CSC8626 Data Visualization: Summative Assignment Report Sheet

Name: Subin Jung

Part One Task	Description of how your submission achieved this.
Fit to task: does the visualization allow the identification of areas most and least in need of aid.	Generally, based on the *population vs. impact of infections* graph, it is obvious that the population is significantly related to impact of infections. Even if filtered data for each group demonstrates that the more population has the more impact. According to the bar graph of *measure of impact of infections*, the *above 10* group has the highest impact of 3.2. Also, after applying filter to display only *above 10* group, the *average of all simulations* shows that it has 22.73 number of infections and the impact of infections is 3.19 even though accuracy is 1.40. Compared to *8 to 10* group, the value is quite high considering that maximum number of infections is 51.75. On the contrast, the *0 to 2* group has the lowest impact of 0.7 and the number of infections is 0.61 even though it is distributed in large area. In conclusion, the group of *above 10* seems like needing the aid the most and the *0 to 2* group at least.
Use of visual channels	The dashboard consists of two big parts, information and graphical part. The information part is placed on the left and the other one on the right. Especially, numerical information is placed on the top left showing *total number of infections, average number of infections* and *average of all simulation*, which can deliver the exact value of information to user. On the right side, four graphs are plotted, a donut chart, a bar chart and two scatter plots. Donut chart explains the ratio of total infected cases by group based on the mean value of four simulation. As donut chart presents the area difference intuitively, users can easily understand which group has the most and least infected cases. In addition, the scatter plot which located right next to the donut chart represents the correlation between population and the infected cases based on the same data, mean value. It may be hard to explain only with scatter points so the regression line is added as a dotted line whether it has significant impact depending on the population. Furthermore, the horizontal bar graph describes the measure of impact of infections to compare the impact between groups at once. The data is based on the standard deviation so that it can measure the impact of infections evenly. Finally, another scatter plot is about the correlation between population and impact of infections. Again, this plot also includes the regression line to express the relationship clearly.
Gestalt design principles	<u>Proximity</u> At the top left, there are three boxes – *total number of infections*, *average number of infections* and *average of all simulations* – spaced apart by the recognisable gap so that user can understand the information indicates different purposes. However, although the two boxes at the very top is displaying the four simulations individually, the last box is the summarise of four simulation – which is average – it is placed on the bottom of them but in the middle. In addition, below the boxes but at the left, filters and slicers are set together since they are all taking charge of the same functionality –

	controlling interactive actions. On the half right, all the charts and plots are placed next to each other to let the users get the graphical information easily at the same time. <u>Similarity</u> At the top left, *total number of infections*, *average number of infections* and *average of all simulations* boxes have the same background colour to highlight that it is only for numerical measurement of the impacts. On the contrary, others' background is different colour than these three. Also, there are 6 different group depending on the number of infected people and the allocated colour for each group is consistent in every graph in the dashboard for grouping. <u>Contrast</u> The title of graph has been written as bold style to let viewers to recognise the graphs individually. For example, the graphs at the right side, the label of x-axis 'Population' and the title of graph below could make confusion to the viewer if the title is not bold.
Use of colour	The colour palette has been chosen considering colour blindness and photocopy. Colour blind safe theme has chosen considering deuteranopia, protanopia, tritanopia and monochromacy. As Figure 2 reflects, all the six colours are distinguishable in the perspective of every colour perception deficiency. Especially, since there exists regression line in scatter plot which overlapping partially, the contrast colour is chosen to make the information clear. Moreover, the final colours are confirmed by attempting to print the dashboard page in black and white. Therefore, the plots can be recognisable in the photocopy as Figure 3 shows. Lastly, if there is a same group appearing in different graph, the same colours are used for each group for consistency of whole page.
Use of interaction	There exist one filter and two slicers in the dashboard. They are located at the very left side to let the user to filter some unneeded and hidden data easily (Sparkman *et al.*, 2022a). For example, among the group, since the *0 to 2* group occupies almost half of the data, it conceals the other data points in the map. In this case, group filter – named *Group (Number of Infections)* – helps user to see the other group's data points individually by selecting specific range of group from the checkbox. Even if the group filter can control the data shown by the range, the users might want to see by specific range. Thus, the *number of infections* filter help user customise the range with specific value to see the impact. However, it is hard to say that every variable is reliable. The coefficient of variance describes the uncertainty of every data so the *accuracy* filter let the users can see the certain data relying on their uncertainty.
Use of language and text	Overall, the mean value is named as either *number of infections* or *infected cases* since it is the mean value of infected cases from four simulations. Also, as the users may not be an expert, the intuitive name has set so that anyone can understands what it represents. Not only that, but the standard deviation value is also called *impact of infections* and the coefficient of variance as *accuracy*. Furthermore, the number of infections is grouped into six categories based on their range since it is difficult to show all the single numerical

	data as a legend in any of graphs. Hence, it is separated by the range, for instance, *0 to 2, 2 to 4* and *so on* until *above 10*. Although the maximum value is 51.75 and the minimum is 0.25, most of the data are distributed under 10 so the range has set until *above 10* to avoid the bias in classification.
Technical aspects: performance, reliability, fit on desktop screen.	Performance analyser in Power BI can figure out the functionality of each element (Isleminger, 2022). When using filters, it takes averagely 1000ms and maximum 4016ms to calculate the *impact of infections*. The result of using slicer also shows the similar results. According to Sparkman (2022b), Power BI can match the size either desktop or mobile. This report is customised to fit on the desktop screen having the ratio of 16:9. Furthermore, since data has single values more than 3000, python is used to make a new column into the data and group the mean value into six group relying on the range, for instance, 0 to 2, 2 to 4 and so on.
Part Two Task	
Fit to task: does the visualization allow the identification of areas most and least in need of aid.	In general, random five datasets – Datafile_247.csv, Datafile_058.csv, Datafile_012.csv, Datafile_101.csv, Datafile_225.csv – are selected from 250 files and combined in Power BI. Likewise, they are all wrangled by python creating new column, group of mean. In accordance with *measure of impact of infections*, the group of *above 10* always has the highest impact during five multiple runs. In reverse, the *0 to 2* group consistently appears to have lowest impact in every runs.
Effective visual representation of the data variations over multiple runs.	Overall, Part 2 report is designed based on the structure of Part 1. However, instead of indicating impact of infections by group and the relationship between population and impact, comparison of multiple runs is added as a bar chart. Unlike Part 1, vertical bar chart is plotted as the bar graph helps viewer to see the amount differences between six groups intuitively. Especially, vertical bar graph helps the viewers read which group has higher or lower impact better than the horizontal graph (Esselman, 2022). Consequently, as this report is based on the five datasets, the vertical bar chart is plotted to compare the differences between multiple runs. X-axis represents the different datasets name and each of them includes all the group showing the impact of infections respectively.

CSC8626 Data Visualization: Summative Assignment Report Sheet

Part 1 Screenshots

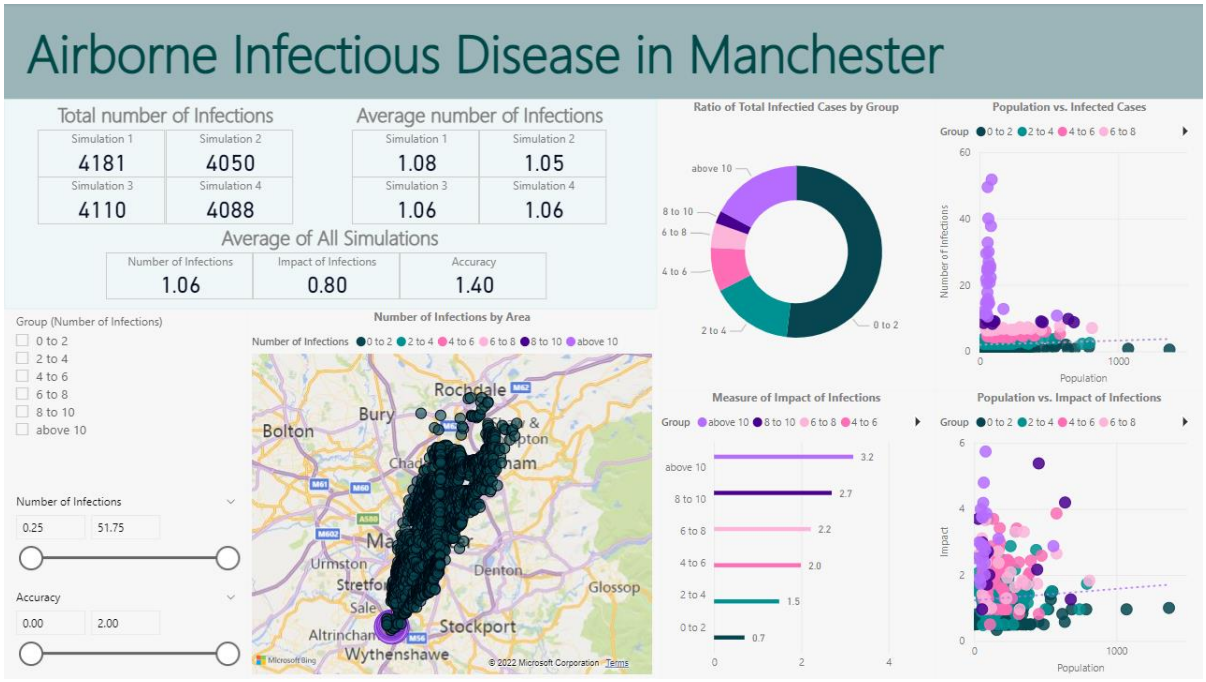

Figure 1: Original dashboard (Part 1)

Figure 2: Aspect of Colour perception Deficiency (Part 1)

CSC8626 Data Visualization: Summative Assignment Report Sheet

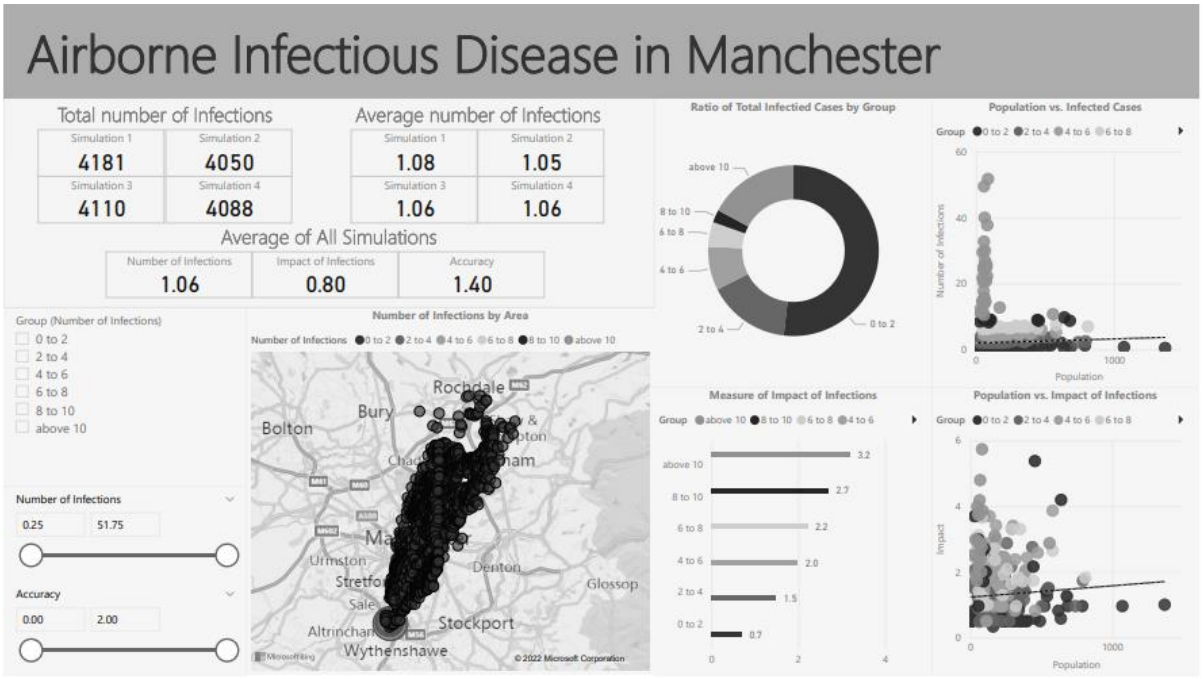

Figure 3: Example of Photocopy (Part 1)

Part 2 Screenshots

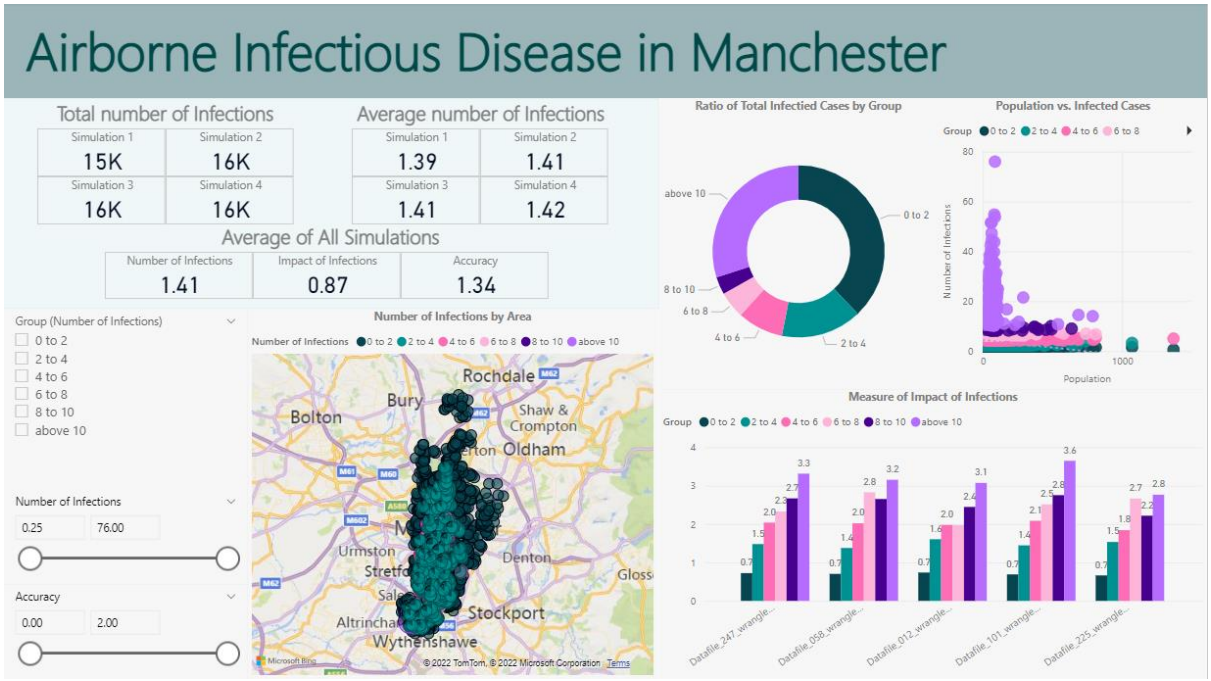

Figure 4: Original dashboard (Part 2)

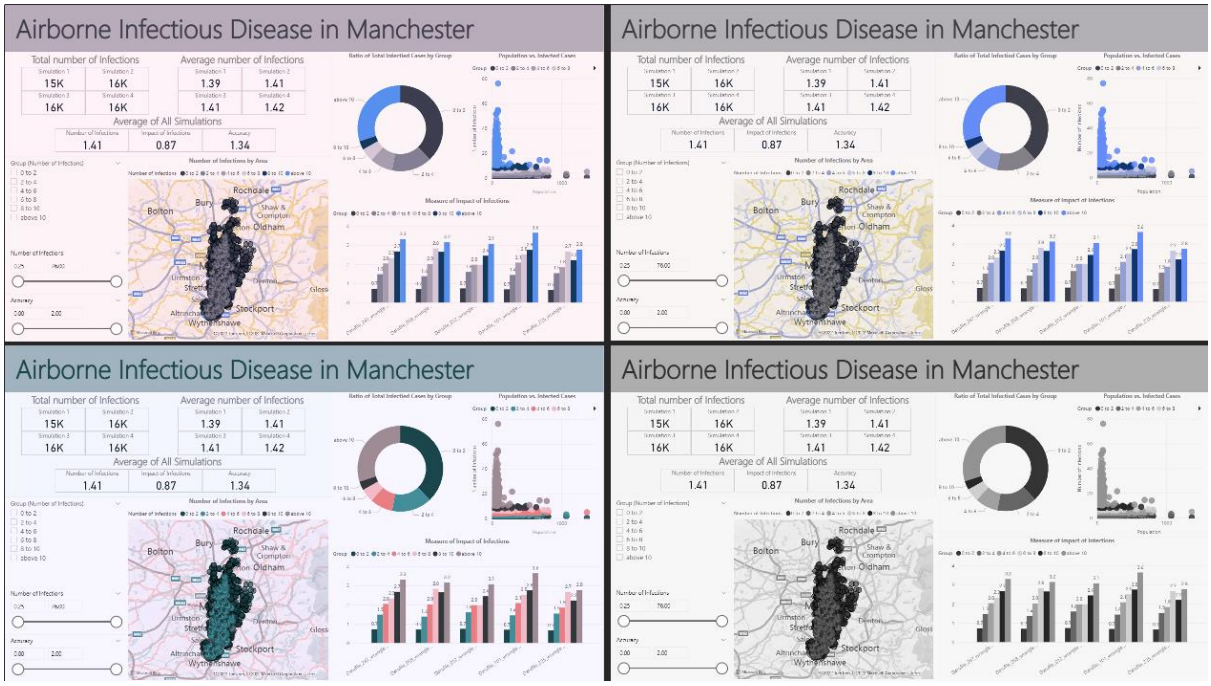

Figure 5: Aspect of Colour perception Deficiency (Part 2)

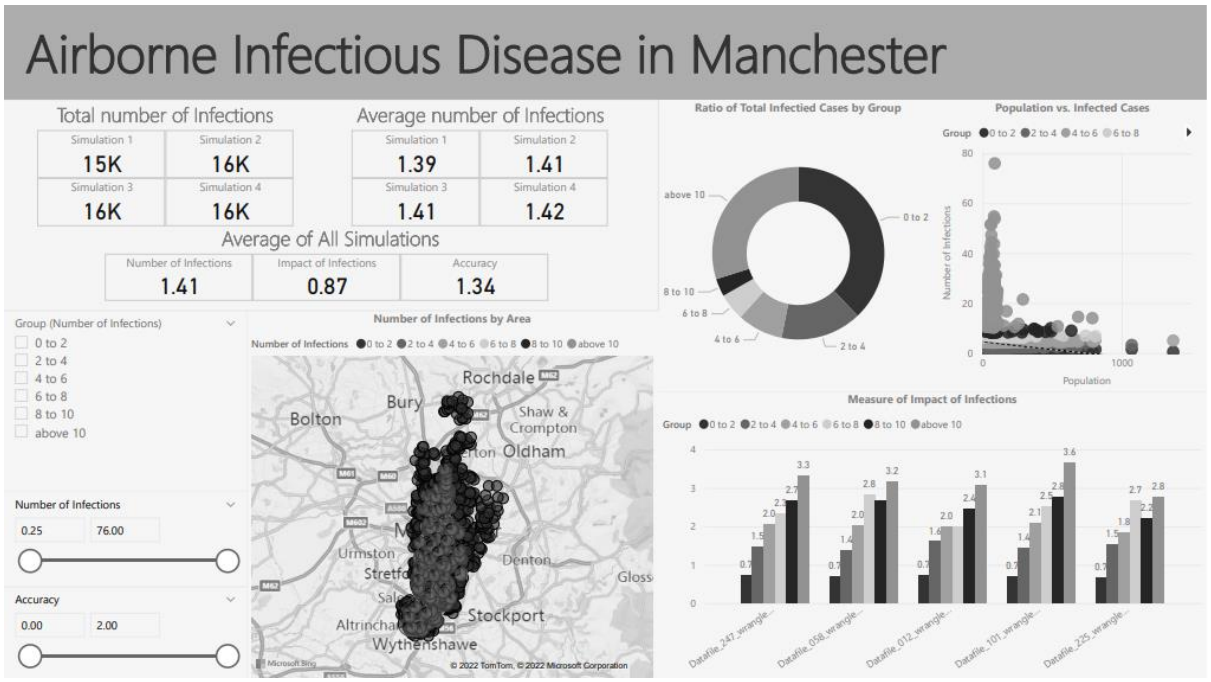

Figure 6: Example of Photocopy (Part 2)

References

- Esselman, A. (2022) *Which bar chart orientation should I use?* Available at: https://www.storytellingwithdata.com/blog/2022/1/21/which-bar-orientation-should-i-use [Accessed 20 October 2022]

- Isleminger, D. *et al.* (2022) *Use Performance Analyzer to examine report element performance*. Microsoft. Available at: https://learn.microsoft.com/en-us/power-bi/create-reports/desktop-performance-analyzer [Accessed 19 October 2022]
- Sparkman, M. *et al.* (2022a) *Slicers in Power BI.* Microsoft. Available at: https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-slicers?tabs=powerbi-desktop [Accessed 18 October 2022]
- Sparkman, M. *et al.* (2022b) *Apply page display settings in a Power BI report.* Microsoft. Available at: https://learn.microsoft.com/en-us/power-bi/create-reports/power-bi-report-display-settings?tabs=powerbi-desktop [Accessed 19 October 2022]