Open Air Walk

A healthy route visualisation of air pollution in Tower Hamlets

BENVGSC4: Spatial Data Capture, Storage and Analysis
Centre for Advanced Spatial Analysis, UCL
May 23 2018
Word Count: 4,901

Supratim Banik
Liam Bolton
Anna Kozłowska
Thomas Weston

Table of Contents

1. Project Context and Aims

1.1. Project Context

According to the World Health Organisation, 9 out of 10 people worldwide breathe polluted air, with around 4.2 million people dying every year from diseases related to ambient air pollution (WHO, 2018). In London, air pollution causes the early death of more than 9,000 Londoners every year, and costs London's economy around £3.7 billion (London Assembly, 2018). Air pollution, the release of harmful gases and particulates into the atmosphere, is detrimental to health, and is therefore a big problem, especially in large, densely-populated cities such as London. The significance of air pollution has made its reduction a major focus of clean-air initiatives within London. These initiatives have been informed by the data collected by the London Atmospheric Emissions Inventory (LAEI) 2013, and have centred around the monitoring of four key pollutants – nitrogen dioxide (NO_2), nitrogen oxides (NO_x), particulate matter with aerodynamic diameter less than 10µm (PM_{10}), and particulate matter with aerodynamic diameter less than 2.5µm ($PM_{2.5}$) – as well as a variety of other pollutant types (Greater London Authority, 2016a).

Cohen et al's 2017 review of existing literature has confirmed the long-known fact that inhaling polluted air negatively impacts upon health (Cohen, A.J., et al., 2017), affecting both the heart and lungs, with the impacts ranging from short term, mild symptoms, to long term, fatal conditions. Air pollution is primarily caused by energy production, and in London, one of the main causes of air pollution is transportation. Policies to reduce emissions in London have been working, as shown by falling levels of all air pollutants. However, there are still large areas of London that exceed the annual mean EU Limit Values, most noticeably for NO_2 levels (Aether, 2017).

The borough of Tower Hamlets was selected for this project for several reasons. In addition to being one of the five boroughs with the highest number of people living in London's worst air quality areas, it is also one of the five boroughs that have the highest proportion of the most deprived population living in these areas of poor air quality (Aether, 2017). Furthermore, it is the borough with the second highest number of people (15%, or 10,917 people in 2017), living in areas that exceed the EU air pollution limit values (ibid., 2017). Containing historically lower-income areas of London, as well as the global financial hub of Canary Wharf, the borough has the second highest unemployment rate in London, as well as headquarters of many major banks, housing residents with a wide range of incomes. Tower Hamlets therefore has a number of features that make it an interesting case study of a London borough, being reasonably representative in the high degree of its diversity.

1.2. Project Aims

This project aims to provide a user-friendly way to visualise healthy walking routes in the London borough of Tower Hamlets. Based on the 2013 LAEI air pollution data, the website is intended to act as a tool to generate insight for users of all backgrounds, without requiring any previous knowledge of air pollution, spatial data, or even Tower Hamlets itself. These walking routes will be defined as healthy within the context of air pollution – the algorithm chosen will calculate paths of least mean pollution. This visualisation provides a tool for local citizens and other users, that will enable the location and selection of the healthiest outdoor routes nearby, allowing the healthy outdoor use of the borough itself. Presented as an interactive visualisation, the tool is intended to be exploratory, engaging, and user-friendly. Different lengths were chosen as people often have different amounts of time to spend outdoors – a short route may be useful to enjoy after a day at work, while a long route may be appropriate at the weekend when more time is available.

1.3. Literature Review and Existing Projects

Platforms such as Google Maps become ubiquitous for many reasons; perhaps one of the main reasons being their ease of use. They provide an incredible amount of information in an easily-digestible format, allowing any user, irrespective of background or prior knowledge, to find exactly what they want. While this project does not aim for Google Maps levels of ubiquity, it does aim to provide an easy way for any user to explore advanced spatial analysis, by creating a simple and user-friendly interface.

This project provides a tool to bridge the gap between existing map-based websites such as Google Maps, and a traditionally presented research paper. The advanced spatial analysis behind the website is probably of little interest to the average user, and as such, ease of use, rather than dissemination of methodology, was chosen as the main focus of the project. As far as similar tools are concerned, 'Londonair', is provided by the Environmental Research Group of King's College London. It is the website of the London Air Quality Network (LAQN) and presents information on pollution levels around London and the south east part of England. While an extremely high quality and informative resource, this website provides no insight for individual users. Londonair visualises air pollution but provides no suggestions on how to avoid pollution in specific neighbourhoods. This project goes beyond the Londonair project by contextualising the air pollution data, showing users not only pollution data in its raw form, but also and more importantly, routes to avoid areas of high pollution and thereby take in the lowest mean amount of air pollution over various route lengths.

Visualisations that combine pre-calculated spatial data with some user-input capabilities, served as the inspiration for this project. A great deal of work has been done within the field of commuting to work within the US, and projects such as 'Average Commute Times' and 'How Americans Get to Work' combine high quality spatial analysis, with an easy to use interface (Auto Accessorise Garage, 2018, Yau, N., 2007). Individual projects such as Au's work on visualising rail network capacity within Germany, and standalone websites such as Google Maps were drawn upon, with the route visualisation for the project based closely upon the simple aesthetic of Google Maps (Au, C., 2017).

2. Design Rationale: Development and Execution

2.1. Data Collection, Handling, Cleaning, and Management

All data was sourced from the London Datastore. The dataset used for this project was the London Atmospheric Emissions Inventory (LAEI), which contains the most recent 2013 pollution surface data for NO_2, NO_x, PM_{10} and $PM_{2.5}$, covering all 33 London Boroughs, and out as far as the M25 motorway (Greater London Authority, 2016b). Upon downloading, the data files were found to be pre-cleaned and processed, so no further cleaning was required. As Tower Hamlets was selected as the borough of focus, the only management of the dataset required before analysis was the extraction of the four key pollutants for this borough.

2.2. Data Analysis and Major Findings

Table 1 summarises the 2013 minimum, maximum, and mean values for each pollutant within the borough, as well as the EU limit values sourced from UK Air Information Resource (UK Air, 2018). We can see from table 1, that the 2013 mean exceeds the EU limit values for both NO_2 and NO_x. While the 2013 means for PM_{10} and $PM_{2.5}$ do not exceed the EU limit values, we can see that the maximum recorded value for each of the four pollutants does exceed the EU limit values. While the maximum 2013 values for PM_{10} and $PM_{2.5}$ exceeded the EU limit values by 26 and 6 μgm^{-3} respectively, the maximum 2013 values for NO_2 and NO_x exceeded the EU limit values by 152 and 754 μgm^{-3}. This huge exceedance, more than 3 times above the limit for NO_2, and more than 25 times above the limit for NO_x, would normally point to an anomalous reading. However, given the accuracy of the sensor network used for measuring the values themselves, this is highly unlikely.

Table 1. 2013 minimum, maximum and mean values, and EU limit values (UK Air, 2018).

	2013 Minimum (μgm^{-3})	2013 Maximum (μgm^{-3})	2013 Mean (μgm^{-3})	EU Limit Values: Annual Mean (μgm^{-3})
NO_2	34	192	42	40
NO_x	51	784	74	30
PM_{10}	25	66	27	40
$PM_{2.5}$	16	31	17	25

Table 2 provides summary statistics for all 45 routes, and provides the minimum and maximum values, as well as the range, mean, and standard deviation of values. These routes are all based on the combined pollutant values. While the minimum pollution values for all route lengths have a relatively small range of 0 to 24, the maximum values range from 13 to 574. This wide variety of values results in some ranges that are 0, and some that are over 570. There is far less variety of mean values, with the lowest mean of 7 and the highest mean of 45.08. Standard deviation ranges from 0 to 50.85.

Within the 45 routes, there are some interesting outliers. There are three routes with no variation between the maximum and minimum values – routes 10, 19 and 23. It is interesting to note that both mean and standard deviation increase gradually over the 1 km to 6 km route lengths, and then at the 7 km, 8 km and 9km route lengths, their values increase substantially.

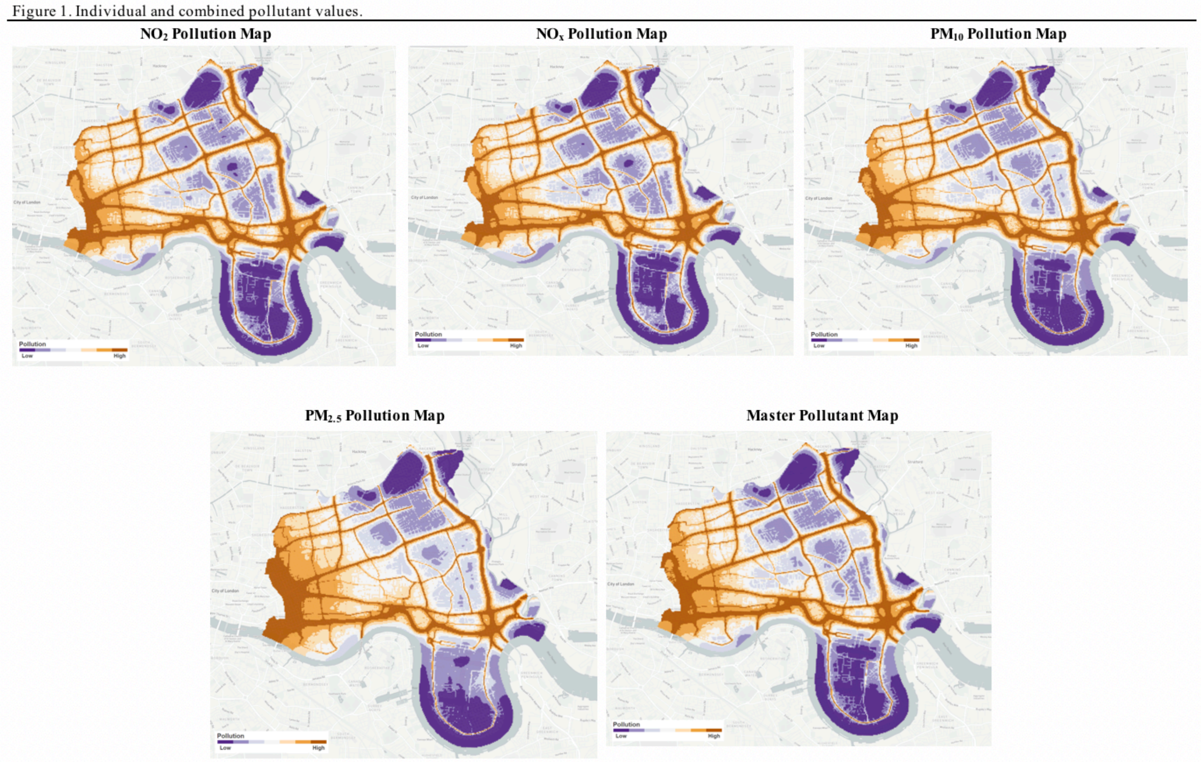
This is, however, not surprising, as shorter routes by their nature have fewer readings, decreasing the chance of extremely high values which inflate both the mean and standard deviation. It is interesting to note that there is no consistent linear relationship between length of route, and minimum, maximum, or standard deviation. Apart from two outliers (routes 8 and 16), the majority of routes under 6 km have broadly similar minimum and maximum values. However, over 7 km, there is a significant increase, and only 3 routes (35, 39 and 40) have a range under 100.

Table 2. Summary statistics for all 45 healthy walking routes.

Route Number	Minimum (µgm-3)	Maximum (µgm-3)	Range (µgm-3)	Mean (µgm-3)	Standard Deviation
1 km					
1	16	85	69	23.51	7.22
2	16	27	11	18.32	2.84
3	11	23	12	15.03	3.00
4	7	17	10	9.25	2.99
5	7	17	10	7.68	1.93
2 km					
6	16	21	5	17.28	1.84
7	16	17	1	16.27	0.45
8	4	105	101	13.45	12.62
9	5	21	16	13.03	6.16
10	13	13	0	13.00	0.00
3 km					
11	5	24	19	14.43	7.77
12	11	20	9	14.03	2.32
13	4	17	13	12.60	5.08
14	10	20	10	12.09	2.50
15	10	13	3	11.58	1.12
4 km					
16	1	347	346	22.14	34.43
17	14	28	14	19.35	4.62
18	4	26	22	17.46	7.94
19	16	16	0	16.00	0.00
20	4	87	83	10.91	7.12
5 km					
21	15	26	11	21.12	3.41
22	5	39	34	17.05	11.75
23	17	17	0	17.00	0.00
24	11	25	14	15.77	4.58
25	11	13	2	12.30	0.80
6 km					
26	24	27	3	25.20	0.98
27	9	70	61	23.35	11.26
28	14	28	14	19.31	4.44
29	15	27	12	17.56	3.97
30	4	13	9	7.00	4.24
7 km					
31	11	453	442	38.73	45.24
32	10	311	301	37.33	35.48
33	1	347	346	32.96	50.85
34	15	186	171	28.78	22.39
35	15	34	19	22.33	8.08
8 km					
36	9	351	342	45.08	36.53
37	9	453	444	43.68	44.27
38	4	186	182	28.17	22.84
39	9	85	76	28.16	22.55
40	13	42	29	18.11	5.23
9 km					
41	15	204	189	39.94	24.62
42	1	574	573	31.44	45.84
43	4	574	570	28.84	37.03
44	1	331	330	28.04	34.52
45	0	574	574	24.49	37.26

Figure 1 shows the values for each individual pollutant (NO_2, NO_x, PM_{10} and $PM_{2.5}$) as well as the combined pollutant values, in map form.

Figure 1. Individual and combined pollutant values.



There are several commonalities across all five maps. Firstly, we can clearly see the major roads within the borough. These are highlighted across all pollutants and illustrate the large impact that all five pollutants have on air quality. Secondly, we can see that there are two areas with the lowest levels of pollutants. These can be found in the north east and south east corners of the borough. This is where the two largest parks are located – Victoria Park in the north east corner, and Mudchute Park and Farm in the south east corner. As would be expected, areas with green, open space do indeed to have lower levels of pollution, as not only do they contain increased levels of pollution-absorbing vegetation, but also fewer sources of pollution (this is particularly the case for Mudchute Park and Farm, which being situated in the loop of the river has far fewer significant roads and hence much lower exposure to pollution creating vehicular traffic).

While all maps illustrate broadly the same patterns, there are also some subtle differences between pollutants. Firstly, we can clearly see a large area of high pollution in the $PM_{2.5}$ map, that extends beyond the individual road paths. Secondly, we can see slightly different pollution patterns across all four individual pollutants, within the Isle of Dogs peninsula. While the NO_2 and NO_x patterns look roughly similar, there is a larger area of low pollution in the PM_{10} map and a smaller, more southern concentration of low pollution in the $PM_{2.5}$ map.

2.3. Methodology

The data for each of the four key pollutants listed above (NO_2, NO_x, PM_{10} and $PM_{2.5}$) were first normalised. Each layer of the pollutant values was converted to a uniform scale, by normalising and transforming the values to a range between 0 and 1. These layers were then added using raster algebra, to form the combined layer. This layer was again normalised and finally scaled up to increase the detail of variation of the final values, therefore increasing the precision of the final routes to be created. This normalisation allowed the combination of each separate pollutant, into a combined air pollution dataset; a single numerical reading for each data point represents all four pollutants.

The first normalisation was performed to remove the scaling differences. The second normalisation was performed to normalise the row-column results, providing normalised values in the range of 0 to 1, making relative comparisons easier.

The analysis of the data focused on the generation of 45 healthy, low pollution routes – five routes for lengths ranging from 1 to 9 kilometres (in gradations of 1km). The five routes for each kilometre length with the lowest mean concentration level for all four pollutants (NO_2, NO_x, PM_{10} and $PM_{2.5}$) were selected, and follow the existing streets of the borough. The combined pollutant concentration layer was therefore used as the combined layer, over which all 45 routes were overlaid. The existing road network was used as the base for estimating the lowest routes. Random points were generated all across the borough, which were incidents on the road network to be used as origin and destination points for those routes. Finally, using these origin and destination points, multiple routes of various lengths passing through streets with the lowest combined level of pollutants were created to connect these origins and destinations.

Once the routes were created, the delineation was used to extract the exposure per pixel of the different pollutant levels along the entire route for each length, by a raster algebra operation of multiplying the routes with the original combined cost surface. The mean value for each of these routes was then used to select the five least polluted routes for each category of length ranging from 1 to 9 km, forming a total of 45 routes. The sum of the values for each of the routes returned the total exposure of the routes to the different pollutants. From these routes, those with the lowest amount of combined pollutant levels were selected for the final routes.

The shortest path algorithm from the scikit-image and GDAL python libraries were selected for this analysis, as they both provide functions for dealing with raster formatted geospatial data. To prevent the algorithm from creating routes that might deviate from the road network into nearby low-pollution areas, the cost surface was created such that the intermediate areas in between the streets with the lowest NO_2 levels were restricted, while the roads indicate the exact NO_2 levels along which the paths were routed. Therefore, the intermediate areas between the streets in the combined cost surface layer are set to arbitrarily high values, while the street concentration levels remain at their actual values. This procedure was estimated using raster algebra, by multiplying a layer with streets having a value of 1 and the rest of the areas with an arbitrary value with the combined pollution surface layer to transfer the values of the combined levels from the pollution layer to the resulting layer, such that the streets have the actual NO_2 values, while the intermediate layers have arbitrarily high values. This treatment ensured that the newly estimated lowest pollutant routes were restricted to the existing streets, rather than crossing into intermediate areas, which in reality, may be inaccessible. After transforming the raster values into pixel values of an image, this layer was used as the final cost surface in the shortest route algorithm, along with the origin and destination points for the routes estimation.

2.4. Data Visualisation and User Integration

This visualisation tool is intended to be interactive, exploratory, engaging, and user-friendly. This intention has been realised by using a clear and minimalistic layout, with extraneous information and options avoided. The interactive data visualisation is responsive and can be viewed on a variety of different devices – computer, mobile, and tablet.

On the Home page, users are easily able to toggle between the pollution data and the healthy routes. The ability to zoom, paired with a search bar that will automatically zoom into areas searched for, enables users to view specific areas of interest in more detail. By clicking on the three horizontal bars in the top left-hand corner of the page, users are also able to view information about website usage, as well as having the option to navigate to the About and Analysis pages. In order to ensure the map could be viewed by all users, including those with visual impairments such as colour-blindness, the colours were chosen according to a 'colorblind safe' palette found within the Colorbrewer2 tool. To maximise visual acuity, contrasting dark and light colours were used, in order to most clearly present information. The colours chosen are qualitative, meaning that there is no ordinal or cardinal ordering, thus making perception easier. Figures 2.1 and 2.2 illustrate the look and feel of the home page.

Figure 2.1. Home page pollution view.

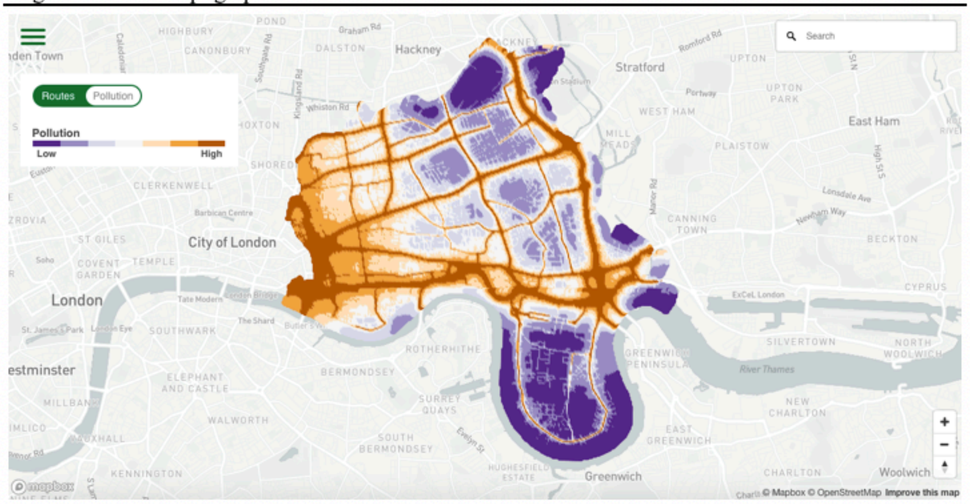
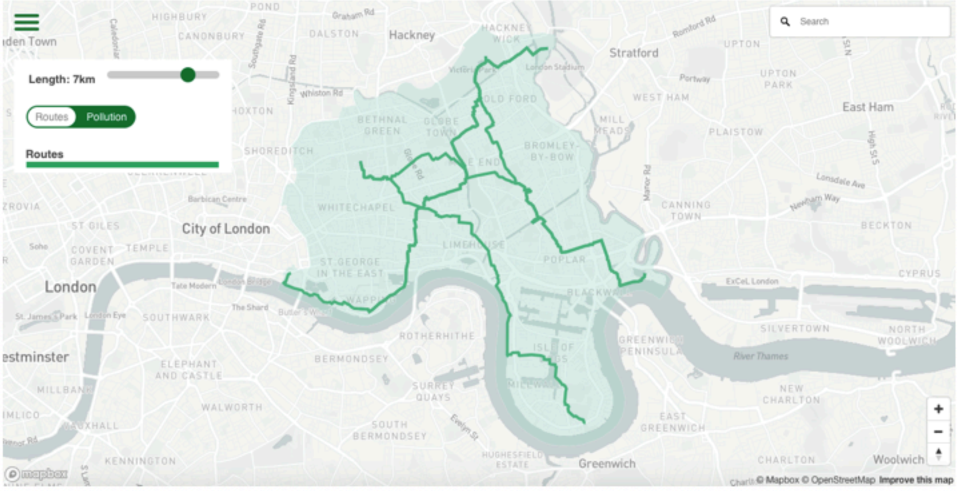
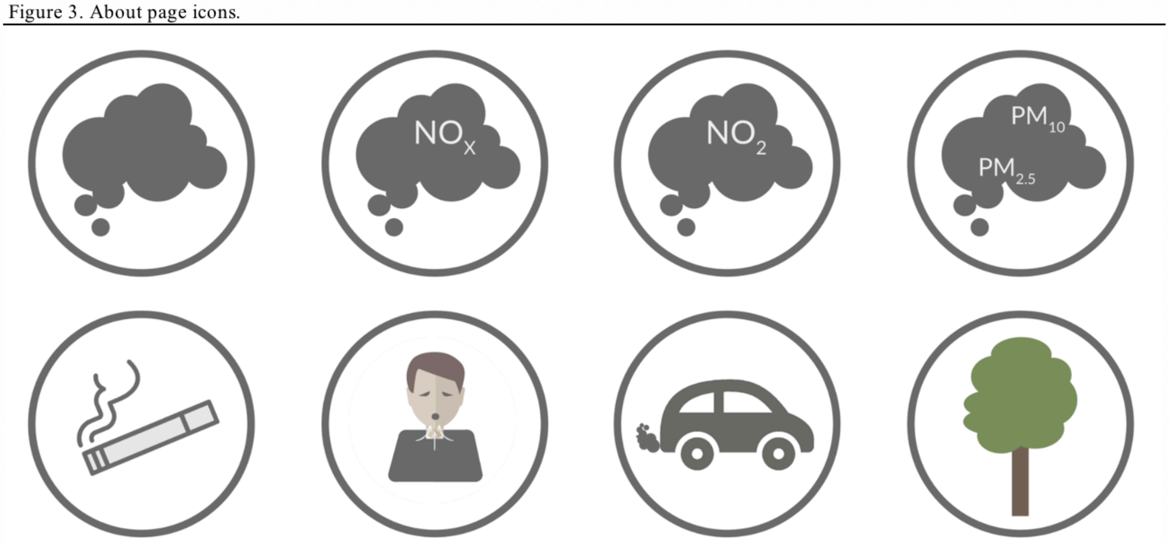


Figure 2.2. Home page healthy routes view.



The About page provides some background information on air pollution within London. In both text and pictorial format, information on pollutants themselves and their contextual importance within London, are provided. The pictorial icons below the text, shown in figure 3, provide pop-up, hover-over interactivity within the About page – by hovering the cursor over each of the eight icons, users can gain further information including contextualisation about poor air quality. Especially interesting are the two icons that provide information about the equivalent toxicity of cigarettes smoked to poor air quality inhaled, as well as the potential reduction of air pollution by the planting of new trees. These two equivalency metrics were selected, in order to provide users with some context regarding air quality in London. The statistic selected for cigarette smoke equivalency was: 24-hour exposure to 200 μgm^{-3} of NO_2 is comparable to a daily dose of 0.3 cigarettes (Van der Zee, S.C., Fischer, P.H. & Hoek, G., 2016). Everyone is aware of the carcinogenic effects of smoking, and providing an equivalency between such an unhealthy activity, and simply breathing in London, illustrates the extent of the poor air quality. If we refer back to table 1, we can see that the maximum recorded 2013 value of 192 μgm^{-3}, is very close to the 200 μgm^{-3} quoted in the statistic.

Figure 3. About page icons.



On the Analysis page, a clear and simple explanation of the project methodology is given. Pictures of the four individual pollutant maps as well as the combined pollutant map are provided (as in figure 1 above). An interactive bar chart displays the minimum, maximum and mean levels for all four pollutants, as well as the EU limit values (consistent with table 1). Users are able to hover over the different columns to view the exact values. Figure 4 illustrates both the bar chart and its interactive capabilities. The decision to allow users to hover over the columns to view the individual values, rather than simply displaying the values at all times, was made to maintain the clean website aesthetic. The design decision to not overload users with information was made early on in the project, and this attitude was balanced with displaying as much information as possible.

Figure 4. Analysis page interactive bar chart.

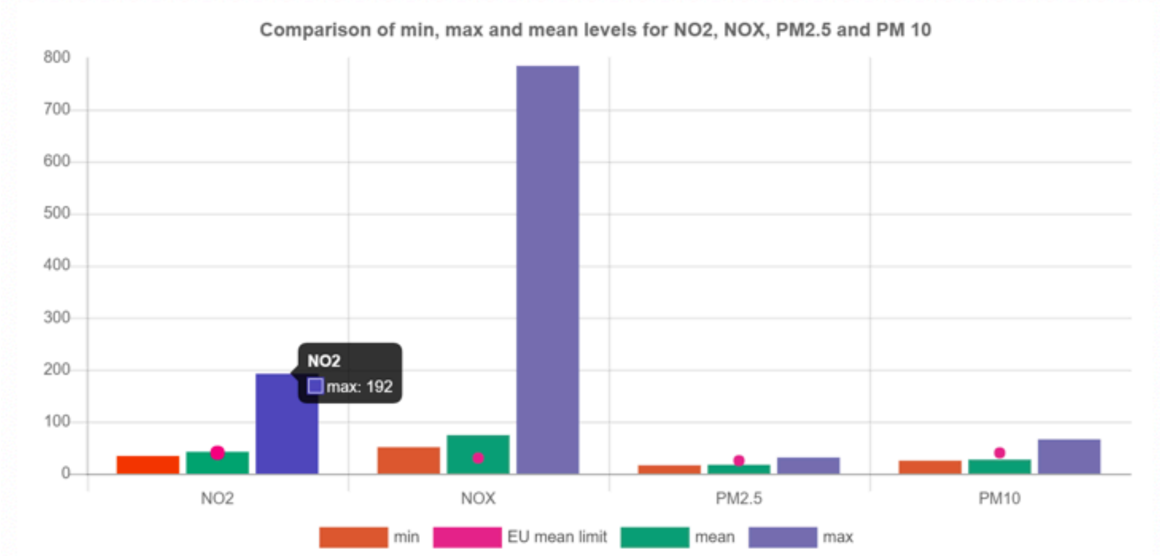
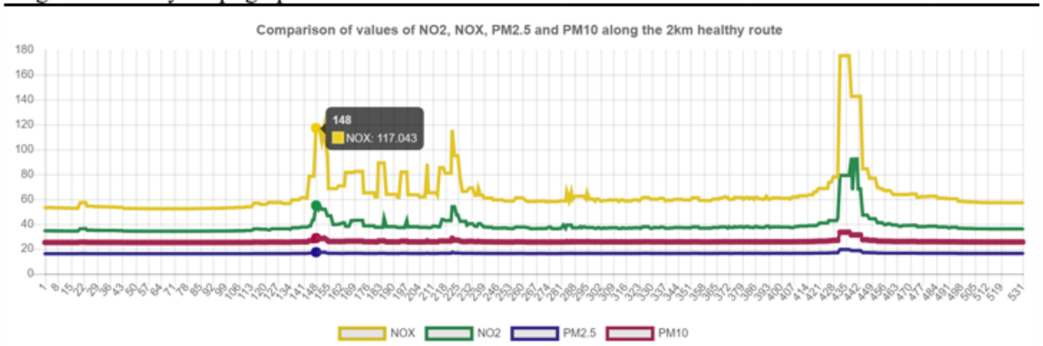


Figure 5 illustrates a line chart of individual pollutant values that correspond to route 7 in table 2. This provides users the ability to see the exact pollutant levels on the individual routes. The x-axis of figure 5 shows the interpolated pollution data for the 531 points used to create the route itself. While the 45 created routes are those with the least mean pollution, figure 5 demonstrates that within these healthy routes, pollution hotspots still occur. It is important to provide a visualisation of how the individual pollutants vary over the course of a specific route, as we can see that some pollutants (NOx and NO2 in this particular example) may vary considerably, while the variation of others is negligible.

Figure 4. Analysis page pollution line chart.



A link to the source data is also provided, so if desired, users can easily download and explore the data themselves. Perhaps most importantly, a link to the source code is provided, enabling users to explore the specific analysis techniques utilised within the project. This was included to provide transparency and reproducibility.

2.5. Technical Integration Between Elements

The website was developed to facilitate technical integration between elements. Developed using Mapbox GL JS, a JavaScript library for creating interactive maps around the open-source mapping platform Mapbox, and jQuery, the website visualises the previously created analysis in an interactive manner.

Firstly, the vector datasets were uploaded to Mapbox, including a geojson file of all 45 routes, and a csv pollution file consisting of NO_2, NO_x, PM_{10}, $PM_{2.5}$, and combined values. Secondly, the layers were visualised using Mapbox GL JS, using a light Mapbox base map. JQuery was then utilised to enable the user to toggle between pollution and routes, and to filter the routes by length with a slider. Further Mapbox functionalities were also utilised, for example, users can zoom into the map itself, allowing the closer examination of specific points. Charts on the Analysis page were created with the latest version of Chart.js.

There are two online repositories for this project. The code is open source and is hosted at: github.com/OpenAirWalk/openairwalk.github.io. The website, published on GitHub Pages, is located at openairwalk.github.io.

3. Conclusions

3.1. Limitations

Currently, the major limitation on the wider application of this project is data availability. While the LAEI is a richly informative and fine-grained dataset, it is a dataset that has only been collected once (in 2013) and is now five years out of date. It would be interesting to see data at the same granularity over time, in order to provide an understanding of how pollution has changed within Tower Hamlets and indeed, across the whole of London. This would enable analysis of the time periodicity and could provide valuable insight into whether (and to what extent) initiatives such as the Congestion Charge (introduced in 2003) have affected air pollution in London, even in areas that do not fall within the congestion charging zone. Unfortunately, as the data has insufficient temporality, it is impossible to visualise the change in air pollution over a 24-hour period. This analysis would be interesting, as pollution fluctuates throughout the course of the day, with on-peak and off-peak traffic flows.

A secondary limitation is the pre-calculated routes themselves. The website provides a visualisation of healthy routes, and while it does allow users to select routes of different distances, it does not allow users to choose their own start and end points. The primary reason for this limitation was time availability. Given the relatively short time allowed for project completion, this project focused on producing a high quality, user-friendly visualisation tool, rather than adding additional functionalities that may or may not have been ready. Allowing users to input specific start and end points would allow a hugely extended range of applications, such as choosing a healthy route to work, a healthy route around a specific neighbourhood, and a healthy route starting and ending at the same point, such as someone's home.

A further limitation, is that some of the routes overlap to varying degrees. For example, for the 1km route lengths, there appear to only be 4 routes, but there are in fact 5. This limitation extends to longer route lengths, which in some instances, simply add together shorter route segments. However, the addition of shorter routes together to create longer routes may not in fact be a limitation, as certain areas of the borough have lower levels of pollution than others, and so are more likely to yield healthier routes, regardless of route length. It should be noted that routes were capped at 9 km in length, as longer routes extend outside the borough boundaries – the borough is not large enough to allow unique routes over 9 km to be created.

3.2. Future Work

The primary opportunity for future work, would be to extend this tool to cover all London boroughs. Due to time constraints, such an extended scope was not attempted for this project. A further opportunity, would be to extend coverage to all cities where sufficiently fine-grained pollution data is available. This would provide comparative capabilities, and not only allow the development a database of healthy walking routes in many different cities, but also to develop a greater understanding of how and why some clean air initiatives succeed whilst others fail. If a longer time series of data was available for a wide geographic area, cities where clean air initiatives were adapted early on, such as the Congestion Charge within London, could provide invaluable case studies of best practices as well as initiatives to avoid, to cities that have yet to adopt initiatives. Such a capability would be particularly interesting in the coming years, with many clean air initiatives soon to be adopted (such as London's Ultra Low Emission Zone), and also with the increasing popularity of hybrid and electric vehicles.

A further iteration of this project, would be to integrate automated real-time sensor data into the tool, and allow the routes to update automatically based on this new data. Readings could be stored in a database at hourly intervals, allowing users and policy makers alike to view times of the day and year that were historically lower in pollution. This tool could be integrated with the existing sensor network in London, the London Air Quality Network (LAQN) to provide a comprehensive information repository for all London residents, allowing the selection of times of day for outdoor activity based on the current air pollution conditions. This sensor network, hosted by the Environmental Research Group at King's College London on the 'Londonair' website, already provides hourly air pollution data for the whole city. This data could be integrated into this project, or vice versa, providing residents of all London boroughs not only the general air pollution conditions within the city, but specific low pollution, healthy routes, based on real time data. These might change depending on the time of day or weather conditions and would give users the ability to make more informed choices.

Another opportunity for a richer user experience, would be to enable user-input start and end points, enabling people to select their own routes, such as a least polluted route to work, or a route taking in specific, multiple locations, such as the functionality that Google Maps allows. An additional functionality would be to enable the creation of loop routes, that start and end at the same point, such as an individual house, while travelling a user-specified distance. This has obvious benefits to the user – currently, the visualisations focus on pre-calculated routes of varying distances, that start and end at specific points. Allowing users to select their own start and end points would allow more people to make the most of the tool, as they would not be limited to routes that may not be easily accessible given their specific location.

In the future, it may also be possible to link to air pollution sensor data to the traffic light network. This would enable traffic flow management based on air pollution, which could be a new way of managing the combined impact of pollution and congestion. Either by utilising designated high-flow routes, or simply by changing the traffic signals so that congestion is mitigated against, congestion and concentrations of pollutants could possibly both be reduced. It may not be particularly advantageous to disperse traffic more widely across the borough, as this may increase the air pollutants where they are currently low, but by integrating the traffic light network and the air pollution sensors, there is the possibility to reduce congestion and therefore overall levels of pollution.

The line chart visualisation displayed on the Analysis page for the pollutant values of route 7, provides a further opportunity for future work. At the moment, this visualisation was provided to demonstrate the capability to display specific pollutant values for an individual route. In the future, this could be extended to another page of a website, providing a side by side view of the individual route (as on the Home page), and the line chart (as on the Analysis page). This would allow users to see how the pollutant values vary along each individual route. Given more time, this would be enabled by a hover over functionality, as in figures 4 and 5 currently displayed on the Analysis page. While hovering over the route, a hover-over point would appear at the corresponding location on the line chart, and vice versa.

3.3. Conclusions

This project was intended to create a tool to visualise spatial data in a way that enables any user to generate useful insight. In its current iteration, its target audience is both users and policy makers within the borough of Tower Hamlets. The tool provides both groups with information about the borough's air pollution, contextualised within the idea of healthy walking routes. This is useful for users of the borough, as they can see where different length routes of low pollution are located, and therefore both the best places to enjoy the borough's open spaces and the places to avoid whilst outside. Policy makers can view these routes and gain an understanding of the areas that currently yield healthy routes, as well as areas that require targeted initiatives to improve their air quality.

The tool succeeds in achieving its objectives, as well as providing great opportunities for further functionalities and future work. Given more time, the limitations mentioned above could be mitigated and the future work could be completed. It would then be possible for the tool to either function on its own as a standalone product or serve as a more user-friendly extension of the 'Londonair' website hosted by the Environmental Research Group at King's College London. For dissemination purposes, serving as an extension of a pre-existing website may enable access to a greater number of people, but as a standalone website, its original simple, user-friendly interface may appeal to a wider audience.

4. Bibliography

Aether, 2017. Updated Analysis of Air Pollution Exposure in London: Report to Greater London Authority. [online]. Available from: https://www.london.gov.uk/sites/default/files/aether_updated_london_air_pollution_exposure_final.pdf [Accessed 1 May 2018].

Au, C., 2017. Peak Spotting: Visual Tools for Managing the Capacity of Germany's Rail Traffic Network. [online]. Available from: http://www.nand.io/projects/clients/peak-spotting/ [Accessed 1 May 2018].

Auto Accessorise Garage, 2018. Average Commute Times. [online]. Available from: https://www.autoaccessoriesgarage.com/average-commute-times [Accessed 1 May 2018].

Cohen, A.J., et al., 2017. Estimates and 25-year Trends of the Global Burden of Disease Attributable to Ambient Air Pollution: An Analysis of Data from the Global Burden of Diseases Study 2015. The Lancet, 389(10082), pp. 1907-1918

Greater London Authority, 2016a. London Atmospheric Emissions Inventory (LAEI) 2013: Methodology. [online]. Available from: https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory-2013 [Accessed 1 May 2018].

Greater London Authority, 2016b. London Atmospheric Emissions Inventory (LAEI) 2013. Available from: https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory-2013 [Accessed 1 May 2018].

London Assembly, 2018. How We're Cleaning up London's Air. [online]. Available from: https://www.london.gov.uk/what-we-do/environment/pollution-and-air-quality/how-were-cleaning-londons-air [Accessed 1 May 2018].

UK Air, 2018. UK and EU Air Quality Limits. Department for Environment, Food & Rural Affairs. [online]. Available from: https://uk-air.defra.gov.uk/air-pollution/uk-eu-limits [Accessed 1 May 2018].

WHO, 2018. 9 out of 10 People Worldwide Breathe Polluted Air, But More Countries are Taking Action. [online]. Available from: http://www.who.int/news-room/detail/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action [Accessed 1 May 2018].

Yau, N., 2007. How Americans Get to Work. [online]. Available from: http://flowingdata.com/2015/01/20/how-americans-get-to-work [Accessed 1 May 2018].

Van der Zee, S.C., Fischer, P.H. & Hoek, G., 2016. Air Pollution in Perspective: Health risks of air pollution expressed in equivalent numbers of passively smoked cigarettes. Environmental Research, 148, pp. 475-483.

5. Appendix

5.1. Procedure to create the routes of lowest mean pollutant.

In [15]:
```
import gdal, osr
from skimage.graph import route_through_array
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import rasterio
```

In [2]:
```
def raster2array(rasterfn):
    raster = gdal.Open(rasterfn)
    band = raster.GetRasterBand(1)
    array = band.ReadAsArray()
    return array
```

In [3]:
```
def coord2pixelOffset(rasterfn, x, y):
    raster = gdal.Open(rasterfn)
    geotransform = raster.GetGeoTransform()
    originX = geotransform[0]
    originY = geotransform[3]
    pixelWidth = geotransform[1]
    pixelHeight = geotransform[5]
    xOffset = int((x - originX)/ pixelWidth)
    yOffset = int((y - originY)/pixelHeight)
    return xOffset,yOffset
```

In [4]:
```
def createPath(CostSurfacefn,costSurfaceArray,startCoord,stopCoord):

    # coordinates to array index
    startCoordX = startCoord[0]
    startCoordY = startCoord[1]
    startIndexX,startIndexY = coord2pixelOffset(CostSurfacefn,startCoordX,startCoordY)

    stopCoordX = stopCoord[0]
    stopCoordY = stopCoord[1]
    stopIndexX,stopIndexY = coord2pixelOffset(CostSurfacefn,stopCoordX,stopCoordY)

    # create path
    indices, weight = route_through_array(costSurfaceArray, (startIndexY,startIndexX), (stopIndexY,stopIndexX),geometri
    indices = np.array(indices).T
    path = np.zeros_like(costSurfaceArray)
    path[indices[0], indices[1]] = 1
    return path
```

In [5]:
```
def array2raster(newRasterfn,rasterfn,array):
    raster = gdal.Open(rasterfn)
    geotransform = raster.GetGeoTransform()
    originX = geotransform[0]
    originY = geotransform[3]
    pixelWidth = geotransform[1]
    pixelHeight = geotransform[5]
    cols = array.shape[1]
    rows = array.shape[0]

    driver = gdal.GetDriverByName('GTiff')
    outRaster = driver.Create(newRasterfn, cols, rows, gdal.GDT_Byte)
    outRaster.SetGeoTransform((originX, pixelWidth, 0, originY, 0, pixelHeight))
    outband = outRaster.GetRasterBand(1)
    outband.WriteArray(array)
    outRasterSRS = osr.SpatialReference()
    outRasterSRS.ImportFromWkt(raster.GetProjectionRef())
    outRaster.SetProjection(outRasterSRS.ExportToWkt())
    outband.FlushCache()
```

In [6]:
```
def main(CostSurfacefn,outputPathfn,startCoord,stopCoord):

    costSurfaceArray = raster2array(CostSurfacefn) # creates array from cost surface raster

    pathArray = createPath(CostSurfacefn,costSurfaceArray,startCoord,stopCoord) # creates path array

    array2raster(outputPathfn,CostSurfacefn,pathArray) # converts path array to raster
```

In [2]:
```
CostSurface  = 'E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/pyhtonCoding/1-NewMaster/THamStAll4Friction1.ti
```

In [8]:
```
if __name__ == "__main__":
    CostSurfacefn = CostSurface
    startCoord = (536552.739718,184199.345531)
    stopCoord = (533542.17254,180680.0646)
    outputPathfn = 'E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/pyhtonCoding/1-NewMaster/test1.tif'
    main(CostSurfacefn,outputPathfn,startCoord,stopCoord)
```

In [16]:
```
StartpointsEndpoints = pd.read_csv('E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/pyhtonCoding/1-NewMaster/po
StartpointsEndpoints.head()
```

Out[16]:

	FID	X	Y
0	0	536552.739718	184199.345531
1	1	535730.226088	182879.641548
2	2	534644.392062	180140.151694
3	3	537923.475729	182736.088233
4	4	537011.546346	179307.633403

In [17]:
```
for i in range(len(StartpointsEndpoints)):
    if __name__ == "__main__":
        CostSurfacefn = CostSurface
        startCoord = (StartpointsEndpoints.loc[i,'X'],StartpointsEndpoints.loc[i,'Y'])
        for j in range(len(StartpointsEndpoints)):
            stopCoord = (StartpointsEndpoints.loc[j,'X'],StartpointsEndpoints.loc[j,'Y'])
            outputPathfn = 'E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/pyhtonCoding/1-NewMaster/output/pat
            main(CostSurfacefn,outputPathfn,startCoord,stopCoord)
```

In [84]:
```
for i in range(3):
    for j in range(3):
        print("file_" + str(i)+str(j) + ".dat")
```

```
file_00.dat
file_01.dat
file_02.dat
file_10.dat
file_11.dat
file_12.dat
file_20.dat
file_21.dat
file_22.dat
```

In [29]:
```
!gdal_calc.py -A CostSurface -B CostSurface --outfile='E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/pyhtonCo
```

```
Traceback (most recent call last):
  File "C:\Users\Blusubmarine\Anaconda3\Scripts\gdal_calc.py", line 4, in <module>
    __import__('pkg_resources').run_script('GDAL==2.2.4', 'gdal_calc.py')
  File "C:\Python27\ArcGIS10.5\lib\site-packages\pkg_resources\__init__.py", line 2927, in <module>
    @_call_aside
  File "C:\Python27\ArcGIS10.5\lib\site-packages\pkg_resources\__init__.py", line 2913, in _call_aside
    f(*args, **kwargs)
  File "C:\Python27\ArcGIS10.5\lib\site-packages\pkg_resources\__init__.py", line 2940, in _initialize_master_working
_set
    working_set = WorkingSet._build_master()
  File "C:\Python27\ArcGIS10.5\lib\site-packages\pkg_resources\__init__.py", line 635, in _build_master
    ws.require(__requires__)
  File "C:\Python27\ArcGIS10.5\lib\site-packages\pkg_resources\__init__.py", line 943, in require
    needed = self.resolve(parse_requirements(requirements))
  File "C:\Python27\ArcGIS10.5\lib\site-packages\pkg_resources\__init__.py", line 829, in resolve
    raise DistributionNotFound(req, requirers)
pkg_resources.DistributionNotFound: The 'GDAL==2.2.4' distribution was not found and is required by the application
```

In [2]:
```
import pandas as pd
NO2Histogram = pd.read_csv('E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/data/Histogram Table/HistogramNO2.c
```

In [16]:
```
NO2Histogram.head()
NO2Histogram = NO2Histogram.rename(columns={1:48})
NO2HistogramTest = pd.read_csv('E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/data/Histogram Table/HistogramN
NO2HistogramTest.head()
plt.hist(NO2HistogramTest['NO2'])
plt.hist(NO2HistogramTest['NO2'], color = 'blue', edgecolor = 'black', bins = int(160/5))

sns.distplot(NO2HistogramTest['NO2'], hist=True, kde=False, bins=int(92/4), color = 'blue',hist_kws={'edgecolor':'black
```

Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0xa689828>

In [68]:
```
TowerNO2 = ('E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/pyhtonCoding/rasterCalculation/pollutants/TowerNO2
TowerNOx = ('E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/pyhtonCoding/rasterCalculation/pollutants/TowerNOx
TowerPM10 = ('E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/pyhtonCoding/rasterCalculation/pollutants/TowerPM
TowerPM25 = ('E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/pyhtonCoding/rasterCalculation/pollutants/TowerPM


with rasterio.open(TowerNO2) as TowerNO2:
    NO2 = TowerNO2.read()

with rasterio.open(TowerNOx) as TowerNOx:
    NOx = TowerNOx.read(1)

with rasterio.open(TowerPM10) as TowerPM10:
    PM10 = TowerPM10.read()

with rasterio.open(TowerPM25) as TowerPM25:
    PM25 = TowerPM25.read()



NO2Normal = (NO2 - NO2.min()) /(NO2.max() - NO2.min())

NO2Normalized = 'E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/pyhtonCoding/rasterCalculation/pollutants/NO2N

with rasterio.open(NO2Normalized, 'w',**TowerNO2.meta) as output:
    output.write(NO2Normal)
```

```
C:\Users\Blusubmarine\Anaconda3\lib\site-packages\rasterio\__init__.py:160: FutureWarning: GDAL-style transforms are
deprecated and will not be supported in Rasterio 1.0.
  transform = guard_transform(transform)
```

In [ ]:

```
with rasterio.open('R.byte.tif') as src:
arr = src.read()
arr = (arr * 9) / 5 + 32


 myTestlayer = myTestlayer + 1000
    xMax = myTestlayer.max()
    xMin = myTestlayer.min()

xMax
xMin


with rasterio.open('output.tif', 'w',**src.meta) as dst:
    dst.write_band(1, arr[0])


NO2Normalized = 'E:/2018/CASA/Term 2/4-Spatial Data Capture/0-coursework/pyhtonCoding/rasterCalculation/pollutants/NO2N

with rasterio.open(NO2Normalized, 'w',**TowerNO2.meta) as output:
    output.write(NO2Normal)

NO2Normal = (NO2 - NO2.min()) /(NO2.max() - NO2.min())
```