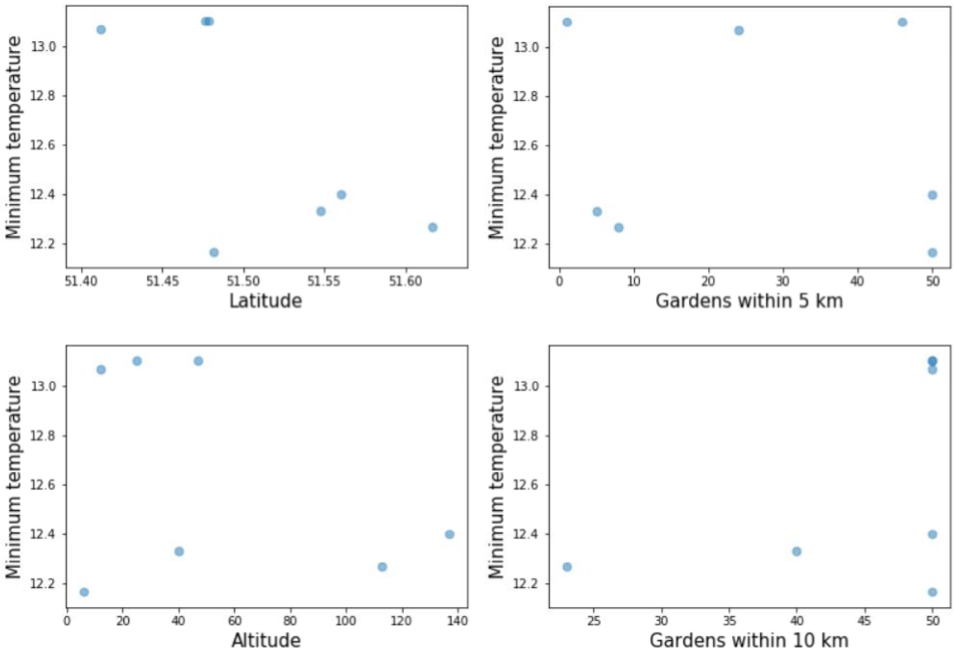

Fig. 4 The relationship between latitude (°), altitude (m a. s. l.), number of gardens within 5 km and number of gardens within 10 km with minimum temperature (°C)

4.3 Simple linear regression

4.3.1 Maximum temperature

For maximum temperature, the number of gardens within a 5-km radius (Tab. 3) only explained about 1.5% of the variance (see R^2). It was slightly better for the number of gardens within a 10-km radius with the explained variance of 23%. From the slope, it was clear that for every 10 gardens within the radius of 10 km maximum temperature should rise by 0.2°C.

The topographic variables were much better predictors, for instance, latitude could explain about 47% of the variance and maximum temperature would actually decrease by -4.4°C for every degree of latitude.

Tab. 3 The values of intercept, slope and R^2 for linear regression between maximum temperature and latitude (°) altitude (m a. s. l.), number of gardens within 5 and 10 km radiuses

Predictor	Intercept	Slope	R2
Latitude	237.7188	-4.3697	0.4743
Altitude	12.8061	-0.0032	0.1404
Gardens within 5 km	12.6967	-0.0024	0.0153
Gardens within 10 km	11.7227	0.0204	0.2346

4.3.2 Minimum temperature

Despite the higher correlation coefficient for minimum than maximum temperature (Tab. 2), the explained variance for linear regression between the number of gardens within a 5-km radius and minimum temperature was lower than 1 % (Tab. 4). Similarly, the explained variance for gardens within a 10-km radius and minimum temperature was 18%, which is lower than for maximum temperature. The reason might be that either quadratic (gardens within 5 km) or exponential relationship (gardens within 10 km) might actually fit minimum temperature better. But again, if the minimum temperature was to increase exponentially in case the number of gardens within 10 km increased, we would soon reach values impossible for the air temperature on Earth.

For minimum temperature, altitude was a great predictor with an explained variance of 68 %. From the slope, it could have been assumed that for every 100 m in altitude the minimum temperature would decrease by 0.8°C (because the slope is for 1m). Latitude explained 29 % of the variance.

Tab. 4 The values of intercept, slope and R^2 for linear regression between minimum temperature and latitude (°) altitude (m a. s. l.), number of gardens within 5 and 10 km radiuses

Predictor	Intercept	Slope	R2
Latitude	214.1716	-3.7282	0.2860
Altitude	22.5462	-0.0077	0.6795
Gardens within 5 km	22.1588	-0.0012	0.0029
Gardens within 10 km	21.2422	0.0198	0.1841

4.4 Multiple linear regression

It was apparent that when all four variables were used as predictors (Tab. 5), explained variance increased both for minimum and maximum temperature. Actually, in both cases, it was possible to explain more than ¾ of variance.

Tab. 5 The values of intercept, slopes and R^2 for multiple linear regression between maximum and minimum temperature with latitude (°) altitude (m a. s. l.), number of gardens within 5 and 10 km radiuses

Temperature	Intercept	Slope_lat	Slope_alt	Slope_g5km	Slope_g10km	R2
Maximum temperature	-308.8022	6.4159	-0.0122	0.0001	0.0247	0.8171
Minimum temperature	488.8560	-9.2486	0.0073	-0.0105	0.0013	0.7641

Slope_lat: slope for latitude, Slope_alt: slope for altitude, Slope_g5km: slope for the number of gardens within 5 km, Slope_g10km: slope for the number of gardens within 10 km

The question was whether this was just an effect of combining the topographic variables or if the number of gardens influenced the size of R^2. As can be seen from Tab. 6, if only altitude and latitude were used as predictors, the explained variance was not as high as when the number of gardens within 5 and 10-km radiuses were included.

Tab. 6 The values of intercept, slope and R^2 for multiple linear regression between maximum and minimum temperature with latitude (°) and altitude (m a. s. l.)

Temperature	Intercept	Slope_lat	Slope_alt	R2
Maximum temperature	-94.7177	2.2791	-0.0101	0.7189
Minimum temperature	359.3077	-6.7343	0.0040	0.5552

5. Discussion and Conclusions

This work aimed to show what was the influence of green spaces on local climate in London. It is a well-known fact that green areas are supposed to help lower the air temperature in cities, for instance, because of their ability to withhold water and later cool by the process of evapotranspiration.

Surprisingly, in this study, we have found the strongest relationship for both maximum and minimum temperature to be positive, i.e. the more green spaces, the warmer climate. One of the reasons for that might be that during extreme heatwaves, after a few days the trees stop evapotranspiration and are actually one of the warmest areas in the whole city.

We have examined in more detail one of the strongest negative relationships for green spaces between minimum summer temperature and the number of gardens within 5 km area. The relationship was very weak and actually insignificant, yet it was found that when combined with other predictor variables it can improve prediction of minimum temperature.

It was found that latitude and altitude were better predictors for both minimum and maximum summer temperature. Yet, the strength of the relationships (based on R^2) had increased when both topographic variables and the number of green spaces were used in multiple linear regression. In the case of multiple linear regression with both topographic and green spaces variables, slightly more variance was explained for maximum than minimum temperature. The reason for that could be in the fact that minimum temperature during the day usually occurs around sunrise and Urban Heat Island is the most effective during the late afternoon and early night. Therefore, it is more likely to affect the maximum temperature.

Considering the radius, the largest impact had the number of gardens within a 10-km radius. That is very surprising because the closest area around the station should affect the local climate the most. The reasons for that are not yet quite understood.

Finally, the question was why in this case the well-known relationship “more green spaces – colder climate” did not work. The answer might be in the type of data we used. In other words, having the area of green spaces instead of just their number could affect the results of the analysis greatly. Furthermore, some of the climatic stations lie within a large green space themselves (e.g. Kew Gardens); it is, therefore, questionable whether this green space, which only counts as one in the predictor variable, does not influence the temperature much more than it was possible to find out in this study.