6/22/2020

FIT5147 – Data Exploration and Visualization

Semester 1, 2020

Narrative Visualisation Project

Air Quality in India and its Causes - Report

Submitted by: Abhilash Kale [30254140]
Activity Number: 04-P1
Tutor Name: Farah Tasnuba Kabir

Table of Contents

Table of Figures

1. Introduction

Pollution, climate change, global warming and environmental crisis are the growing fears for everyone in today's world. Globally, air pollution has been one of the most dangerous in terms of health risks. The growth of industries and the rise in the numbers of vehicles on road have major contributions towards it. According to the World Health Organization, 9 out of 10 people breathe the air which comprises high amounts of pollutants, and unfortunately, about 7 million people die each year due to the exposure to polluted air (Osseiran & Lindmeier, 2018). India, the world's second most populous country, is highly endangered because of its rising air pollution. Moreover, a news article also suggests that India's air quality is even more lethal than the leaders in global population, China (Anand, 2017).

As the time has passed, the growth in industrialisation and modernisation has been exponential in India, resulting in a massive increase in the air pollution. A narrative visualisation of these pollutants – NO2, SO2, SPM, RSPM, resulting into a single valued Air Quality Index (AQI), and their causes through the different states in the country, will help us understand the story of its rising levels and the effects it might create, in India.

This report supports the narrative visualisation on India's ambient air quality data. The visualisation focuses on the major air pollutants and their levels which participate in polluting the air in India. The narrative visualisation considers the primary causes of air pollution in India like the industrial pollution, Diwali festival fireworks and vehicular pollution. The visualisation application aims to aid the target audience, i.e. the general public or the citizens of India and other countries, to learn the reasons and trends behind India's air quality, easily. The open data retrieved for this visualisation was previously wrangled and explored in regards with the proposed research questions and other discovered findings during the process.

Based on the exploration, the narrative visualisation is developed through R Shiny, to understand the findings with a visual and interactive experience for the users.

2. Design

Visualising data for novice users requires special efforts in order to make them easy to use and understandable. Achieving the high quality along with the visualisation's ability to convey information easily, is not a walk in the park.

As a developer, many challenges are faced while designing and creating a visualisation tool or application for the end-users. It is very important to interact with the users, recognize their requirements, design visual solutions, implement them in real-time and then finally evaluate them (Roberts, 2011). In order to achieve this seamlessly, a specific process needs to be followed. Sketching the designs on paper has been proven to be a helpful way of planning and searching solutions. Low-fidelity prototyping is one of the known approaches, which saves time, money and meets good solutions faster (Roberts, Headleand, & Ritsos, 2015). Thus, product design wants the developers to ensure a good quality design of the tool or application using easy, inexpensive and fast methods.

The Five Design Sheet (FdS) methodology enables developers to create the visualisation interfaces through low-fidelity approaches. The developers themselves can play the role of designers here for achieving the FdS. Ideas are sketched, planned while expressing various different possibilities for the application's design and then finally think through these ideas to come up to a final design solution for the visualisation.

The FdS methodology consists of:

- Sheet 1 (Brainstorm) – Sketch and plan ideas with the potential effectiveness of the ideas
- Sheets 2, 3 and 4 (Alternative Designs) – Draw three possible principle design solutions
- Sheet 5 (Realisation) – A final realisation design which is ready to be implemented

The development of the narrative visualisation of Air Quality in India has been achieved by following the Five Design Sheet methodology in order to consider and brainstorm through every aspect of the case study and data and design the best possible implemented solution. The five sheets for this visualisation are further discussed and explained in detail.

2.1 Sheet 1 – Brainstorm

Sheet 1 is the first step in the design process. This sheet is all about ideation and brainstorming of all possible visualisations, graphs, plots to convey the story of our data. The sheet is divided into five major components to be achieved before starting to design the possible design solutions.

The five components are:

- *Ideate* – Designers to sketch maximum possible ideas
- *Filter* – Designers to remove any duplicate, irrelevant, impossible and infeasible ideas
- *Categorise* – Designers to understand similar ideas and different ideas and categorise accordingly
- *Combine & Refine* – Designers to organise and imagine the brainstormed mini-ideas towards bigger solutions
- *Question* – Designers to analyse, discuss and reflect on what has been ideated

Firstly, the datasets used are explained with their pros and cons. The intended audience of the visualisation application is clarified, along with the desired outputs of the data to be shown to them.

Ideas like spatial map, heatmap, bar graphs, line graph, area graph, tree map and sunburst chart have been sketched on paper for the proposed narrative visualisation of Air Quality in India. These plots were chosen as per the questions to be answered and the story to be conveyed to the users. The ideas were further filtered to be suitable or not, to be a part of the probable design solutions. Sunburst chart was filtered out in the process as it is not great to the human eye for understanding the information it conveys, easily. Similarly, tree map and the idea of different shapes for different pollutants were rejected as they are not useful for the planned visualisation story to be conveyed through. The remaining graphs were then chosen to move forward and be a part of the next steps.

The ideas were then categorised into 3 types in regards with 3 major sub-topics to be covered for the data's visualisation. Similar ideas were group together and boxed into the same category. Next, the ideas were organised properly to be combined and refined and their usage to be imagined for the bigger picture. The ideas were now clearer and more had a better purpose to serve. The steps carried out in the sheet are questioned and evaluated before finalising, and also explained with the required information wherever required.

2.2 Sheets 2, 3 and 4 – Alternative Designs

These three design sheets allows the designer to sketch the possible alternative design solutions. These are used as a guidance for the designer to reach the final design solution (Roberts et al., 2015). The ideas in first sheet need to be converted to solid designs by expanding on each one of them. The main idea here is to develop three standalone designs answering all the questions asked. Then the three designs are focused upon, analysed, operation listed and discussed for acquiring detailed insights. These sheets are divided into five major components to attain maximum information from them.

The five components are:

- *Information* – Designers to add the appropriate meta-information
- *Layout* – Designers to sketch an interface with details, menus, buttons, etc.
- *Operations* – Designers to list down all the operations and functionalities of the design
- *Focus* – Designers to showcase the core concept of the visualisation and the major interactions by showing the flow diagrams
- *Discussion* – Designers to discuss the positive and negative sides of the design or the advantages and disadvantages as a critique to check its suitability and potential

2.2.1 Sheet 2

The second sheet of the Five Design Sheet methodology used for this visualisation, mainly consists of a combination of choropleth map and symbol proportional map. The choropleth map will be used to understand the distribution of the air quality and the individual pollutants all over the nation. Moreover, the same map uses a car symbol on top of each state which denotes the count of vehicles in every state, by making its size proportional to the actual count. And this map has its further dependent plots showcasing the factors that affect the air quality in India. The dependent plots are displayed as bar graphs for each of the factors, namely, the type of area affecting the air quality along with the effect of Diwali festival as well on the air quality. Also, a ranked bar graph is also displayed to see the top ten worst affected states in India for a particular selected pollutant or the overall air quality index.

As for the focus of the design, a dropdown has been provided to select a pollutant for a separated interactive experience for the user to understand their patterns deeply. And the information gained from the map is through hovering on each state or the car marker to view tooltips displaying the state's name, pollutant level's value and the vehicle count respectively. Along with the hover interaction, click interaction for each state is included to display the dependent plots accordingly. The dependent plots are drill-downs for each selected state, displaying the effect caused individually.

The design led to its discussion where a fair share of pros and cons were discovered. It can be seen that a spatial map is very helpful for the end-users to gain knowledge from and also understand meaningful patterns based on the actual location of states. Also, multiple dimensions can be included in a map to gain more knowledge. But, some cons like clutter or too much interactivity should not confuse the user.

2.2.2 Sheet 3

Next, the third sheet also tried to answer the asked questions, but by different set of plots and visualisations. Use of heatmap along with a combination of a line chart and scatter plot has been done in this design. The line plus scatter plot is used to visualise through three dimensions of the data. An evolution through the time period, along with the pollutant level's value and count of vehicles for a particular state can be understood in this graph. The size of the scattered points on the graph denote the pollutant's value for the air quality. To support the plot, a heatmap to list all the states through the years is also plotted which shows the pollutant's values listed next to each other.

Dropdown have been provided to select a particular pollutant for visualising about it, along with a state dropdown for the line plus scatter plot to aggregate data for a particular state. Also, tooltips on hover have been designed to showcase more information and more insights identifying further patterns, for both the plots. The focus is clear in this design as the interactions and the plots themselves, lead to provide maximum meaningful information to the users.

And, while discussing the design, the share of negative outcomes has been higher than that of the positive outcomes. Even though the design shows a good solid trend over the time period of the data, and also covers all the important dimensions required, it is not conveying enough information to the users. Additionally, a heatmap should not be preferred over an actual spatial map. A heatmap cannot convey the amount of information and trends that a spatial map conveys, as the states are simply listed with blocks of colour values. Whereas, a spatial map shows the true locations of the states on the globe, which reveals much more patterns and behaviour than a heatmap. And, talking about the line plus scatter plot, it is not an appropriate choice for novice users, who have limited knowledge about the subject or domain. Alternative and easy to understand plots can be a better choice for such audience.

2.2.3 Sheet 4

The design in third sheet showcases an interesting and a unique visualisation with the use of a comparative analysis between two states. The aim of this design is to display a comparison between a user selected state and the worst affected state for air quality in the country. This approach tells the users a story of each individual state while evaluating it with the worst levelled state with its measures already displayed on the screen. Use of area charts have been done to achieve the comparative analysis design solution. An evolution of the air pollutant's level through a year's time aggregated for all the available values can be shown for better understanding of the time dimension. Also, a seek bar for the users to select a year to drill down in time can be used. Factors affecting the air quality like

vehicles, Diwali and types of areas can be enforced through the user interface by adding multiple respective buttons and dropdowns for selection, to change the view accordingly.

The users have multiple dropdowns, buttons and a seek bar for interaction with the visualisation, to focus upon the desired topic or question of the case study. Major dropdowns include the selection for state and pollutant. Tooltips, again playing an important role in this design, as they focus to gain the important information of the necessary pollutant's or dependent dimension's values.

As the approach remains unique, interesting and effective on the end-users, the design faces both sides of discussion. A comparison with the worst affected state leads to a direct evaluation and a clear message in the user's mind for the selected state. A trend can be easily identified and understood by the users when the time evolution is through the months of a year. But, without displaying the actual location of the states through a map or showing just a single visualisation as a solution for all the topics and questions, might lead to an incorrect analysis in the user's mind. All aspects might not be covered by the user and might take wrong or incomplete insights from the visualisation. Also, is this enough to be a standalone design? Instead, it may be used as a part of the whole visualisation focusing on a particular topic.

2.3 Sheet 5 – Realisation

The fifth and the final sheet of FdS is the realisation and final solution of the entire process of design. The developers consider this sheet as the final concept before the implementation and delivery (Roberts et al., 2015). This sheet is exactly how the final visualisation application or tool should be developed. This design can be a same design as one on one of the sheets, i.e. 2, 3 or 4. Or it can also be incorporation of the concepts displayed on these sheets. The sheet is categorised into five similar components of the previous sheets, except for the last one. Namely,

- *Information*
- *Layout*
- *Operations*
- *Focus*
- *Detail* – Designers to include the main algorithms or patterns used with the technology used for the development. Also, the time and cost estimates of the application development along with other dependencies, if any.

The realisation sheet for this visualisation includes an amalgamation of ideas and concepts, mainly from the sheets 2 and 4. The sheet was designed after a thorough analysis of all the three alternative designs sheets. The pros and cons of each sheet are considered to finally arrive on the final design solution. This design aims to conveniently convey the story of India's air quality with a seamless and pleasant experience to the users.

The visualisation will be showcased through 4 chapters or tabs in the application. It contains two maps, one being just a choropleth map, and the other one being a combination of choropleth and symbol proportional map. The first map, in the first tab, aims to convey a story of the pollutants throughout the country by a user selected pollutant. While the second map, in the fourth tab, adds the vehicle count dimension on top of the air quality index being showcased as the choropleth part of it. The vehicle count in each state can be understood by the difference in the car markers' sizes on the map as they are proportional to the actual count of the vehicles. Dependent plots exist along with the first, i.e. choropleth map, which cover the factors like types of areas and Diwali festival which affect the air quality in India. These are visualised with the help of bar plots along with the map, as they are drilled down based on the selection of a state on the map, by the user. A ranked bar graph showing the top ten worst affected states for the selected pollutant is also displayed under the map. A comparative

analysis from the fourth sheet is included in the second tab of the design. The worst affected state in India is compared with all the other states based on the user's selection. And, as the last plot, a time series is displayed for each pollutant for each state, through 2003 to 2015 of the data. The time series graph unfolds the evolution of the pollutants through the years. The seek bar here can also help the user drill down the view as desired.

The focus of the design changes as the tabs are changed in the application. For the first tab, the focus remains on the dependent plots, when they change as the user clicks on any state for drilling down further. The ranked graph is also focused upon based on the user selected pollutant. Along with the click interactivity, hover interaction for every plot in the design has been provided to display the respective details of values of the graphs. Necessary dropdowns for each tab have been provided for interaction, and a seek bar to travel and drill down through time for the time series graph.

Lastly, the details of the design have been noted, where the technology to be used for the development is listed as R Shiny along with the details of datasets used as source. Also, detailed estimates of time and cost required are calculated and mentioned accordingly.

3. Implementation

The visualisation for this case study required multiple data sources, then to be merged, and then easily used for plotting on screen. The data sources used are:

- **Ambient Air Quality for India Dataset**
  The analysis is primarily based on the Historical Daily Ambient Air Quality Data for India during the years 1987 to 2015 ("Historical Daily Ambient Air Quality Data," 2017). This large data with 435,742 rows and 13 columns, is contributed by the Ministry of Environment and Forests and Central Pollution Control Board of India under the National Data Sharing and Accessibility Policy (NDSAP).
- **Dates for Diwali Dataset**
  A date dataset from the web, for India's festival of lights – 'Diwali', has been used to check its influence towards the air quality ("Diwali Date List," 2020).
- **Registered Vehicles in India Dataset**
  Registered Motor Vehicles in India during the years 1951 to 2013, has been used as a contributing cause towards India's increasing air pollution ("Total Number of Registered Motor Vehicles in India during 1951-2013," 2016). This data is contributed by the Ministry of Road Transport and Highways under NDSAP.
- **State-wise Registered Vehicles in India Dataset**
  Another vehicle dataset, State-wise Total Registered Motor Vehicles in India from 2001 to 2011, has been used to know the counts of vehicles in the different state in India distinctly ("State-wise Total Registered Motor Vehicles In India," 2015).
- **Shapefiles – India**
  Shapefiles for India have also been used to visualize the nation's air pollution story on a map ("Download data by country," 2020).

The data has already been wrangled, merged and explored to gain important insights and findings. Based on the exploration, the data has been filtered to exclude the records before the year 2003, for this visualisation to be more insightful. Moreover, the data is further wrangled and stored in separate CSV files as per the design's requirement. This ensures an improved performance of the application as the master dataset is huge and its direct usage might lead to inefficiency and slowness.

3.1 User Interface

The final designed solution for this visualisation was planned to be implemented in R Shiny. R consists of some very well reputed and powerful libraries which are apt for data visualisation. *Shiny* is such package which allow the developers to create interactive web applications through R. And, as the application ought to be end-user interactive, *Shiny* gives the developers a powerful and open source framework to build the web applications. Moreover, additional libraries which are useful for data visualisation, are used along with *Shiny* to invoke the graphs on the application. Styling is an important aspect of an application, and R Shiny comes with a ready library filled with some standard styles using inbuilt CSS, *shinythemes*. This library has been used to give the application a standard and sophisticated look throughout. Additionally, libraries like *shinycssloaders* and *shinyWidgets* have been used to add richness to the user interface of the application. *shinycssloaders* library has enabled the application to have loading wait animation to the plots until the graph is rendered on the interface. While, with the help of *shinyWidgets* library, an attractive and soothing background colour throughout the application has been assigned.

The application is divided into 4 major tabs and 2 additional supporting tabs. The tabs have been categorised based on individual relevant topics, which answer the research questions, and conveys all the necessary knowledge through high-quality visualisations.

3.2 Maps

As the application was designed, it contained a large variety of graphs, plots and maps which require a range of libraries in R. Plotting these require different libraries in R, and eventually to be invoked in the Shiny application. Starting with the maps, both of them, the choropleth and the combination of choropleth and symbol proportional map were plotted using the *Leaflet* library. Leaflet is among the very powerful and highly used open source library which is used to develop user interactive maps. The choice of using maps and *leaflet* for maps go hand in hand, as maps on leaflet are very interactive and provides clear knowledge to the human visual system. Any other method than spatial maps, for providing the same information would not give the data and its trend enough justice.

Functions like *addPolygons*, to plot the shapes for choropleth, and *addMarkers*, to plot the symbol markers for the proportional symbols, were keys while developing the maps. Choropleth maps require a *Spatial Polygon Data Frame* of the India's map to be imported, which can be achieved by merging the shapefiles for India along with the actual clean and merged dataset during the exploration. To read these shapefiles and convert them to the required format, the *sf* library is used. *sf* library enables easy functionalities on the shapefiles for displaying the final map on *leaflet*.

The maps have been enabled with additional features like hover and click events. On hovering on different states, they are highlighted, and relevant informative tooltips are displayed for conveying important information. And, a zoom and view feature has been enabled in the map when clicked upon a state, to set the view on the respective clicked state, zooming in a little. Also, drilled down dependent plots are displayed to change on click events, to aggregate the plots for the respective clicked state.

3.3 Plots / Graphs

For plotting all the non-map graphs in the visualisation, variety of other libraries have been used. The plots include dependent and supporting bar graphs, area charts for the comparative analysis, and a time series graph through line and area chart. Here, the *ggplot2* library has been used to plot the bar graphs and area charts. Although, the plots are later rendered into plotly graphs in the Shiny application. Hence, the *plotly* library has also been used wherever ggplot2 plots have been coded. The *plotly* library enables easy user interaction with hovering and tooltip functionalities for the graphs. And, a dedicated *dygraphs* library has been used to plot the time interactive series evolution for different pollutants in different states. This library enables the developer to create a high-quality visualisation which helps the users to drill down in time to select their desired time period for the selected pollutant and state and understand its trend by hovering through and the tooltip at the top-right corner.

The choice of graphs has been pretty straight forward and has aimed to be simple and easily understood by the users. Bar graphs has been a clear choice here, when the map needed additional dependent and supporting graphs to convey important information. The users should not find it difficult and should not spend more time in understanding the story, rather than actually browsing through the visualisation application. Any other graphs could have created high amount of distractions, when cluttered with multiple visualisation in the first chapter's tab. The remaining plots are chosen based on high research of graphs, which are effective through all aspects towards the users.

3.4 Colours

Moving ahead, an appropriate use of colours has been done for all plots to ensure an eye soothing experience to the users. The *RColorBrewer* library has been used, to generate manual colour palettes and assigning them respectively. Use of manual colour bins has also been done to assign colours to a range of values for visualising the choropleth map. This ensures that the user understands the story behind the map and depicts the accurate meaning of it. Similarly, all the plots used in the visualisation are given appropriate colours which will be helpful for faster and easier transfer of knowledge to the users.

3.5 Data Processing

The *dplyr* library is used for any processing required before the actual import of data in a particular visualisation's code. The library is used for faster and easier data modification, and setting the data quickly based on the user's selection through the various user interface options provided. The *scales* library has also been used for processing the time data. Here, the months in dates while plotting the trend through time as area charts for the comparative analysis, are processed and modified as required. The months which are stored in numbers in the date data type, are converted and shown by the actual month names, i.e. January, February, etc., on the axis. Lastly, an instance of the *xts* library has been used to convert the dataset consisting of a date column to an extensible time-series object, making it ready to be imported in the code required for plotting the *dygraph's* time series graph.

```r
# import the required libraries
library(shiny)
library(shinythemes)
library(leaflet)
library(dplyr)
library(plotly)
library(shinycssloaders)
library(shinyWidgets)
library(dygraphs)
```

```r
# import the required libraries
library(shiny)
library(sf)
library(leaflet)
library(RColorBrewer)
library(scales)
library(ggplot2)
library(dplyr)
library(plotly)
library(dygraphs)
```

Figure 1. Libraries used for development in ui.R & server.R respectively

4. User Guide

The R Shiny application developed for visualising the story of India's air quality consists of many features and functionalities. Some features are related to the application's interface, while some features are related to the various visualisations plotted in it. To make it easy for browsing through the different topics within the case study, the application is categorised into 4 major chapters, i.e. 4 tabs in the interface. This section of the report will go through every functionality of the Shiny application in detail, to guide the user through it.

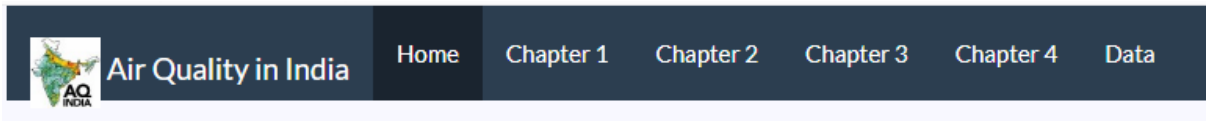


Figure 2. Tab panel of the visualisation

The *Home* tab here is an introduction tab to the visualisation, where the user is familiarized with the case study and the topics which are focused upon in the further tabs. And a *Data* tab is also added to the interface for the users to browse through the resource links of the original open data used for the entire visualisation.

4.1 Chapter 1

The first tab of the application focuses on the distribution of the 4 major pollutants contributing towards the air pollution in India through a choropleth map of India. And the distribution of a resultant Air Quality Index derived from the 4 pollutants is also plotted here, based on the user's choice through the dropdown for pollutant selection as shown in figure 3.

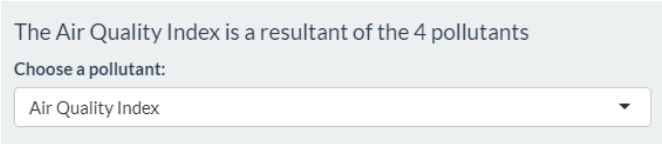


Figure 3. Dropdown for pollutant selection

Further, the map is developed with hover and click event functionalities to provide the users with important insights. As shown in figure 4, the map has no dependent plots as there has been no click event on any of the states. But, once the user clicks on any state, the data is aggregated for the respective state, and drilled down dependent bar graphs, showcasing the effect of factors like types of areas and Diwali, are displayed. Also, the map is zoomed a little and its view is set on the clicked state for focus. Figure 5 gives a clearer idea of the interaction with the map. The combination of these plots helps the users to browse through the map, allowing them to hover and click on each state to focus upon. Another bar graph showcasing the rank, i.e. the top 10 worst polluted states for the user selected pollutant, is also displayed under the map, as shown in figure 6. Each of these plots are enabled with mouse hover functionalities to display the relevant and necessary values and information for better understanding. This chapter is aimed for the user to browse through and get an idea of the spread of pollutants, and thus, the pollution status of all the states in India.

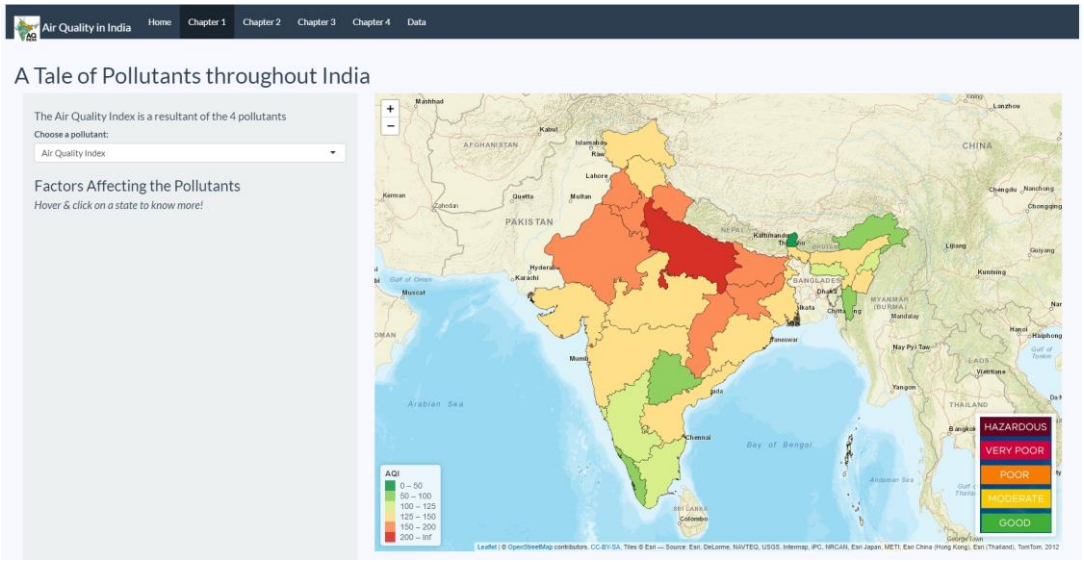


Figure 4. Choropleth map of India before any click interaction

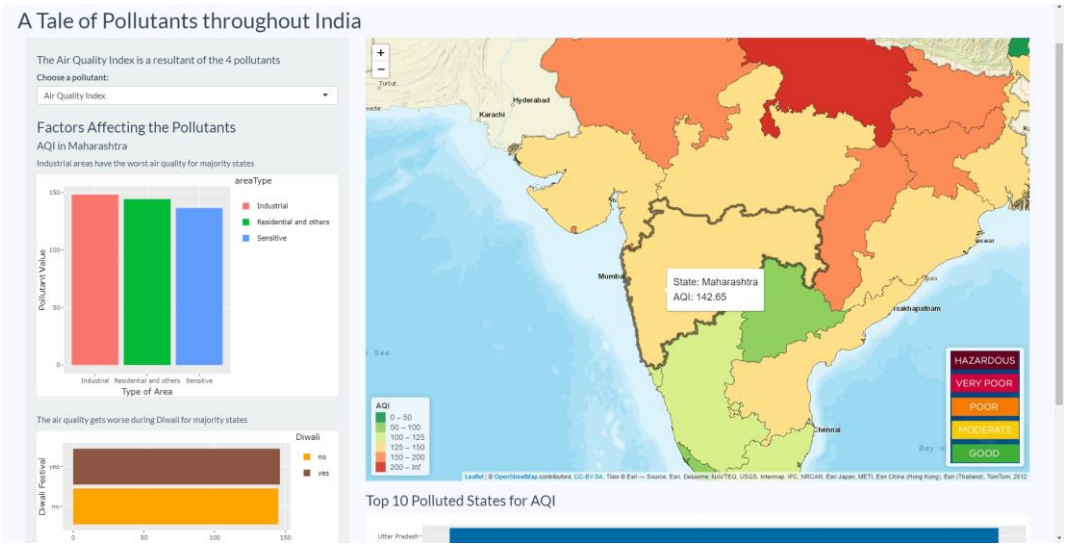


Figure 5. Choropleth map of India after hover and click interaction

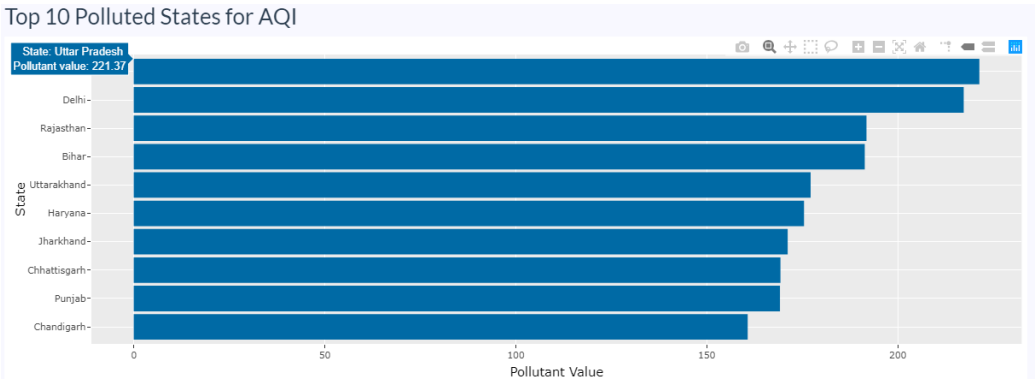


Figure 6. Bar graph for the top 10 polluted states

4.2 Chapter 2

The second chapter focuses on the comparative analysis for a user selected pollutant and state, with the highest polluted state in India, as learnt in the first chapter. A subplot of area charts to show the comparative trends of the states are plotted here.

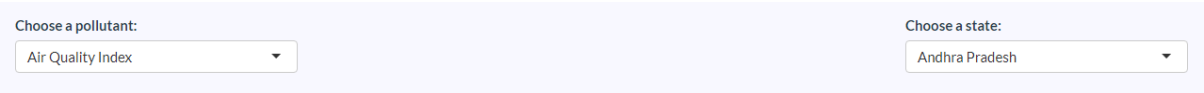


Figure 7. Dropdowns for pollutant and state selection

As shown in figure 7, couple of dropdowns have been provided for the users to select their desired pollutant and a state which will be compared with Uttar Pradesh, i.e. the worst affected state.

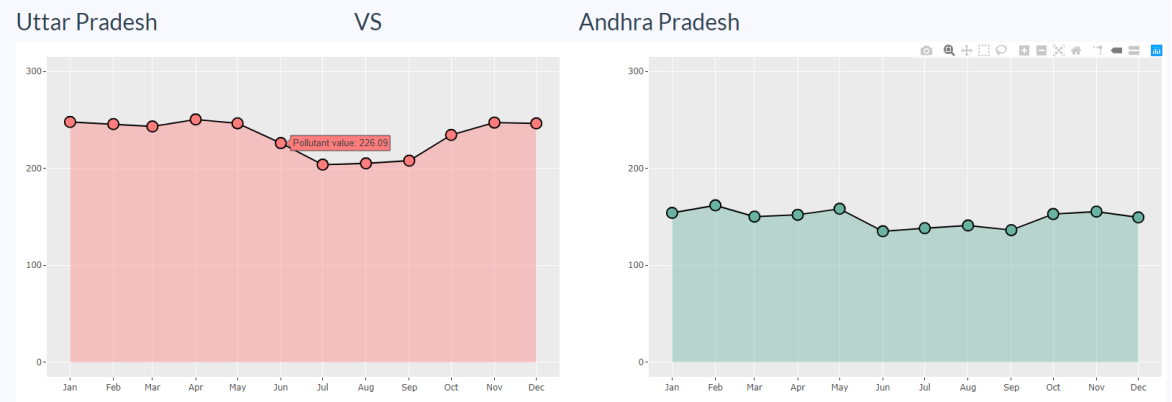


Figure 8. Subplot of area charts for comparison

As figure 8 shows, the selected state will be compared for the respective selected pollutant. Both the area charts are enabled with mouse hover interactivity for users to learn the exact value of the pollutant for that particular month of the year. This average trend through a year will also help the users to know a pattern of times when a state faces high or low air pollution.

4.3 Chapter 3

The third chapter of this visualisation is a display of the evolution of the pollutants through a time period from 2003 through 2015, for each state. A dygraph in R has been used to plot this time series graph.

Figure 9. Dropdowns for pollutant and state selection

As shown in the above figure, dropdowns for the users to select a pollutant and a state for viewing their trend in time have been provided for interaction.

And again, mouse hover interactivity has been provided for the graph to gain the exact values for the data point. A seek bar to select the user desired time frame to gain time targeted insights has been provided under the graph. The seek bar can also be used to drill down in time to select shorter and specific time frames for a deeper understanding of the trend and behaviour of the pollutant. Figure 10 shows us a view of the graph and the interaction provided.

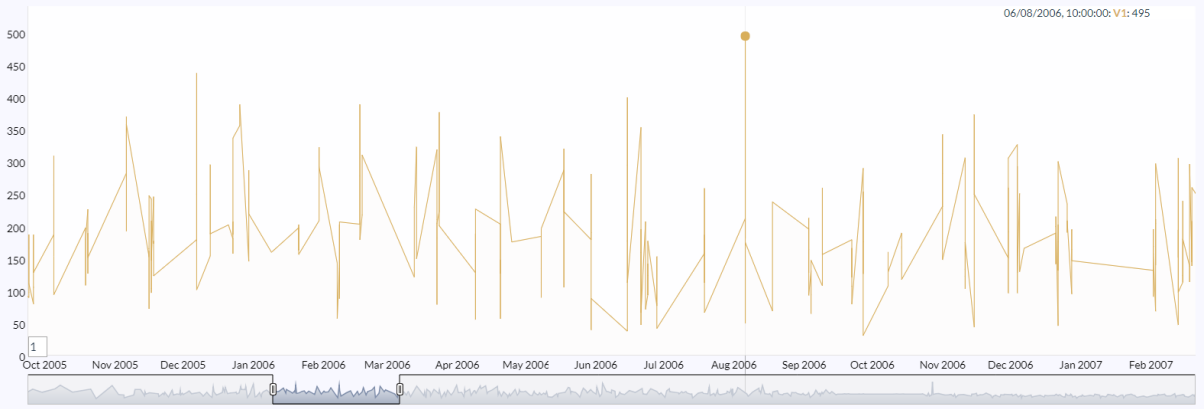


Figure 10. Time-series graph of the pollutants in a state

As for the exact values on hovering, a reactive tooltip has been provided at top-right corner of the graph. The tooltip displays the exact date and the respective value for the pollutant selected. Figure 11 gives a focused view on the tooltip.

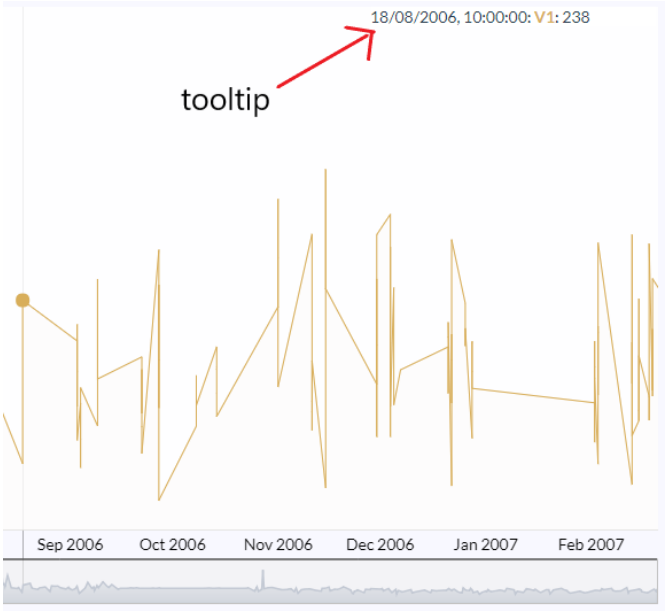


Figure 11. Tooltip for the time-series graph

And, the below figure shows the seek bar used for the graph as mentioned earlier.

Figure 12. Seek bar for drilling down in time

4.4 Chapter 4

The final chapter of the visualisation is related to the effect on the air quality of India due to the rising count of vehicles in all states over the years. The data for the year 2015 has been chosen for this visualisation, with the Air Quality Index considered for the plotting the air quality on the map. A combination of choropleth and symbol proportional map has been used to convey the relation of the factor. The map comes with a similar hover and click functionalities, with the zoom and focus on click feature, for user interaction, as mentioned for chapter 1. The proportional symbols are car markers placed on the map; whose sizes are proportional to the count of vehicles in the respective state. The combination used in map shows the users a direct comparison and relation between the two attributes. A reactive text has also been displayed in the sidebar panel based on the state clicked. The figure below explains the interactivity better, that is provided for the map.

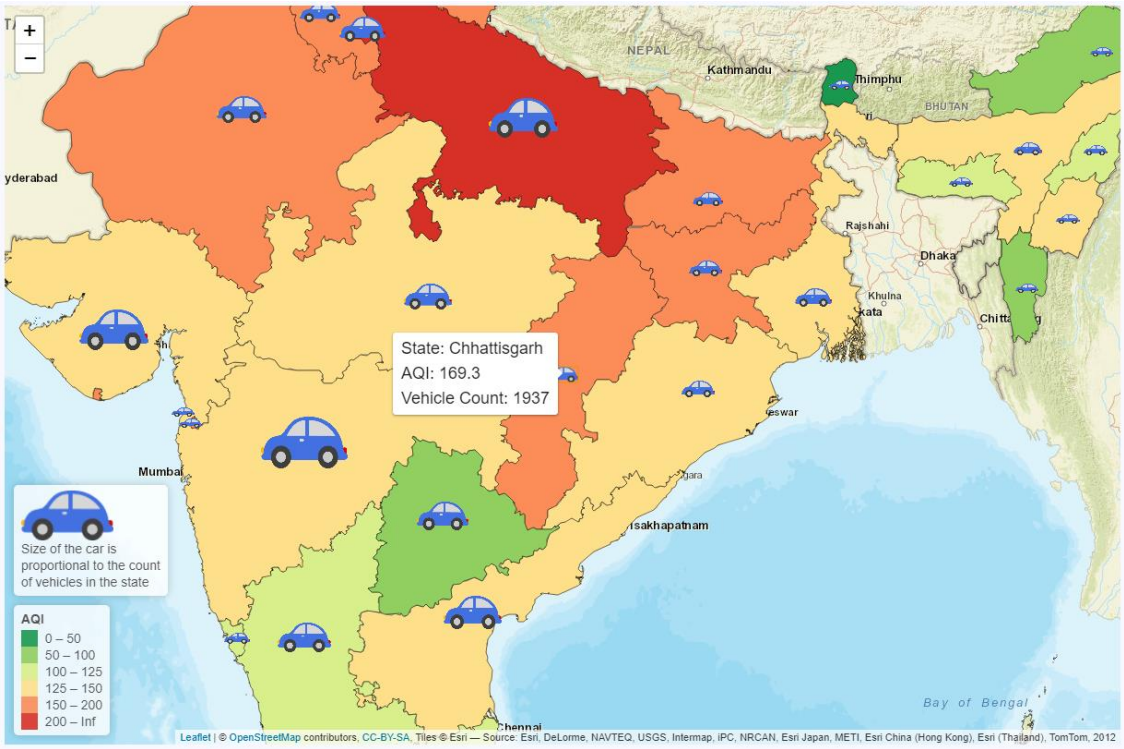


Figure 13. Choropleth + symbol proportional map for plotting vehicle count and air quality

And, figure 14 below shows the reactive text which is dependent on the respective state which the user clicks.

Figure 14. Reactive text for click interactions

5. Conclusion

The Historical Daily Ambient Air Quality data for India had plenty of stories to be told after it was wrangled, merged with the dependent datasets and finally explored to gain findings and insights. Communicating the findings through visualisations was a challenge, as there were multiple topics which were covered along with the India's air quality. The various factors that affect the air quality of India played a crucial role in communicating the story to the end-users. The audience being the general public and citizens around the country as well as the world, it was necessary for the visualisations to be effective and convey the necessary knowledge with least possible efforts. In order to achieve that, easy to understand and commonly found visualisations like spatial maps, bar graphs and area charts were used and developed in a complex way, with concepts like drilling down based on state selection and comparative analysis of the states. A drill down based on time for the time series was also achieved which provides chunks of information to the users based on their desired interaction. A complex visualisation like the choropleth plus symbol proportional map is very informative, but I, as a developer, needed to be careful of the car symbol used on the map as it may hide some states and their information.

Leaving the users filled with complete and accurate information from the visualisations was a key in this design and development process. Studying the human visual system and its way of grasping the images, shapes and colours of the visualisations helped me in developing the application in a seamless manner, while considering a user centric design. Developing the solution as designed from Five Design Sheet methodology was also challenging, considering the feasibility and possibility of the plots, graphs and the user interface the technology provides with. Various topics could have created a clutter of visualisations and information for the users but inter-linking them all together made it better for easy depiction. Sufficient and easy to use interactivity provided with the visualisations to the users makes them stay involved with the application and extract maximum possible knowledge.

The plots and visualisations unfold many trends, patterns and behaviours from the data, as found during the exploration. But the factors covered in this case study are not all. There are more factors which affect the air quality in India, like crop burning. Including this data with the current scope could have been more informative, but due to insufficient resources, it was not possible to create a relation. Also, considering from the visualisation's perspective, use of more complex graphs and plots could have been possible. But at the same time, development of a user centric design needs to communicate maximum information accurately to the users, especially when they are not familiar with the topic, the subject and the domain. Hence, the choice of graphs and plots still remains dependent on the audience and their ability to grasp the information.

6. Bibliography

References

Anand, G. (2017). India's Air Pollution Rivals China's as World's Deadliest. *The New York Times*. Retrieved from https://www.nytimes.com/2017/02/14/world/asia/indias-air-pollution-rivals-china-as-worlds-deadliest.html?_r=0

Diwali Date List. (2020). Retrieved from http://www.world-timedate.com/holidays/kali_puja_deepavali_date_list.php

Download data by country. (2020). Retrieved from http://www.diva-gis.org/gdata

Historical Daily Ambient Air Quality Data. (2017). Retrieved from https://data.gov.in/catalog/historical-daily-ambient-air-quality-data

Osseiran, N., & Lindmeier, C. (2018). 9 out of 10 people worldwide breathe polluted air, but more countries are taking action [Press release]. Retrieved from https://www.who.int/news-room/detail/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action

Roberts, J. C. (2011). *The Five Design-Sheet (FdS) approach for Sketching Information Visualization Designs.* Paper presented at the Eurographics (Education Papers).

Roberts, J. C., Headleand, C., & Ritsos, P. D. (2015). Sketching designs using the five design-sheet methodology. *IEEE transactions on visualization and computer graphics*, *22*(1), 419-428.

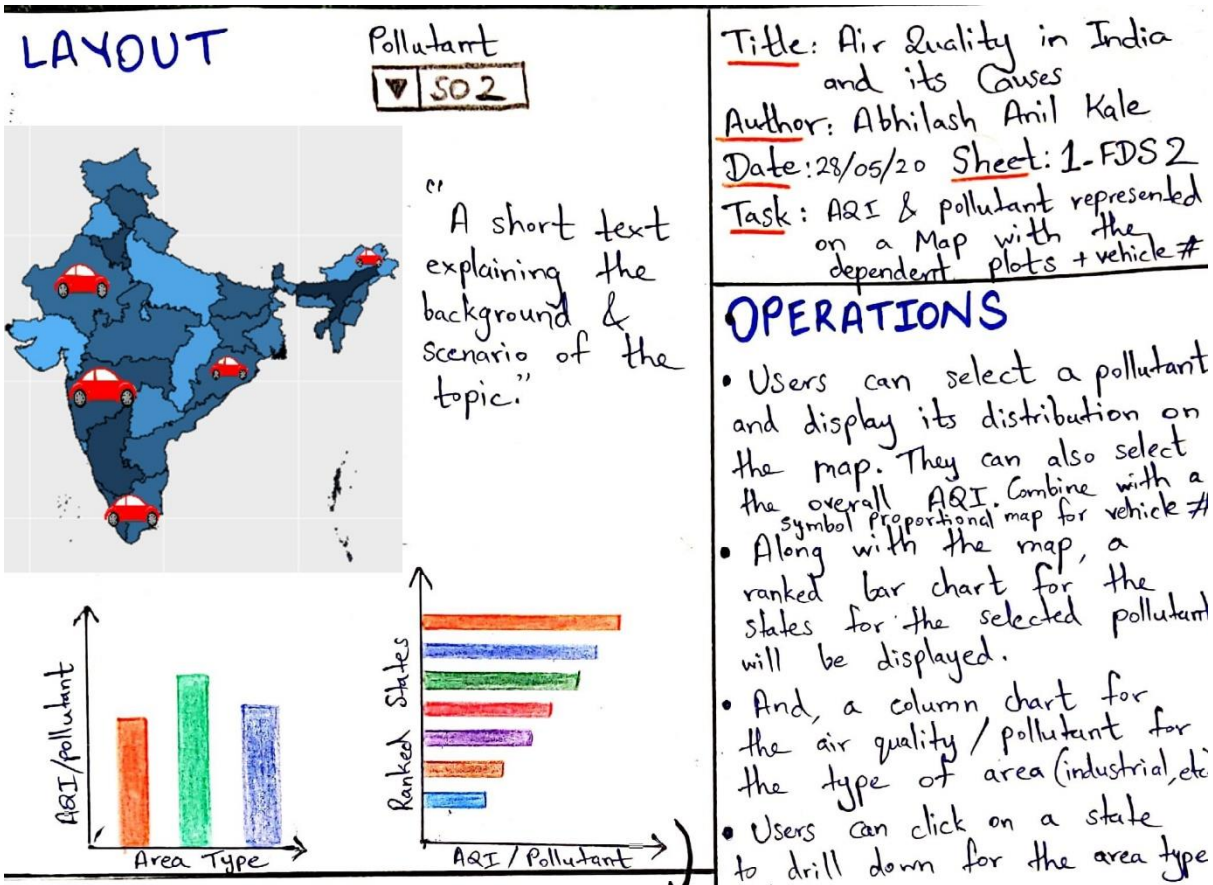
State-wise Total Registered Motor Vehicles In India. (2015). Retrieved from https://data.gov.in/catalog/state-wise-total-registered-motor-vehicles-india

Total Number of Registered Motor Vehicles in India during 1951-2013. (2016). Retrieved from https://data.gov.in/resources/total-number-registered-motor-vehicles-india-during-1951-2013

Appendix

1. Sheet 1 – Brainstorm

2. Sheet 2



DISCUSSION

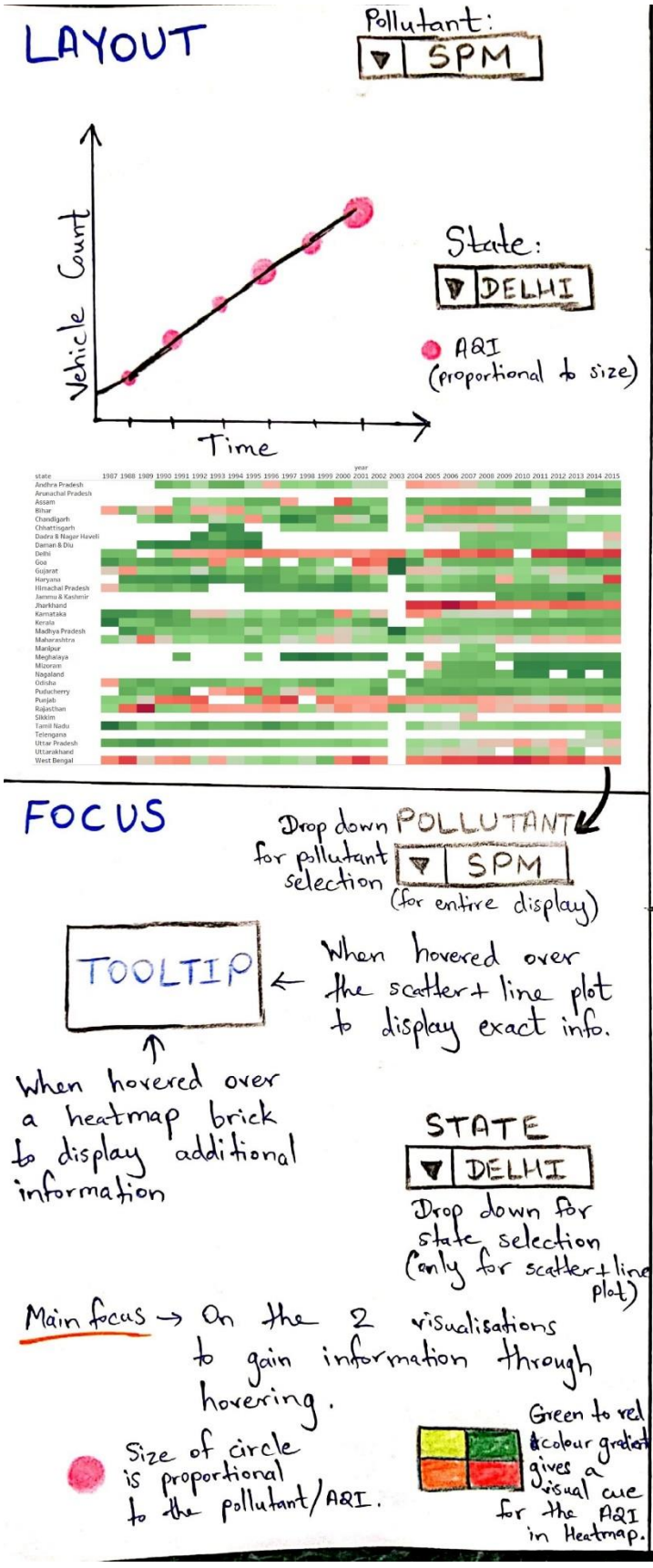
+VE

1. Spatial map is very helpful for users.
2. We can identify various patterns and behaviour from a spatial map.
3. We can add more dimensions to the map to gain more information.

-VE

1. Is it a clutter for the user?
2. Are separate screens better to include everything?
3. Will too much interactivity confuse the user?

3. Sheet 3



Title: Air Quality in India and its Causes

Author: Abhilash Anil Kale

Date: 28/05/20 Sheet: 2-FDS 3

Task: AQI & pollutants represented on a line + scatter plot with the vehicle count. Heat Map to support.

OPERATIONS

- Users can view a line + scatter plot which shows us the distribution of vehicles & AQI, by hovering.
- Users can also view & hover on the heat map to obtain necessary information.
- Users can select the pollutant to view on the heat map and the line + scatter plot.
- Users can also select the overall AQI for the display.

DISCUSSION

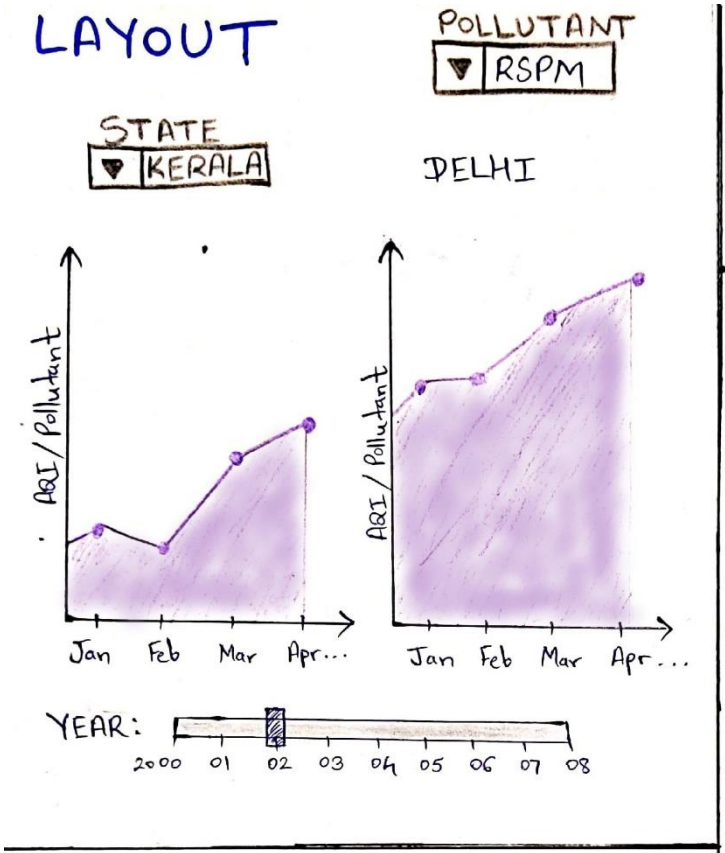
+VE

1. Shows a good trend over the time period.
2. The plots cover important dimensions.
3. Users can explore through the plots easily gain knowledge.

-VE

1. Are they conveying enough information?
2. Is a heatmap preferrable over a spatial map?
3. Can enough trends/behaviours be identified through a listed heatmap?
4. Is it user-friendly?
5. Is a scatter + line plot helpful to a novice user?

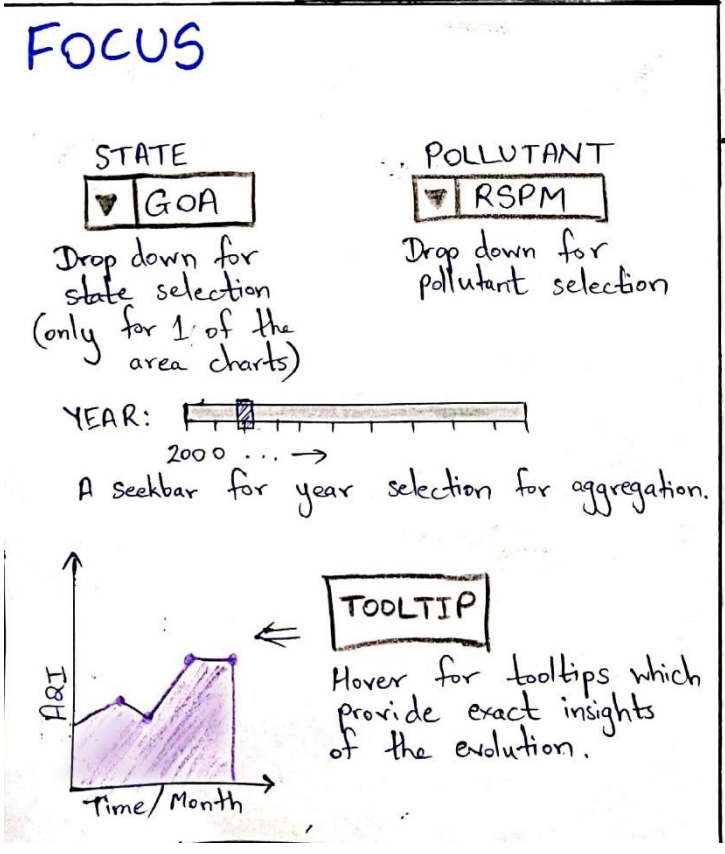
4. Sheet 4



Title : Air Quality in India and its Causes
Author: Abhilash Anil Kale
Date : 02/06/20 Sheet: 3-FDS4
Task: Comparison within states for AQI through time

OPERATIONS

- Users can select a state to view the AQI trend.
- Users can hover on the area charts to gain the exact insights of the evolution.
- The comparison will be an animated trend going from January to December for both states.
- The comparison will be between user selected state vs Delhi, which is among the top polluted regions in the world.



DISCUSSION

+VE

1. Comparison is a good way to identify trends in the data.
2. Comparison with the worst affected state may be a good way to evaluate.
3. A trend within a year's time for a state may help the users gain more insights.

-VE

1. Is it enough to be a stand-alone visualisation?
2. A good way for analysis for the users, but is it covering all the aspects?

5. Sheet 5 – Realisation