Invisible Cities – Air Quality

Group 3 – Qiuwen Ju, Tom Keel, Zixin Lyu & Jay Wilson

Total Word Count: 6000

Project outputs

Website scripts	Main index: index.html History page: history/history.html Globe vis: globe/globe.html Region vis: region/region.html Street vis: street/street.html
Website address	https://naughty-kepler-88ce41.netlify.com (*automatically generated https*)
Presentation file	presentation_Group3.pptx
Supporting scripts	General: stories.js; buttons.js; index.css Globe Vis: globe.js; globe.css World Vis: leaflet_functions.js; map.js; buttons.js Region Vis: region.js; region.css Street Vis: street.js; street.css Data-preprocessing: in 'notebook' folder
Supporting data files	Globe Vis: airPollutionDeaths.geojson; world-110m.json World Vis: stationData folder; stationIDs.csv Region Vis: stationData folder Street Vis: On Mapbox
Links to supporting sites and technologies	Github: https://github.com/CASA-DV-Group3/AirQuality-0/tree/master World Air Quality Index API: https://waqi.info/

Table of Contents:

Table 0. Outline of contributions of group members per section of website and report

Task Name	Major contributors	Additional contributors	Relevant report section
Concept development	All members		All
Data preparation	Lyu	Jay	3.1
Data Integration	Lyu, Tom	Qiuwen	3.1
Introduction & Background pages	Jay, Qiuwen	Tom	3.4
3D Global Health Impact Visualisation	Tom		3.2.1
World Air Quality Monitoring Visualisation	Qiuwen,Tom	Jay, Lyu	3.2.2
Regional-Level Animated Air Quality Monitoring Visualisation	Lyu		3.2.3
Street-Level Air Quality Visualisation	Qiuwen		3.2.4

Collaboration & hosting	Tom		3.5
Design	Qiuwen	Jay	3.3

Individual Contributions:

Jay: (website) After failing to get the right raster data from ESA's Copernicus website and not having a 3D Unity street to use cartoon faces on, I focused on engaging the visitor with text to get them thinking about air, illustrating the human cost of air pollution by providing a history of disasters and government responses, and the components and effects of pollution. Too late I thought of a way to use bounding boxes for satellite measurements of ozone or vegetation, which would have been good additions to the globe.
(report) I wrote and planned the initial layout of the report while the others in the team produced a successful website: my sections for the introduction, literature review, early inspiration, user interaction, user engagement, narrative, challenges & solutions, improvements and conclusions were then amended as required for the final version.

Lyu: (website) I took charge of the data collection and processing part, as well as the Regional-Level Animation Visualisation page of our website.
(report) I wrote the methods of collecting and handling the air quality data from The World Air Quality Index project, the methods of visualising the regional-level dynamic air quality data in Part 3-Methods. In addition, I also helped modifying the Part 1-Introduction.

Qiuwen: (Website) I took charge of the overall design of the website, which includes building the initial structure of website and helping to keep data representation and map layout of each part adopting a same design scheme. I did all work for creating Central London Air Pollutant Concentration Map. In the programming work of World Air Quality Monitoring Visualisation, I contributed to write a Javascript function to draw line chart of historical AQI data.
(Report) The sections I wrote include 3.1.3 London Atmospheric Emission Modelling Data, 3.2.2 World Air Quality Monitoring Visualisation – feature and style 3.2.4 Central London Air Pollutant Concentration Map and 3.3 Design.

Thomas: (website) I was the technical lead for the project, setting up the structure and outline of the website. My responsibilities included creating and writing descriptions of the global health visualisation and the working on the global monitoring visualisation with Qiuwen. I also contributed to text and description on the project's landing page. In the data pre-processing phase I helped with the integration of the data from the API into the website and cleaning and merging the global health data. (report) For the report, I wrote sections relating to my contribution on the website and served as an editor of the final version of the report.

1 Introduction

In our daily lives, the air we breathe is inherently part of a larger, yet invisible, system: the atmosphere – the contents of which, play an essential role in our physical health (Gao *et al.*, 2016). A clean environment including a atmosphere producing clean air is a basic requirement for human health, yet (outdoor) air pollution remains one of the World Health Organisation's (WHO) major *global* causes of preventable deaths (WHO, 2005; Linou, 2018).

Across the planet, the air that we breathe is commonly being polluted by harmful chemicals such as carbon monoxide, nitrogen oxides, sulphur oxides and particulate matter which are produced mostly by human activities (i.e. industrial practices and fossil-fuel powered vehicles). As such the health and welfare of humans, animals and plants is being severely threatened (Park & Allaby, 2017). In 2016, it was estimated that 91% of the world's population were inhabiting both urban and rural environments that did not meet suggested global air quality guidelines (set in 2005 by the WHO; WHO, 2018).

Despite the devastation of air pollution and the overall need for good air quality in our environment, it has still hard been hard for decisions to be made (at any political scale) regarding them (Kutlar *et al.*, 2007). Simply raising awareness of the dangers of air pollution level has been challenging – probably owing to the *'invisible'* nature of air. This project, thus aims to visualise the impact of air pollution and at different spatial and temporal scales. Specifically, this project's website moves down through the global-scale health impacts of pollutants and the spatial coverage of monitoring stations, before focusing in on increasingly fine scales (i.e. regional–city–street levels).

The objectives of the visualisations included in the project primarily focused on three key areas:

1. Provide a platform for the dissemination of the impact of air pollution.
2. Highlight the current state of monitoring – the global visibility of air quality.
3. Examine and generate insight into the space time dynamics of air pollution.

The extent to which these objectives are achieved by this project and its outputs are discussed in Section 4.

This structure of this report is as follows: it first examines the relevant literature relating to the air pollution and its visualisation, including an overview of the motivation for this project in section 2. After this, in section 3, we overview the visuals and discuss key design and narrative based decisions, before summarising and discussing the project's overall outputs, challenges and solutions in section 4.

2 Literature Review

Following advances and reduction of costs in monitoring technology in recent years, the global capacity to provide highly sophisticated air pollution measurements has improved (Pattinson et al., 2014; Rai *et al.,* 2017). As such, there has been an increasingly fine-scale approach to studying pollutant concentrations compared to in previous decades – where air quality reports were often aggregated and expressed in terms of 'smogginess' (Schwartz & Dockery, 1992). These technological advances have fuelled further research into air pollutants, although this is a process that remains within the academic realm (Kelly & Fussell, 2015). Indeed, our review of literature find no real example of web visualisations which aim to disseminate the state of air pollution monitoring to the public directly.

Policy relating to dangerous air quality levels have been relatively limited globally, although there are some regional and national successes i.e. the reduction in pollution and thus pollution-related diseases across the EU (Guerreiro et al., 2019). There are a few examples of success globally, such as the Montreal Protocol in 1997 (and subsequent related protocols) or the World Health Organisation's: Air Quality Guidelines introduced in 2005 which set out guidelines for maintaining safe levels of air pollution policy (Chipperfield et al., 2015). Because of the importance of the context of historical air pollution policy and monitoring, we thus have made a decision to accompany our website with a short history of major pollution events.

Motivation:
As there are not many visualisations that directly focus on the state of air pollution at different scales, we have instead drawn inspiration from projects which focus on the general visualisation and thus dissemination of natural processes (such as those relating to climate). These visualisation favour presentational communication of their information aimed at academic groups, as opposed to ones that are directly insightful for any user (i.e. visualisations that walk the user through them) (Kirk, 2016). Thus this project's visualisations attempt to be first informative and, secondly, educational. We hope to achieve this by offering a high level of user-interactivity in the maps, something that has become more achievable with increasingly sophisticated web-mapping technologies (Smith, 2016).

One major inspiration for our initial *global-scale* visualisation comes from *nullschool.net*, which displays live modelled satellite data spanning across a 3-dimensional globe (Beccario, 2019). We believe this visualisation is effective in translating the inter-connectivity of a physical process (such as air pollution) and we aim to emulate this by using a combination of 3D, 2.5D and 2D visuals.

Another inspiration and, one of data sources, comes from the World Air Quality Index Project (WAQIP, 2019), which plots live air pollution data. This visualisation is dynamic and remains relevant as it constantly changing (i.e. depending on the time and day). Furthermore it offers additional interactivity if users want to discover more in the data, which is achieved in the form of popup on top of each data point (Kirk, 2017).

3 Methods

3.1 Data Collection & Handling

3.1.1 Global Health Data (used in the *global health impact visualisation*)

Two types of data, relating to general exposure to, and death rate (per 100,000 persons) from the presence of local concentrations of particulate matter with particle size of less than 2.5 µg (PM2.5). Both measures derive from the results of the *Global Burden of Disease Study 2017* available from GBDCN (2018). Owing to the difficulty with ascertaining deaths directly from PM2.5, the data has its drawbacks, as the GBDCN broadly define death rate as the number of people who die prematurely with pollution-related diseases in an area (Ritchie & Roser, 2017). Exposure is measured in local air concentrations of PM2.5 (in µg/m³), determined using a model combining global satellite data and wind regimes (Ritchie & Roser, 2017).

This data is grouped into 165 countries, covering the years 1970–2017, later reduced to 1990–2017 due to missing data. The data were read into Python using geo*pandas* and merged together with a GeoJSON of the boundaries of world countries. Due to slight changes in the names of the countries between the two types of data sets, the *difflib* library was used to programmatically look for the similarity of the two names in a string format i.e. the names 'United States' and 'United States of America' have a similarity score of 0.71. Looping through similar strings, a user prompt was created which allows the input of 'y' or 'n' to change or keep the country name string and is based on this similarity scores.

Finally, *shapely* was used to find the coordinates of the centroids for each country's boundary, so that the global health data has a location to be projected onto the globe visualisation. This stage of pre-processing is detailed in the 'pm_deaths_and_exposure.ipynb' notebook.

3.1.2 World Air Quality Index Data (used in the *global and regional visualisations*)

Air quality data was retrieved via an API from the World Air Quality Index Project (WAQIP, 2019), which provides transparent air quality information for more than 88 countries. The data covers more than 11,000 stations in 1000 major cities, which is the most extensive and abundant open data source in the field of air quality.

The air quality data from 11,772 air quality monitoring stations all over the world and this can be accessed through the JSON API and by specifying the ids of the monitoring stations, this returns the corresponding air quality data, including the AQI (Air Quality Index), the corresponding station name, station geolocation and time of collection. The data is real-time and update every hour.

The data were then pre-processed and cleaned, as detailed in Figure 1 below, for problems such as error messages, null values, duplicates and unreasonable collection times. This reduces the amount of valid air quality station to 8,990 which update every hour.

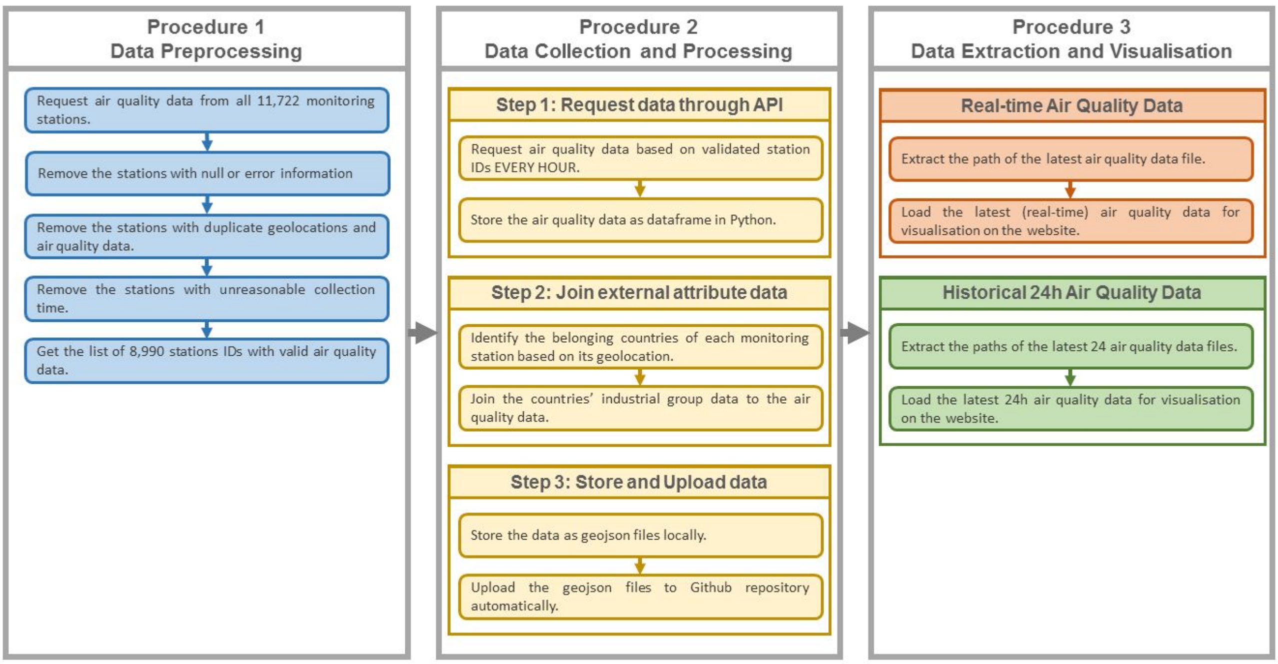

Figure 1. Data Collection and Handling process

These methods are specifically carried out using Python scripts and Jupyter notebooks (see "AQI_API_Retrieval_and_Processing" under the "notebooks" directory). We run one of these scripts locally to request air quality data from the API every hour so that collect the real-time data (a better solution would be to have a dedicated server pulling data from the API).

After collection, the *pandas* and *geopandas* libraries are used to merge the data with a GeoJSON detailing each country and its *standardised* name (see 3.1.1) and a dataset containing information about the level of industrialisation of each country (specifically, the percentage of GDP that the industrial sector contributes to from each country) from the World Bank (2017).

Finally, a dataset detailing every hour is then exported into a GeoJSON and automatically uploaded to Github through the Github API for use by the front-end of the project.

3.1.3 London Atmospheric Emission Modelling Data (used in the *street-level visual*)

A dataset containing simulated air pollutant concentration of Greater London is applied to create a street-view map. To address the need of Greater London Authority (GLA) and Transport for London (TFL) to improve air quality in London through policy-making, King's College London's (KCL) Environmental Research Group developed a model to estimate and predict accurate atmosphere emission across Greater London (Beevers et al, 2016; London Air, 2019). Based on the pollution London Atmospheric Emissions Inventory from 2008, 2011 and 2015, spatially detailed emission is calculated across a 20m*20m grid spanning London (GLA, 2015). The output data series contains five categories of air atmospheric emissions: annual average emission of NO2, NOX and PM10 as well as the emission of PM10 and PM2.5 that exceeds EU limits. The figure shown below is the example of NO2 annual mean emission data.

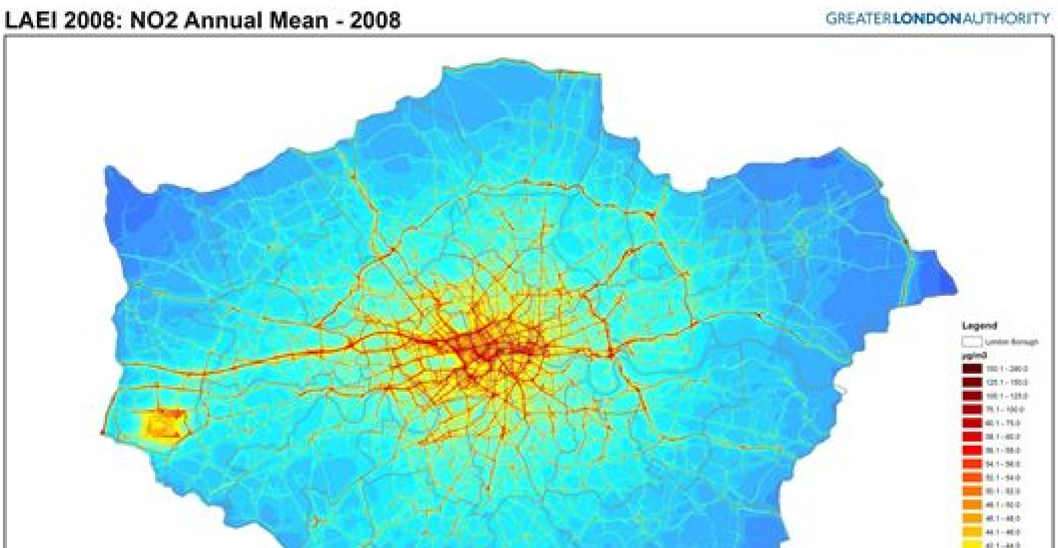

Figure 2. Example of raw emission modeling data (*source: GLA, 2015*)

Because of the high resolution of the data, difficulties can be caused in the processes of both hosting and loading. At the same time, as shown in Figure 2 and as expected, high concentrations of pollution often concentrates in central areas. As such for the street-level visualisation in this project, we select a subset of approximately 45,057 20m*20m cells. This decision was taken to ensure:

1. collaboration with Mapbox GL;
2. fast and smooth user experience;
3. efficient data dissemination and visualisation.

The data itself are five shapefiles contains emission data of three discontinuous years (2008, 2011, 2015). Using ArcGIS, all files are filtered down so that only central London areas remained. Then all five shapefiles are merged into one with all fields represent emission are kept and renamed according to pollutant type and year. The final outcome is one shapefile with 15 fields of estimated pollutant emission – five categories in three years. An overview of this stage of the data processing flow for London modelling pollution data is shown in figure 3.

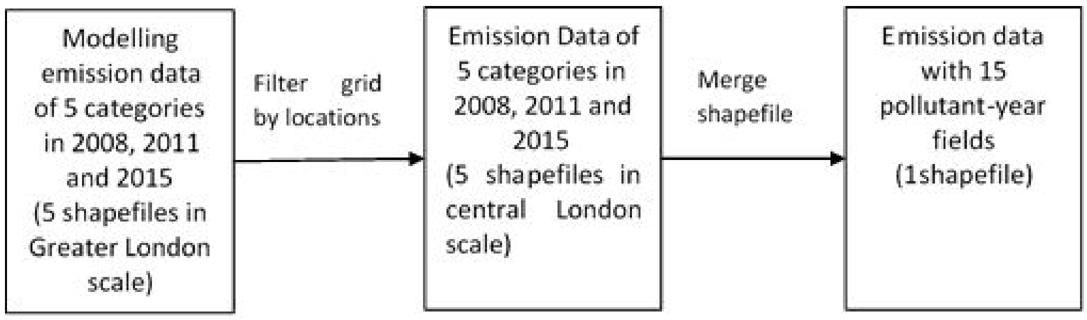

Figure 3. The processing flow of modelling emission data.

3.2 Visualisations

In this section we discuss, provide a methodology and highlight key cartographic and design decisions made with each of the visualisations and narrative text that appear in the website.

3.2.1 Global Health Impact Visualisation

Why visualise air pollution in this scale?
The first visualisation of this project examine the impact of air pollution to *global health* as it is, itself, a major source of preventable deaths globally (Linou, 2018). We decidedly found it important to visualise the quantifiable damage of air pollution, first, before evaluating the state to which it is monitored (3.2.2+3) and distributed (3.2.4) in latter visualisations – it provides the user some context into the devastation of poor air quality before viewing the rest of the project.

One major cartographic decision for this visual was to project the data (see 3.1.1) onto the 3-dimensional globe as opposed to a 2-dimensional map. This decision is important when considering the data itself, which details PM2.5 – a major global pollutant whose concentration can be carried to across the globe (Fuzzi *et al.*, 2015). Any projection on a 2D form of a map would remove the effect that the visualisation has i.e. if using a UK-centred world map this would hinder those from periphery, to the map, regions such as New-Zealand or Alaska. Then again, the data is a visualised in the form of a graduated symbol map, so it must be noted that this visualisation is not explicitly spatial, and instead serves as a thought-provoking interactive infographic (Meirelles, 2013; Kirk, 2017).

Methodology:
This visualisation contains two major components, which communicate with each other: a rotating globe and a static line graph. The globe is rendered using D3.js and TopoJSON and is adapted from source code from Vogel (2017). For this component, functions control the rendering and movement of the globe and the projection of the PM2.5 data as *GeoCircle* markers onto it (indicating country centroids; see 3.1.1).

For the line graph, Plotly.js is used to plot a dual time series containing the change in both exposure and death rate data for a given country between the years 1990-2017. And the *JStat.js* library is used to calculate a correlation (Spearman's rank-coefficient) between them. Including correlation is used as it elevates the analytical aspect given by this visual and can roughly be used to determine the relationship between exposure and deaths related to PM2.5 at per-country level (Kirk, 2017).

The *d3.json* function is used to load both the country and deaths and exposure data into the HTML canvas where event listeners wait for the user to *drag* or *click* on the visualisation, which respectively adjust the user's view of the visualisation or pause the rendering of the globe. Additional listeners wait for the user to *hover-over* or *double-click* on the country GeoJSON layer which updates the line graph to contain subset data for that country. Finally, two buttons are used which can update the data projected onto the globe.

Styling & Features:
This visualisation exist at a cross-roads of presentational and exploratory information as it contains both interactive and static presentational features (Smith, 2016). Despite this, it is hoped that interactivity is not essential for the narrative of the globe. There are aspects of the visualisation without direct reference and relying purely on user's implicit interaction i.e. hovering over the time-series plot or descriptive text reveals further information for users interested (Atterer *et al.*, 2006).

The two separate univariate data sets used in this visualisation (PM2.5 exposure and deaths per 100,000) are differentiated by colour on the globe and also by line style in the line graph to allow for ease of detection (Meirelles, 2013). Due to the sensitivity of the data being projected onto the globe and the continuous nature of the data, a one-colour scale is used (more serious colours).

With this type of visualisation, a graduated symbol map, we are aware that the number of classes used to divide the data can significantly influence the perception of the visual and, for this visual, the radius of the circle marker, so we keep this at 6 classes for both datasets (Meirelles, 2013). Where the radius of each circle marker (in pixels) is calculated according to square root of these classes (after Meirelles, 2013). A drawback from the styling is that the categorical breaks chosen for the legend are relatively arbitrary i.e. 5 groups of 20 and 1 group of 20+, in reality the distribution of both these continuous variables contain long-tails (see figure 4). Finally, the brightness of the globe itself is reduced to emphasise the lightness of the data projected onto it (Kirk, 2017;). The opacity of the outer GeoJSON country layer has been reduced to give the effect of highlighting certain regions on user's hover-over.

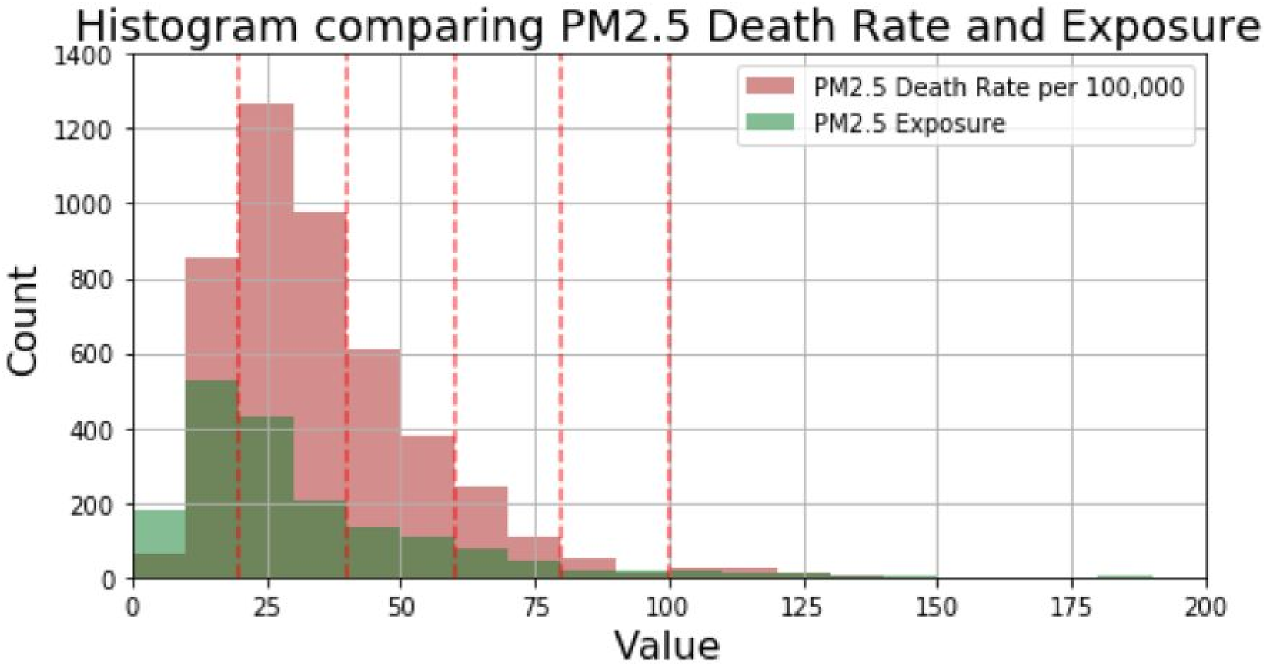

Figure 4. Histogram comparing the distribution of the two continuous variables used in the *Global Health Impact* visualisation (red lines indicate breaks for classes chosen; *n=165*).

3.2.2 World Air Quality Monitoring Visualisation

Why visualise air pollution in this scale?
In the second visualisation of this report, we want to emphasise how the invisible nature of air pollution is one of the major factors which hinders its comprehension, and thus, control across the world (Monks *et al.*, 2009). This visualisation is unique in the report as it *requires* the user's direct interaction to reveal different section of the map – "making the invisible visible". This revealing process is intentional, however, this visualisation is hoped to inspire the user through a process of insight generation (Roth, 2013). An example of one such insight could be the conclusion that there is not, as of May 2019, enough monitoring stations to record air pollution across the world (Kutlar *et al.*, 2015).

The world view is chosen as the entry point to this visualisation as the user will be presented with all the information at once. As such at the global level, this visualisation contains one main feature: a map. However, this visualisation is not exclusively global and if the user wants to zoom into smaller areas, we haven't prevented further exploration. We believe this makes the map a platform for research and further insight generation for the user (Roth, 2013).

Methodology:
A Leaflet map with a MapBox tile layer forms the basis for this visualisation and this is handled by the *Leaflet.js* library. The *Leaflet-Easy-Button.js* plugin is used for the buttons where each contains two states ('checked' and 'unchecked') depending on whether button has been clicked. In the code, the buttons push an item relating to their industrial grouping ('High', 'Medium', 'Low') to global array which then controls which subset of data is loaded onto the map layer.

Each of these industrial group layers contains a collection of Leaflet circle markers relating the coordinates of Air Quality monitoring station, from the latest version of data received from API (see 3.1.2). A leaflet popup is then bound to each marker containing information about station, the most latest pollutant record and line plot of the historical records from the last X amount of hours. To improve performance of this visualisation, no popup initially contains a historical plot, instead a local GET request is fired when clicking a popup which retrieves information from locally cached data and then generates the plots 'on-the-go'. The plot itself is produced using *D3.js.*

Styling:
For this world-scale visualisation, a *Mapbox* tileLayer is in a web Mercator projection (ESPG: 3857) is used, meaning that the planet's higher latitudes are emphasised versus the lower latitudes (Meirelles, 2013). This decision stems from the limitation of the Mapbox not supporting alternative projections and thus this hinders this visualisations impression on the user and our second key objective: to evaluate the *extent* of monitoring across the planet (see section 1).

In this map, air quality monitoring stations are represented as circle markers and both circle radius and colour scheme to indicate the value of AQI. We choose circle markers and adjust their opacity to 75%, instead of using full opaque 'sign-post' markers used by World Air Quality Index Project Network to project the same data (see WAQIP, 2019), as this allows the user to view through markers and see the extent of spatially clustering of the monitoring stations (Kirk, 2017). The colour scheme for these markers (green through red to purple) representing good through bad to hazardous is used to visualise air quality and directly stems from the world-standard color scheme used to represent levels of danger from air quality from the World Air Quality Index Project (WAQIP, 2019).

Illustrating a dataset containing over a thousand objects, this map is equipped with several interactive functions to encourage users to find their own insight with this exploratory visualisation (Kirk, 2017). However, we have tried to reduce the complexity of the visualisation as much as possible to not distract the user from its overall purpose (Roth, 2013).

3.2.3 Regional-Level Animated Air Quality Monitoring Visualisation

Why visualise air pollution at this scale?:
Apart from visualising air quality data statically at a global level, we are also interested in the changing dynamics of air quality at the regional level and across time. It holds that, spatial-temporal dynamics of air pollution change when we adjust both the spatial and temporal units that we or the end-users are examining, a common example of the 'modifiable areal unit problem' (MAUP) and 'modifiable temporal unit problem' (MTUP) (Openshaw, 1984; Cheng & Adepeju, 2014).

Our basic assumption is that, considering that most of the industrial production activities concentrate within daytime, the air quality will be worse during daytime and then improve again during the nighttime. To verify this assumption, an iterative animation visualisation is selected as an appropriate method to observe these spatial-temporal dynamics. This visualisation thus serves as a communication metric where the user can visualising the latest 24-hour historical data and discover any hidden patterns of air quality changes.

Animation is specifically chosen at the region scale, as opposed to the global or street level, as industrial heavy regions or large cities have the ability to produce large quantities of air pollutants that can be (invisibly) carried across national boundaries (Tilt, 2019). At the global scale this type of animated information is too coarse for the user to interpret and at the city scale (i.e. city) too specific. Over time the animation may display temporally lagged trends. Moreover, it is often at the regional level, as opposed to global-level, where common policy on air pollution is shared i.e. in the EU (Kutlar *et al.*, 2017).

Methods & Design
This animation visualisation involves the techniques of *MapboxGL.js, p5.js* and *Mappa.js*. *MapboxGL.js* is a JavaScript library that uses WebGL to render interactive maps from vector tiles and Mapbox styles, which is used to create the basemap of this visualisation section. *p5.js* is a JS client-side library for creating graphic and interactive experiences. To integrate *MapboxGL.js* and *p5.js* the Mappa.js library is used. Specifically, this facilitates communication between the canvas element and existing map layer through an API. *Mappa.js* provides complete access to the base map library that its original properties methods can be used on the canvas of p5.js, which makes the new combinations of visualisation techniques much easier. For example, this visualisation combines the advantages of p5 animation and the "flyto" method of Mapbox Gl as follows:

First, the 24-hour historical data were pre-loaded in the preload() function of p5.js;
Second, a Mapbox GL basemap was created through Mappa.js instance in the setup() function of p5.js.
Then, a for loop was called within the continuously executing draw() function, the key foundation to create the iterating animation, to run through the historical data and plot them on the basemap. Not only the size and color of the markers on monitoring station would change according to the AQI value, but also the corresponding collections time would be shown on the upper-left corner of the window;

Additionally, several "flyto" HTML buttons were created and shown on the lower-left corner, directing to the world view and several continent regions with meaningful amount of monitoring stations, including Europe, North America, South Asia and East Asia. When the buttons were clicked, the corresponding "flyto" method of Mapbox GL JS would be called to change the central coordinates and zoom level of the map. Through clicking these "flyto" buttons, the users can switch their view to their interest regions for closer observation. Still, the users are able to zoom in wherever they like with their mouses at the same time.

3.2.4 Street-Level Air Quality Visualisation

Why visualise air pollution at this scale:
The last visualisation section include a 2.5D map illustrating the estimated pollutant concentration at the street level. When focusing in on a city and its streets more detail is provided about the spatial distribution of pollution, than available at the global or regional scale. A 1996 study by Koussoulakou finds that the multidimensional nature of air pollution becomes relevant when the scope of visualisation is within urban areas.

Specific dynamics governing pollutants emitted into the air (i.e. by a city's vehicles) mean that pollution often persists and concentrate along certain streets, owing to factors such as proximity or height of buildings, amount of traffic and presence of trees (Pugh et al., 2012). This phenomenon is known as the 'Urban Canyon effect' (ibid.). Understanding where this occurs street-by-street is important for pedestrians, especially as the 'street' is a location in which groups are primarily exposed to air pollutants. This form of visualisation is important, as a result, as it serves as platform to visualise various harmful pollutants at the street level and can provide insight to how air pollution affects people's life (i.e. which routes around London they should avoid walking along).

Methods:

The data applied in this part is from the London Air Initiative (London Air, 2019) and relates to atmospheric emission modelling data, which is mentioned in section 3.1.3. Mapbox GL JS, Javascript and HTML are the key programming languages applied in this part.

The Mapbox fill-extrusion style enable this map to create 2.5D model of a city from 2D maps. The extruded building layer, a default Mapbox layer derived from OpenStreetMap, illustrates the basic 3D structure of central London and creates a visual model of real London in order to help audience navigate it. Meanwhile, the shapefile of the annual emission of each pollutant is reformed both in colours and heights. Each spatial unit is a 20m x 20m grid which is extruded according to the amount of emission of one category of pollutant in a year. Further identification of the concentration level is also represented by colouring each data feature by the rule in which green represents low emission, orange is for medium and purple refers to high.

Employing the *map.addLayer* function of Mapbox GL JS and Javascript, the user is given the ability to the switch the pollutant layer between different contaminants and a total of 3 distinct years. A function gets inputs from buttons of pollutants and time, and it calls the layer to display according to inputted year and contaminant name. With this function, the audience can explore all 15 layers by clicking buttons of categories and time. Two buttons that trigger the view changes are linked to *map.flyTo* functions that customise viewpoints, so the audience can explore the map from various viewpoints.

Apart from the data, text is added to pollutant buttons as dropdown contents to delivery extra information about the contaminants, again the user can explore this feature if they desire. Three locations which have performed especially poorly in terms of NO2 concentrations Euston Road, Strand and Oxford Street) are also identified by points and text, because these places are in list of worst air quality in London and even Europe (BBC, 2016; Taylor,2019). These points have been manually created on Geojson.io.

Overall, this map presents a detailed and visible air quality scenario through time and space in central London. A spatial pattern of pollutant can be directly observed from the map: NOX and PM are significantly concentrated around main roads, and the roads with rather higher concentration than others, such as The Strand and Oxford Street. These road, along with a few others in central London are known to share the characteristic which create urban canyons (Pugh et al., 2012).

3.2.5 Narrative and Textual Information

A picture of haze over the River Thames forms the backdrop to the homepage. Once the ominous “Begin” button is clicked we direct the viewer’s thoughts towards air and this begins the distributed narrative of the project – one that may be different for each user(Walker, 2004). The next stage provides them with a real narrative, a journey back in time through air pollution incidents and legal or artistic reactions to them. We hope that this brief history, sets the scene and further engages user attention, helping them relate to the graphical pages that follow.

For the first visualisation, data is displayed on a spinning globe, that does not require any user interaction. The user can toggle between these the two types of data or hover over countries they are interested in, so the user can decide what the visual shows. Displaying data in this globe format is not yet common on websites so we expect the user’s to engage in exploratory actions (Kim & Maher, 2008).

For the second visualisation, the choice of dividing the data into three groupings is used as the driving narrative for the second map, which leads with the tagline: “make the invisible, visible”. As aforementioned, the user is put through a process of physically ‘lighting-up’ the world at this scale. This is in contrast to the next regional map, which uses the same data in a different format. In this, the third visualisation the user is instead bombarded with the same data but it is animated, highlighting the rapidly changing dynamics of the air quality index (AQI) and its large range of values.

Finally, the city-level visualisation displays different years of different pollutant levels across a section of inner-London. This visualisation pushes the user to be more exploratory than prior visualisations and in way this is the most relevant to most users of the website as it is the street where people interact with pollution in space (Pugh *et al.*, 2015).

3.3 Design

The primary goal of data visualisation is facilitating users’ understanding, and data visualisation includes all designs decisions made for visual communication (Kirk, 2016). To achieve the aim to build a platform for visualising the impact and current state of monitoring of air quality and bring this topic to the public's attention. As such the design decisions of the website are built around two key principles:

1. Achieving high efficiency of data dissemination;
2. Improving the user experience and interaction with the project.

Styling
All four visualisation parts are displayed within webpage in format of full-screen slide. A navigation bar with icons is placed to the top-right corner to bring users to contents they are interested in and to notice users which page is viewed.

Each part of visualisation follows the same format of design to ensure the consistency of website style. The background colour for the whole website is set to dark grey while the data displayed within the visualisations and figures use schemes with bright colors and high saturation to emphasise the data (Meirelles, 2013). Moreover, it is hoped that this reduces the disturbance of information in the base maps, which should be of secondary importance to the viewer compared with the information they are trying to disseminated (Kirk, 2017).

Every visualisation contains two types of distinguishing features i.e. colours and styles such as radius size (on the *global* and *regional* visualisations), line-style (*global health impact*) and extrusion height (*street-level*). Except the globe view visualisation, all the maps display air quality and air pollution data. The same colour scheme is applied to every air quality layer to ensure the consistency of style in accordance to the WAQI (2019). The variation of hue and lightness make the deviation between air quality across space distinguishable, and this could be a safe strategy of devising for audience from all groups, especially for colour-blind people (Meirelles, 2013).

Layout
All pages include maps are shown in full-screen size, while necessary elements such as titles, legends, buttons, charts and explanations are placed as an overlay towards the left of each visualisation. The extra information that compensates each scale is put in popups of certain data features, so users can show and hide it according to their needs.

Interactivity
The data visualisation practices we use emphasise a combination of exploratory and analytical functionality, which promote an advanced level of interaction (Smith, 2016). Indeed, we aim to provide conditions which support insight on patterns of air pollutants and its monitoring (Roth, 2013). Equally, the interactivity of each visualisation ensure the audiences can have highly flexible and individual user experience. Finally, buttons are used throughout the project which trigger the loading (or hiding) of extra data layers – for any interested users.

3.4 Technology & Web Hosting

A range of both JavaScript & Python libraries have been used in this project and these are stated along with their required versions in Table A1. For the online hosting of this projects files, Netlify was preferred over GitHub Pages and this decision is based purely on improving the performance.

4 Discussion & Conclusion

4.1 Evaluating Objectives

By combining visualisations across four spatial scales and using design techniques appropriate to each scale, it could be argued that, the visualisations somewhat achieve our first key objective of providing an informative platform for the dissemination of the impact of air pollution. For our second objective, the global and regional-scale maps highlight the physical locations of monitoring stations which goes some way towards improving the 'visibility' of air quality, our second objective. However, we do not develop further on this and as such the effectiveness of these visualisations depend on user insight generation. We propose that the final objective is most confidently achieved as each visualisation contains spatial and temporal information and encourage the user to experience these dynamics through interactive features.

4.2 Challenges & Solutions

With ambitions to create a visualisation website providing advanced insights, we made attempts to include ample sources of data and complex web technologies in early stage. Nevertheless, we encountered difficulties in accessing data and applying visualising technique we originally sought – as such we made a few key compromises.

One example was with the 3D globe visualisation, it was originally hoped to map a surface layer relating to the air quality across the entire globe (similar to Beccario, 2019). The raster form of the Copernicus data is ideal for the global visualisation section, because this dataset includes 30 km square patches of pollutants (Copernicus, *no date*). However, using information from the copernicus satellite was challenging as this included (locally) hosting large amounts of data.

Another key challenge was with the street-level visualisation that appears as the final visualisation in the project. It was hoped that we could integrate a video simulating a 3D street in Unity, but our initial results proved unsatisfactory. It was instead decided to produce a 2.5D version in Mapbox to complete the website in the minimal spatial scale.

An initial design decision to use a webpage that scrolled up and down incompletely from one section to the next was dropped because the transition produced an unattractive jump from one background to another. However, this plan caused discontinuity when users switch between each sections, which affect the overall representation of the website. As a result we chose to the *SmartScroll.js* library.

4.3 Limitations:

Data Limitations

Data reliability hinders each aspect of this project's visualisations. For example, we do not have access to any calculations or models which are used to produce the data used in all the visualisations. In addition, our visualisations do not contain any communication of the level of this uncertainty associated with any of the data, something which is essential when presenting scientific information as the project does (Winkler, 2015). As such, this form of epistemic uncertainty hinders any conclusions that can be drawn from visualisations.

Methodological Limitations

For the second visualisation all stations were divided into three groups according to a country's industrial sector as a % of GDP. Although this serves to divide groups into categories where we expect the most industrial nations to pollute more. This method was mainly used to solve the slow loading speeds associated with large leaflet marker layers. It is appreciated that this may lead to a user misunderstanding of the purpose of this visualisation.

The other limitation of methods is the absence of an automatic data refreshing function. In the initial stage, the world map is planned to present real-time data that is updated every hour. But hosting web server to automatically update API data need is not free. For the consideration of financial cost, we chose not to use web server to keep data on our website updating in real-time.

Technology Limitations

Ultimately this project suffered from a slow performance, as it takes around 10 seconds to load the project's content. General website performance was something we struggled with throughout the project development, owing to the large geoJSON files that were needed to create some visualisations. One such way that performance was improved was to store or cache the required data in local memory to prevent it from being continually reloaded.

Finally, there are a few bugs and performance issues that are known about this project. For example the smart-scrolling of website is disabled when clicking on a visualisation, to return to the smart scrolling click on the icons at the side. Rending the globe can massively slow down the map, this is because the d3 canvas is not deleted correctly, simply reload the webpage to improve performance. Besides, when using the buttons on the global health visualisation to change between death rate and exposure, the markers only load when the globe is spinning.

4.3 Further Improvements

A major improvement for the visualisation using data regarding monitoring station from the WAQIP (2019) is to include a mechanism that is constantly updates these map with live measurements. This would make the visualisation more dynamic and thus relevant.

Another improvement is to optimise the performance of the project's website and its data. This could be achieved by having a dedicated server to host the files and data online such as a cloud computing technology.

To evaluate our second objective, having a count of available stations per country would provide more context about the state of monitoring. Finally, if the regional animation visualisation contained more complete historical data (i.e. at a seasonal scale), this website would provide more insights into the spatio-temporal dynamics of air pollutants at different scales.

4.4 Conclusions

The choice of using air quality for the topic of “Invisible Cities” provided plenty of interesting avenues to explore and we felt it important to present this subject in the most informative the appropriate light. The nature of our chosen subject for the visualisation meant it was likely we would get caught up in the distributed narrative that we created for our viewers. At various points each of us experienced heightened environmental awareness and hoped that websites similar to ours would be educating and changing attitudes. Our individual reactions are part of an established individual psychological process in global environmental change and to be expected. It only remains to record that the way we were affected in actually creating these visualisations about is how we hope visitors will be affected too.

5. References

Atterer, R., Wnuk, M., & Schmidt, A. (2006). Knowing the user’s every move. *Proceedings of the 15th International Conference on World Wide Web - WWW ’06*. doi:10.1145/1135777.1135811

BBC (2016), *Reality Check: Is Oxford Street the world's most polluted?* [press release] 19 August 2016. Available at: https://www.bbc.co.uk/news/uk-politics-37131138 [Accessed 20 May 2019].

Beccario, C. (2019) Earth [*Online*]. Available at: https://earth.nullschool.net/ [Accessed 20th May 2019].

Brauer, M., Amann, M., Burnett, R. T., Cohen, A., Dentener, F., Ezzati, M., … Thurston, G. D. (2012). Exposure Assessment for Estimation of the Global Burden of Disease Attributable to Outdoor Air Pollution. *Environmental Science & Technology*, 46(2), 652–660.

Cheng, T., & Adepeju, M. (2014). Modifiable temporal unit problem (MTUP) and its effect on space-time cluster detection. *PloS one*, *9*(6), e100465.

Chipperfield, M. P. *et al.* (2015) Quantifying the ozone and ultraviolet benefits already achieved by the Montreal Protocol. *Nature Communications.* 6:7233

Copernicus Sentinel-5P ozone boosts daily forecasts / Sentinel-5P / Copernicus / Observing the Earth / Our Activities / ESA (Copernicus) (no date). [*Online*]. Available at: https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-5P/Copernicus_Sentinel-5P_ozone_boosts_daily_forecasts [Accessed: 15 May 2019]

Gao, J., Yuan, Z., Liu, X., Xia, X., Huang, X., & Dong, Z. (2016). Improving air pollution control policy in China-A perspective based on cost-benefit analysis. Science of the Total Environment, 543, 307–314.

Greater London Authority (GLA) (2015). London Atmospheric Emissions Inventory (LAEI) Concentration Maps. [*Online*] GLA. Available through: London Datastore https://data.london.gov.uk/dataset/laei-2008-concentration-maps?resource=159b2b40-d881-4ee5-88e2-09cbf0e1c012 [Accessed 10 May 2019].

Guerreiro, González Ortiz, A., de Leeuw, F., Viana, M., Colette, A. (2019). Air quality in Europe — 2018 report. *European Environment Agency.*

Kelly, F. J., & Fussell, J. C. (2015). Air pollution and public health: emerging hazards and improved understanding of risk. *Environmental Geochemistry and Health*, 37(4), 631–649.

Kim, M. J., & Maher, M. L. (2008). The impact of tangible user interfaces on spatial cognition during collaborative design. *Design Studies*, 29(3), 222–253.

Kirk, A (2016) *Data Visualisation: A Handbook for Data Driven Design*. 1st ed. SAGE Publishing Limited: London.

Kutlar J., M., Eeftens, M., Gintowt, E., Kappeler, R., & Künzli, N. (2017). Time to harmonize national ambient air quality standards. *International Journal of Public Health*, 62(4), 453–462.

London Air (2019) London Air Annual Pollution. [*Data*]. Available at: https://www.londonair.org.uk/LondonAir/Default.aspx [Accessed 16th May 2019].

Meirelles, I. (2013) *Design for Information: An introduction to the histories, theories, and best practices behind effective information visualisation.* 1st ed. Rockport Publishers: Beverly.

Monks, P. S., Granier, C., Fuzzi, S., Stohl, A., Williams, M. L., Akimoto, H., … von Glasow, R. (2009). Atmospheric composition change - global and regional air quality. Atmospheric Environment, 43(33), 5268–5350.

Openshaw, S. (1984) Ecological fallacies and the analysis of areal census data. *Environment and Planning*, 16, 17–31.

Park, C. & Allaby, M. (2017), 'Air Pollution'. *A Dictionary of Environment and Conservation, 3rd* ed. [Online]. Available at: https://www.oxfordreference.com/view/10.1093/oi/authority.20110810104326671 [Accessed 20th May 2019].

Pattinson, W., Longley, I., & Kingham, S. (2014). Using mobile monitoring to visualise diurnal variation of traffic pollutants across two near-highway neighbourhoods. *Atmospheric Environment*, 94, 782–792. doi:10.1016/j.atmosenv.2014.06.007

Pugh, T. A. M., MacKenzie, A. R., Whyatt, J. D., & Hewitt, C. N. (2012). Effectiveness of Green Infrastructure for Improvement of Air Quality in Urban Street Canyons.

Rai, A. C. *et al.* (2017) 'End-user perspective of low-cost sensors for outdoor air pollution monitoring', *Science of The Total Environment*, 607–608, pp. 691–705.

Ritchie, H. & Roser, M. (2017) Air Pollution [*Online*]. Available at: https://ourworldindata.org/air-pollution [Accessed 20th May 2019].

Roth, R. E. (2013). Interactive maps: What we know and what we need to know. Journal of Spatial Information Science, 6(6), 59–115.

Schwartz, J., & Dockery, D. W. (1992). Increased Mortality in Philadelphia Associated with Daily Air Pollution Concentrations. *American Review of Respiratory Disease*, 145(3), 600–604.

Smith, D. A. (2016). Visualising world population density as an interactive multi-scale map using the global human settlement population layer the global human settlement population layer. *Journal of Maps,* 57(2016), 106-117.

Snyder, E.G., Watkins, T.H., Solomon P.A., Thoma, E.D., Williams, R.W., Hagler, G. S. W., Shelow, D., Hindin, D.A., Kilaru, V.J., and Preuss, P.W. (2013) The Changing Paradigm of Air Pollution Monitoring *Environmental Science & Technology 47* (20), 11369-11377.

Taylor, J. (2019), Worst air pollution in London: Earl's Court, Camden and Southwark top the list of 500 places breaching air quality limits, *Evening Standard,* 27 February 2019 [Online].. Available at: https://www.standard.co.uk/futurelondon/cleanair/air-pollution-toxic-air-pollution-crisis-a4077936.html [Accessed 20 May 2019].

Tilt, B. (2019). China's air pollution crisis: Science and policy perspectives. Environmental Science and Policy, 92(November 2018), 275–280.

Vogel, J. (2017) Create a 3D globe with D3.js, Jorin's Logbook. [Online]. Available at: https://jorin.me/d3-canvas-globe-hover/ [Accessed 12th May 2019].

Walker, J., 2004. Distributed narrative: Telling stories across networks. *Internet Research Annual*, *3*, pp.91-104.

Winkler, R. L. (2015). The importance of communicating uncertainties in forecasts. *Risk Analysis*, *35*(3), 349–353.

WHO (2005). Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide. *World Health Organization*, Geneva, Switzerland.

WHO (2016). Ambient air pollution: a global assessment of exposure and burden of disease. World Health Organization. http://www.who.int/iris/handle/10665/250141 [accessed 26th May 2019]

WHO (2018) Ambient (outdoor) air quality and health [*Online*]. Available at: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health [Accessed 20th May 2019].

World Air Quality Index Project (WAQIP) (2019) World's Air Pollution: Real-time Air Quality Index [*Online*]. Available at: https://waqi.info/ [Accessed 10th May 2019].

World Bank (2017) Industry (including construction), value added (% of GDP). [Data]. Available at: https://data.worldbank.org/indicator/nv.ind.totl.zs [Accessed 18th May 2019].

Appendices

Appendix 1 – JavaScript & Python Libraries & Versions

Table A1. Table containing all external libraries needed for JavaScript and Python.

JavaScript		
Library	**Versions**	**Use in project**
Bootstrap 4 Material Design (Material-Kit.js)	2.0.5	General Styling
Bootstrap-Slider.js	10.6.1	Slider Styling
Bootstrap-Awesome-Font	4.7.0	Fonts and Icons
D3.js	v4	General Plotting
D3-Simple-Slider.js	0.1.2	Slider controls
D3-TopoJSON	1.0.0	Geographic plotting
Versor.js	0.0.4	Globe projection equations
Plotly.js	1.4.7	Plotting
JStat.js	1.7.1	Statistical functions
Leaflet.js	1.4.0	Mapping
Leaflet-Easy-Buttons.js	2.0.0	Buttons for World Map
MapboxGL.js	0.54.0	Mapping
JQuery	3.4.0	DOM Manipulation
SmartScroll.js	2.5.0	Scrolling of index web page
Lethargy.js	1.0.7	Scrolling of index web page
P5.js	0.7.3	Animation
P5-Sound.js	0.7.3	Animation
Python 3.6		
requests	2.20.0	Data Collection
datetime	in-built	Data Collection by schedule
apscheduler	3.2.0	Data Collection by schedule

pandas	0.24.2	Data Handling
numpy	1.16.2	Data Handling
bs4	4.7.1	Data Handling
re	2.2.1	Data Handling
string	in-built	Data Handling
difflib	in-built	Data Handling
shapely	1.6.4.post1	Data Handling
geopandas	0.4.1	Data Handling
json	2.0.9	Data Storage
geojson	2.4.1	Data Storage
PyGithub	1.43.7	Data Storage