University of Exeter Business School

Assignment Cover Sheet	
Candidate Number	027029
Module Code	BEAM031
Module Name	Analytics and Visualisation for Managers and Consultants
Assignment Title	Pollutions effect on the English people's health

Within the Business School we support the responsible and ethical use of GenAI tools, and we seek to develop your ability to use these tools to help you study and learn. An important part of this process is being transparent about how you have used GenAI tools during the preparation of your assignments.

The below declaration is intended to guide transparency in the use of GenAI tools, and to assist you in ensuring appropriate referencing of those tools within your work.

The following GenAI tools have been used in the production of this work:

[please specify] ……..

☐ *I have used GenAI tools for brainstorming ideas.*

☐ *I have used GenAI tools to assist with research or gathering information.*

☐ *I have used GenAI tools to help me understand key theories and concepts.*

☐ *I have used GenAI tools to identify trends and themes as part of my data analysis.*

☐ *I have used GenAI tools to suggest a plan or structure of my assessment.*

☐ *I have used AI tools to give me feedback on a draft.*

☐ *I have used GenAI tool to generate images, figures or diagrams.*

☐ *I have used AI tools to proofread and correct grammar or spelling errors.*

☐ *I have used AI tools to generate citations or references.*

☒ *Other [Help with programming issues]*

☒ *I declare that I have referenced use of GenAI tools and outputs within my assessment in line with the University referencing guidelines.*

Table of Contents

1. Introduction

This paper walks through the process of creating a dashboard focused on pollutions effect on people's health in the different regions of England. First the data collection and process are introduced, proceeded by the design process. Afterwards the dashboard is presented, with arguments for the final decisions. Lastly, a discussion and conclusion are presented, finalising the paper.

2. Data collection

The data was collected through available sources from the UK government. The pollution data was collected from the GOV.UK, where the health data was collected through the Department od Health & Social Care.

2.1. Pollution data

There were four types of pollution data available: Ozone, Nitrogen Dioxide, Particulate Matter 2.5, Particulate Matter 10. Only three of them were used, as the Particulate Matter 2.5 dataset only consisted of data from 2009. (Department for Environment, Food & Rural Affairs, 2023)

As the data was formatted as a report, extensive cleaning needed to be done. As the areas given in the dataset were not the regions of England, these needed to be converted. This was done by creating a separate document (Area_to_Region.xlsx), where the regions of the different areas were found and connected. This document was then used to run through all the rows of the individual datasets, replacing the area with the corresponding region. Additionally, the data from before 2001 were removed, as to match with the health data. Lastly, all datasets were merged for easier use, and the percentage of each pollution type, given the total sum of pollution, was calculated.

2.2. Health data

The health data chosen for the dashboard were mainly divided into two categories: mortality rate, all ages; mortality rate, under 75. Here data was collected for cardiovascular disease, lung cancer, respiratory disease, and chronic respiratory disease, resulting in eight datasets. (Office for Health Improvement & Disparities, 2024)

The health data was already cleaned and in csv-format, making it easy to work with. As the data included more information than necessary, they were greatly sorted, only including the year, area code, area name, age, sex, and value. The datasets were merged, whereafter the area code was changed through a function, to ensure it would match up with the map data later on. Lastly, the age range between <75 and all ages were calculated, to give a broader insight into who's is affected by the pollution.

3. Design process

The design process is created based on CAST, presented by Malik, West, and Sykes in the book *Stories that move mountains: storytelling and visual design for persuasive presentations*. (Malik, West, & Sykes, 2012)

3.1. Content

3.1.1. Why

Pollution affects people's health, and increases the chance of getting lung cancer, respiratory disease, and heart disease (How can air pollution cause cancer?, 2024). Visualizing the link between pollution and health in different regions can showcase the true effects, and support arguments as to why current regulations needs to be changed.

3.1.2. What

Regulations can be made into pollution control, as well as energy generation, lessening the proportion of pollution, and thereby lessening the health effects of it.

3.1.3. How

By showing the development between pollution and health in the last 20 years, the effect can be seen. Additionally, with the option to focus on the different types of pollution and their potential effect, more direct action can be taken.

3.1.4. What if

Changing regulations that include pollution will affect many companies, potentially crating a backlash. However, the effort of the dashboard is still important, as the health of the population in the end will affect the workforce of the country negatively, leaving them temporarily or permanently out.

3.2. Audience

3.2.1. Who

To make any changes, the audience needs to be someone that has the ability to change regulations in England. Because of this three general people has been selected:

- Wes Streeting MP: Secretary of State for Health and Social Care (The Rt Hon Wes Streeting MP, n.d.)
- Ed Miliband MP: Secretary of State for Energy Security and Net Zero (The Rt Hon Ed Miliband MP, n.d.)
- Louise Haigh MP: Secretary of State for Transport (The Rt Hon Louise Haigh MP, n.d.)

These secretaries cover different aspects of the goal - Wes Streeting would be interested in the affects on health, as he is responsible for the health segment. Ed Miliband are the secretary of energy and net zero, adding the energy aspect from the visualisation. Lastly, Louise Haigh who is the secretary of transportation, would be interested as part of the air pollution created are done through cars and other forms of transportation.

3.2.2. Learning and decision styles

Given that the audience is working within their individual fields, the focus would not be on learning, but instead on creating decision points and a foundation for change. The audience would need to gain a specific understanding of the relationship between pollution and health issues, for them to make a decision. Here it would be necessary to dig deeper into what type of pollution affects the populations health the most, as this would make it easier to create specific regulations.

3.3. Story

3.3.1. Structure

The structure focuses on the development of disease and pollution over time. Here key numbers are presented immediately, as for the audience to gain a quick overview of the data. Additionally, it is possible to change the data shown, and dig deeper into the story through time, with the map.

3.3.2. Character

The audience is already focused on the general population of England, given their position, but digging deeper into who is affected by pollution – looking into early sickness or death – would illustrate the real and grave effect of pollution.

3.3.3. Sense of urgency

With pollution greatly affecting health, and resulting in early death in some cases, the sense of urgency focuses on saving the population, and sparing them of sickness that could lead them to an early death.

3.3.4. Delivery plan

The delivery of this dashboard would be through an introduction meeting, where it is presented and the purpose of it is explained. The dashboard will afterwards be delivered to each audience member, who will get the opportunity to explore the different elements of the dashboard. Lastly, an additional meeting will be held where a discussion of each finding will take place.

3.4. Tell

3.4.1. Design

The design was created through a rough brainstorming session, where post it notes were used to easily move the different graphs around and test different design ideas (see Figure 1). Here two metrics from each of the three health categories, had their individual graph, as well as the pollutions, that was collected within one graph. A map would be at the right side of the dashboard, making it possible to select the individual region to focus on. Besides the pollution graph, a simple table with the pollution metric for the given year, was placed, with a slider in the bottom selecting the year.

Figure 1: Brainstorm

To ensure that this design was readable, it was recreated in PowerPoint. Here it was clear that the main point of comparing the health and pollution was not obvious, and the overall design was too cluttered and confusing to gain a quick understanding of the goal. Instead, a new sketch was created through PowerPoint, noting all the difficulties the original design had, and improving on them (see Figure 2).

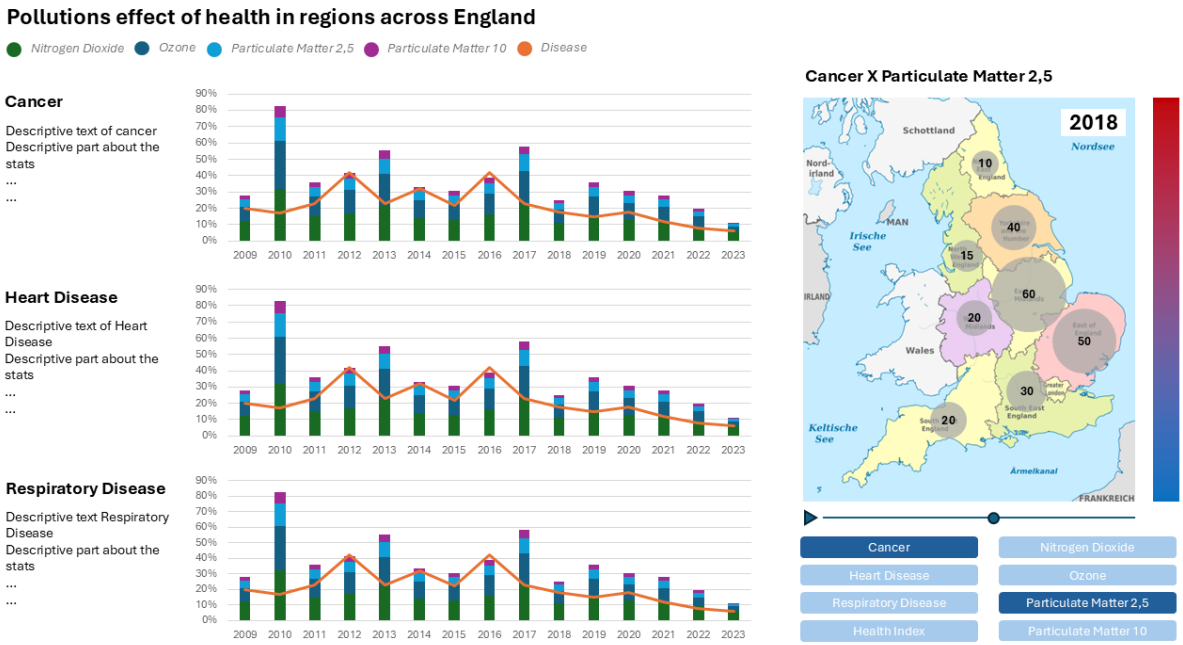


Figure 2: Clean sketch

Here the disease graphs have been reduced to three in total, one for each main health metric, where the pollution is represented in each graph as a stacked bar chart, with the trend of the health being a line chart going on top of it. To minimise clutter, the legend has been added under the title, with only one point for the disease, as it is named by each graph. Additionally, a small summarising text has been added under each chart title, addressing the trend and pollution throughout the years. Lastly a map of the regions in England has been added to the right, that compares the different health metrics with the pollution in the area. The pollution is represented by the gradient in each region, and the bubbles indicates the health metrics. As there are a number of health metrics and pollution measurements, these can be selected below the map, as to still give a full insight into the change throughout the year. To give the map more appeal, it is made interactive, with the user being able to play through the time, as well as hovering over the regions to get more specific data.

3.4.2. Test

Because of the limited time and no clear target group to test on, this has not been included for the design phase. However, in case of possible design phase, professionals within both health and pollution would be consulted, to ensure the data is presented accurately. Afterwards, a target group would be introduced to the dashboard, to test if it is intuitive and easy to understand, without additional support.

4. Dashboard

4.1. General design

The dashboard is divided into five major elements. First, the boxes that create a general overview of what the dashboard shows. Second, the line and bar graph, describing the development of the different diseases and pollutions over the years. Third, the map over England, showing the level of pollution in the different regions, with the number of cases for the given disease. Fourth, the selection of variables, that allows the user to select the region, disease, pollution, and the age range. Lastly, is the selection of gender, that allows the user to select multiple genders or all to compare (see Figure 3).

All elements are grouped visually, by the law of proximity, and law of similarity (Yablonski, 2024). By ensuring that each pollution and disease have their own colour, they are connected to each other, even though they are not part of the same visualisation. Additionally, all of the text are boxed in with its relevant data or visualisation, creating a clear headline for each.

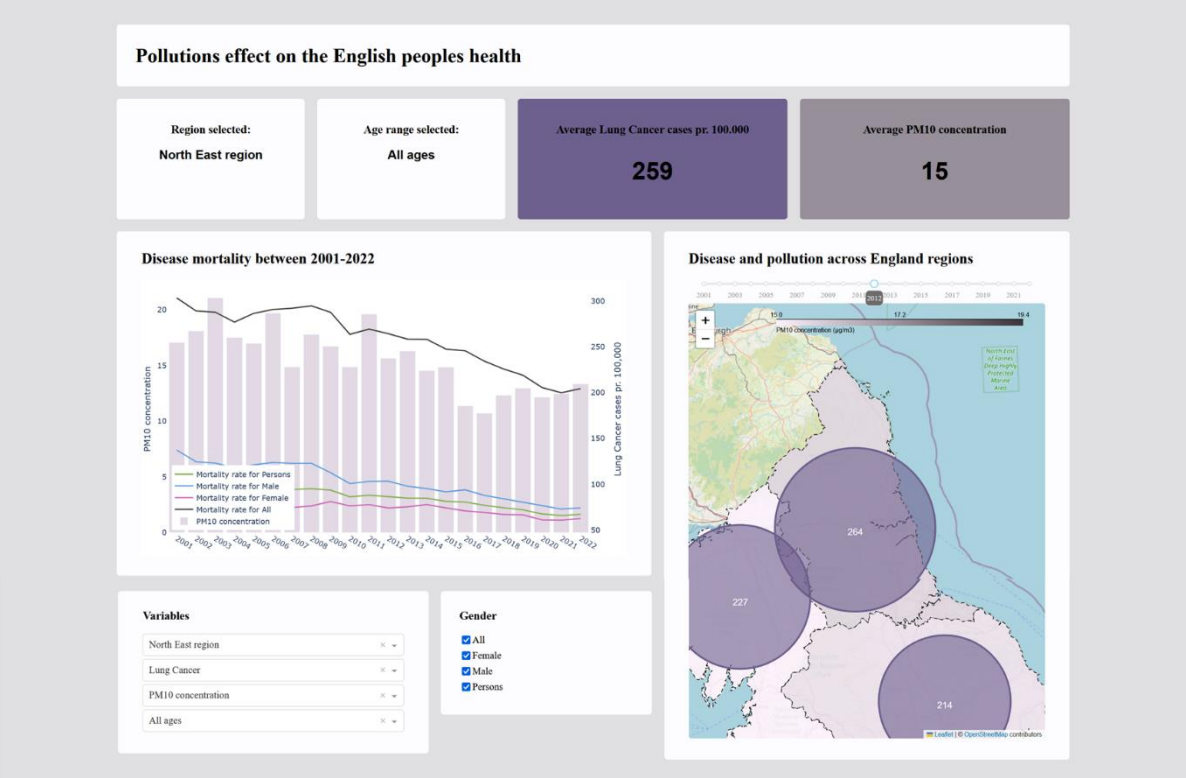
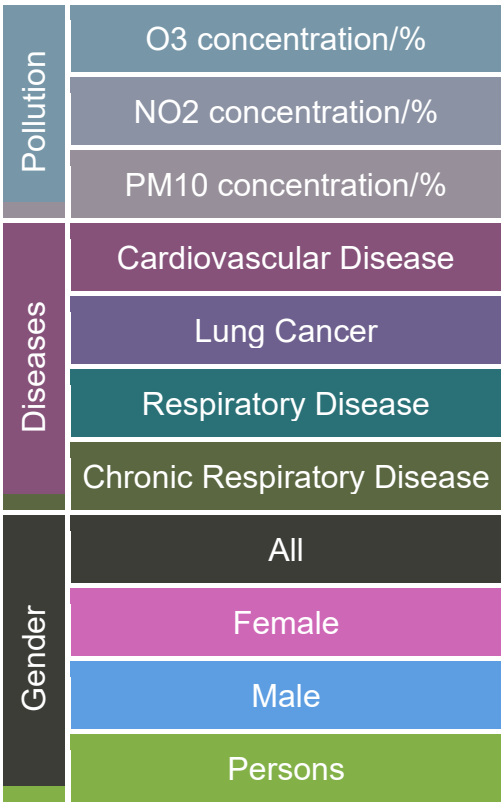


Figure 3: Final dashboard

4.2. Colour choices



A colour scheme has been chosen, to ensure clarity of the visualisations. The different types of pollution and diseases have each received a colour, that is used by both visualisations. As this dashboard is not meant for a broad audience, and the different types of pollution and diseases are not meant to be directly compared, there has not been a focus on using colourblind-friendly colours. Instead, they are used to indicate a change, to make the dashboard more responsive and user friendly. Additionally, colours for the genders have been chosen. Here the genders are represented by colours associated with it, with the “All” and “Persons” category being more neutral in its association with gender. These colours are also more vibrant, as to stand out better in the line graph, as they would otherwise clash with the pollution colours.

4.3. Selection properties

There are two types of selection properties: the drop-down menu, that allows the user to select one element at the time; or the checkboxes, that allow the user to select multiple elements at the time (see Figure 4). These selections carry the functionality of the dashboard, as it makes it possible for all visualisations to update directly, through the dash callback library.

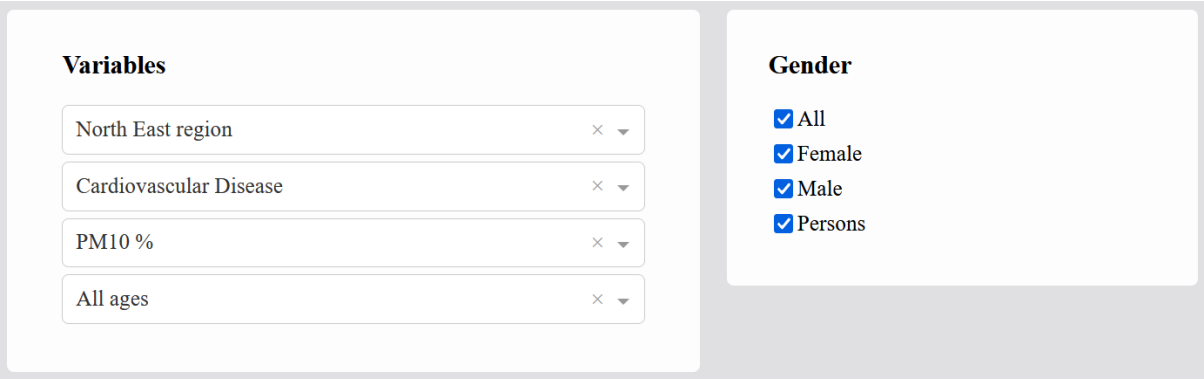


Figure 4: Selection properties

4.4. Info cards

The first element on the dashboard are the info cards. These cards show the region currently selected, the age range selected, the average of the given disease, and the average of the given pollution (see Figure 5). All these cards are dynamic, meaning that their text change given the input chosen by the user.

As both the disease and pollution cards represent an average, this is calculated based on the region and age selection, making it responsive to the user's input. Additionally, the colours for the disease and pollution cards are responsive as well, changing to the corresponding colour in the other visualisations.

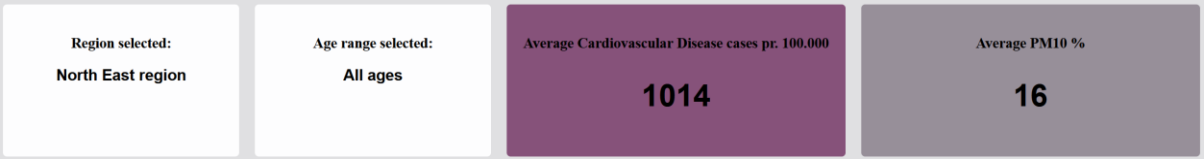


Figure 5: Info cards regarding selected elements

4.5. Line and bar chart

The line and bar chart consists of two general elements: the pollution and the disease selected. Here the disease is split up further into gender, with four types being available: All, Female, Male, and Persons. These are possible to select and de-select, for easy comparison (see Figure 6).

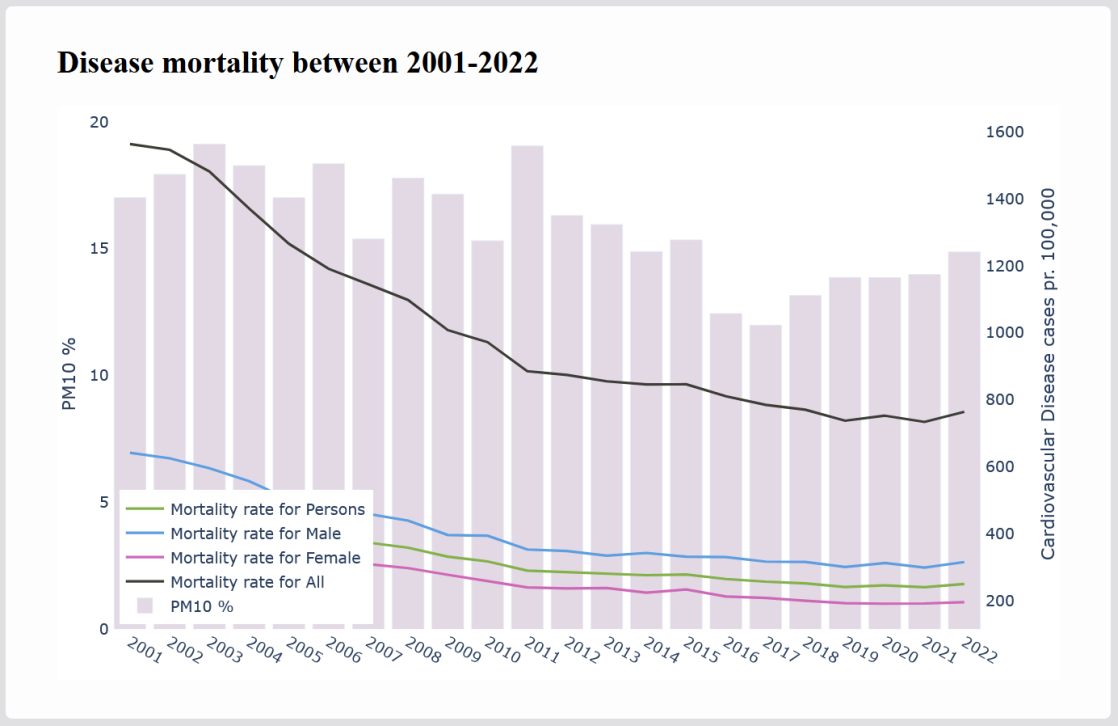


Figure 6: Line & barchart of chosen disease and pollution

4.6. Pollution and disease map

The map shows the chosen pollution and disease in different ways. The pollution is shown through a heatmap, where the disease is shown through a bubble map. Here the size of the bubbles is calculated given the max and minimum value of the disease, to show the difference between disease cases in the regions. Additionally, the number of cases pr. 100.000 is shown in the middle of the bubble, and with further information when hovering the mouse on top of it. This was done as the bubble sizes can be hard to distinguish, making the quick understanding of the visualization harder to grasp (see Figure 7).

Lastly a time slider has been added, making it possible for the user to select a year of interest, where the map will update afterwards.

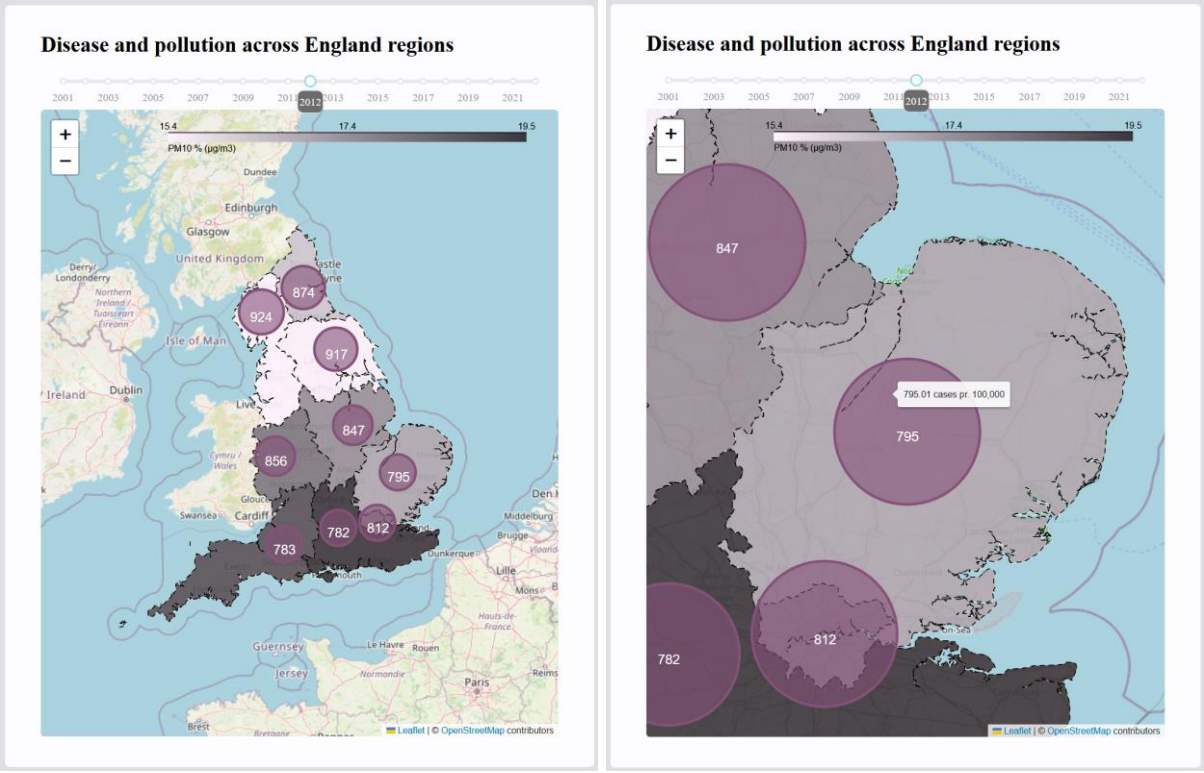


Figure 7: Heat & bubble map

4.7. Tests

While the dashboard was not tested on the target audience or equivalent, it has been shown to a few people, and responses were received. This resulted in an iterative process, where additional design elements were added or changed from the original design. Further knowledge was also gained regarding the options of visualisations and responsiveness, resulting in the design being changed throughout the process.

5. Discussion

The final dashboard and the original design vary in some key parts. The original design includes all the information up front, while excluding essential data, to limit the clutter. Additionally, it is very static, only presenting a timeline and an interactable point. Likewise, the colour scheme was not considered, resulting in the standard Microsoft colours being used (see Figure 2).

Comparatively, the final dashboard greatly reduces the elements introduced on the page, limiting the visualisations to two graphs and an area with information cards. The section for selecting the pollution and disease has been included, but instead of only affecting the map, it now affects the info cards and line chart, making the dashboard more engaging and simpler to look and interact with.

However, there is still room for improvement. The titles for each visualisation would benefit from being more reactive, further showing what is being presented. Additionally, the time scale would have been better as an animated timeline, truly presenting the changes over time. Lastly, the selection area could be improved, by replacing the dropdown menu with coloured buttons, making it more intuitive to work with.

6. Conclusion

In conclusion, the dashboard shows the pollution and health of the people of the England regions, from 2001 to 2021. Selecting the different types of diseases and pollution show a potential correlation, creating a potential fundament for regulation changes. Some changes and improvements could be made with further testing with an audience, and with collaboration with professionals within healthcare and pollution. However, as the dashboard is still within its prototyping phase, these developments and improvements are well within its limits.

7. References

BugBytes. (2021, October 14). *Building Choropleth Maps with Folium and Pandas (Python)*. Retrieved from YouTube: https://www.youtube.com/watch?v=TDlo7s4SZA8

cs95. (2021, October 07). *Turn Pandas Multi-Index into column*. Retrieved from Stack Overflow: https://stackoverflow.com/questions/20110170/turn-pandas-multi-index-into-column

Department for Environment, Food & Rural Affairs. (2023, April 27). *ENV02 - Air quality statistics.* Retrieved from GOV.UK: https://webarchive.nationalarchives.gov.uk/ukgwa/20230801200324/https:/www.gov.uk/government/statistical-data-sets/env02-air-quality-statistics

Folium. (n.d.). *Using GeoJson*. Retrieved from Folium: https://python-visualization.github.io/folium/latest/user_guide/geojson/geojson.html#Styling

Geeks for Geeks. (2024, November 29). *How to Select Rows from a Dataframe based on Column Values ?* Retrieved from Geeks for Geeks: https://www.geeksforgeeks.org/how-to-select-rows-from-a-dataframe-based-on-column-values/

How can air pollution cause cancer? (2024, February 9). Retrieved from Cancer Research UK: https://www.cancerresearchuk.org/about-cancer/causes-of-cancer/air-pollution-radon-gas-and-cancer/how-can-air-pollution-cause-cancer

Malik, N. A., West, M. D., & Sykes, M. R. (2012). CAST and the Visual Story Map. In *Stories that move mountains : storytelling and visual design for persuasive presentations* (pp. 12-21). Chichester, U.K: Wiley.

Office for Health Improvement & Disparities. (2024). *Public Health Profiles.* Retrieved from Department of Health & Social Care: https://fingertips.phe.org.uk

Plotly. (n.d.). *Legends in Python* . Retrieved from Plotly: https://plotly.com/python/legend/

Sambhi, P. (2020, October 18). *ValueError: You are trying to merge on object and int64 columns when use pandas merge*. Retrieved from Stack Overflow:

https://stackoverflow.com/questions/64385747/valueerror-you-are-trying-to-merge-on-object-and-int64-columns-when-use-pandas

The Rt Hon Ed Miliband MP. (n.d.). Retrieved November 14, 2024, from GOV.UK: https://www.gov.uk/government/people/ed-miliband

The Rt Hon Louise Haigh MP. (n.d.). Retrieved November 14, 2024, from GOV.UK: https://www.gov.uk/government/people/louise-haigh

The Rt Hon Wes Streeting MP. (n.d.). Retrieved November 14, 2024, from GOV.UK: https://www.gov.uk/government/people/wes-streeting

wjandrea. (2024, April 15). *How can I iterate over rows in a Pandas DataFrame?* Retrieved from Stack Overflow: https://stackoverflow.com/questions/16476924/how-can-i-iterate-over-rows-in-a-pandas-dataframe

Yablonski, J. (2024). *Laws of UX is a collection of best practices that designers can consider when building user interfaces.* Retrieved from Laws of UX: https://lawsofux.com/