Coursework Description Sheet

Name: Sandra Mishale Niño Arbelaez

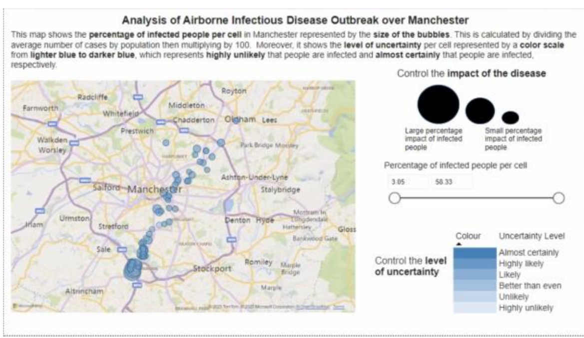
Student ID: 230528581

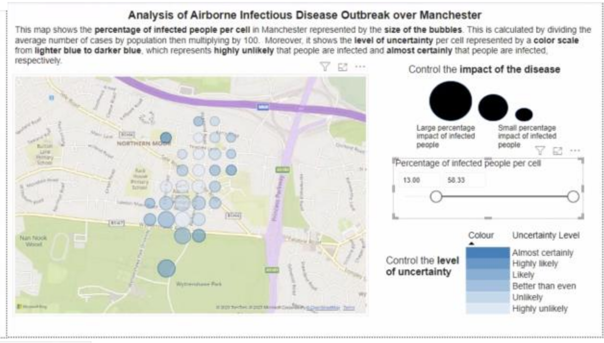
Part One	Description of how your submission achieved this
How does the visualization allow decision makers to understand the overall situation?	In order to understand the impact of the disease over each cell, I calculated a new measure called infected ratio. This measure is calculated by dividing the mean (u) of the four outcomes per cell over the population per cell. Then this value is multiplied by 100 to express it in terms of percentage. Then this measure means how many people are infected in terms of the population per cell. In order to understand the level of uncertainty and assuming our users are people who do not understand statistical methods or mathematics, I chose the standard deviation (sd) and created a new conditional column to represent this quantitative value in an orderly way. I started from the idea that a high value of sd means high uncertainty and a small value is low uncertainty. By doing a previous analysis of the distribution of the values for standard deviation, I created this scale on how certain people are infected: sd >= 4 is Highly unlikely, sd >= 3 is Unlikely, sd >= 2 is Better than even, sd >= 1 is Likely, sd >= 0,5 is Highly likely and sd >=0 is Almost certainly. This approach can scale to other datasets where we can just update the conditions for the standard deviation (or even other uncertainty metric) to assign each level of uncertainty. In this way, I used a map as a graphical representation of the infected people per cell in Manchester. The impact of the disease, that is the infected ratio, is represented in the bubble size, a bigger bubble means more infected people per cell and a smaller bubble means less infected people per cell. It is important to mention that I am showing the cells with the infected ratio greater than 3% to reduce noise in my visualisation of very low percentages. The level of uncertainty is represented in a colour scale from a lighter blue which means high level of uncertainty to a darker blue which means low level of uncertainty.
How does the visualization allow decision makers to decide which areas of the city to target first?	By the law of proximity (Colin, 2013), in the map we can perceive groups because they are close together which indicates that the disease is concentrated in some areas. Then, the decision makers by the law of similarity (Colin, 2013) can look for in those areas for darker-blue bigger bubbles that fall in the levels of almost certainly, highly likely and likely and assign resources in that order to these cells because that means there is a high proportion of infected people in the cell and the level of uncertainty is low. After doing this check in the map and if there are more resources to allocate, then the decision makers can look at separate darker-blue-bigger bubbles in the map to mitigate the spread of the disease in other areas.
Use of visual channels consistent with data variables	Position to encode the location (longitude and latitude variables) of the bubbles in the map. Size to encode the infected ratio. This is to identify how much bigger one mark is from the other, that is bigger bubbles mean high percentage of infected people and smaller bubbles mean low percentage. Colour luminance to encode the level of uncertainty to determine order. This helps us to identify how much darker one mark is from the other.
Use of graphic design principles	The expressiveness principle states that the “visual encoding should express only the information in the dataset attributes” and the effectiveness principle states that the “most important attributes should be encoded with the most effective channels”. (Munzner, 2015). I used the identity channel of position to determine where something is, that is the location of the bubble. I used the magnitude channels of size for a quantitative variable (infected ratio) and colour luminance for an ordered data type (level of uncertainty), which are the

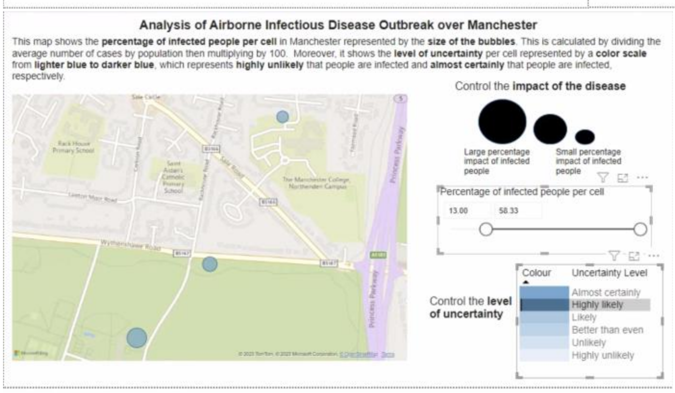
	correct match for each variable and they take into account the effectiveness rank.
Use of colour	To visualise the uncertainty in our data, “the more uncertain an estimate is, the more difficult it is to see becoming less visually prominent” (Yau, 2018). Thus, I am using the colour scale to achieve this by assigning a lighter colour to high uncertainty levels to be less prominent compared to low uncertainty levels which I used darker colours to make them more visible. I used colour luminance because the linear variation of lightness induces ordering. (Kovesi, 2015) The users for this case tend to be male dominant and the fact that around 8% of male are affected by colour blindness, I am using the blue-yellow channel with blue scale.
Use of interaction	By having the following interactions I have a low latency by giving quick feedback to the user when performing an action: a) Slicers to filter the percentage of infected people per cell b) Clickable legend table for the uncertainty levels c) Tooltips when you hover the bubbles. Includes information like location, infected ratio, uncertainty level. d) Zoom in and zoom out on the map.
Use of language and text	Use of a title for the whole visualisation. Brief explanation about the visualisation to understand the elements that are included. Use of legends for the infected ratio (bubble sizes). Use of legend for the level of uncertainty to avoid absolutes when describing the statistical value of sd. Use of titles for the slicers. Use of titles to control the impact and the level of uncertainty.
Reliability of operation, fit on desktop screen.	By selecting the desktop icon in Power Bi in the left down corner, we are ensuring that our visualisation fits a desktop screen and not a mobile screen.
Part Two	
Does the visualization allow an understandable overview of the situation?	I followed the same idea about representing the infected ratio in the bubbles size and using colour luminance for the level of uncertainty (the conditions for the scale are the same). In this part, I labelled each scenario by a letter and I encoded the different scenarios/wind conditions (categorical data) in colour hue, therefore, we are able to identify in the map how the wind direction affects different areas in the city. We have a scale of lighter colour to darker colour based on the colour hue for each scenario. I continued with the idea of being colour blind friendly so I chose the yellow-blue direction by using blue, magenta, orange and yellow colours. It is important to mention that I am showing the cells with the infected ratio greater than 3% to reduce noise in my visualisation of very low percentages.
Effective visual representation of the data variations over multiple runs	The short-term memory is very limited. Thus, I avoided cognitive overload by doing everything in one map and allowing the following interactions: a) Selecting one or multiple scenarios from a slicer or one single scenario from a bar chart (used size and position to encode quantitative and categorical data). b) Slicers to filter the percentage of infected people per cell and population per cell. c) Clickable legend tables for the uncertainty levels for each scenario

Screenshots

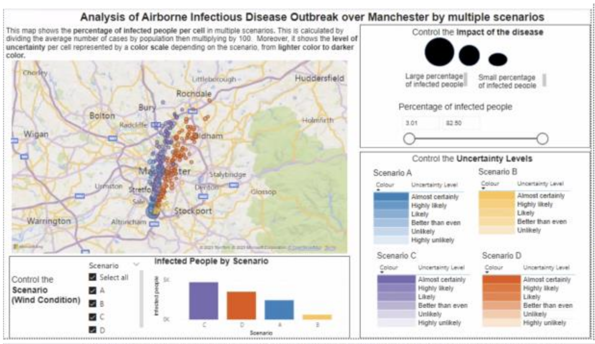
<u>Part 1</u>:

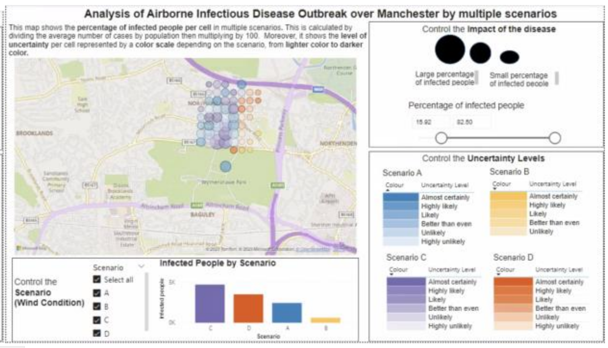


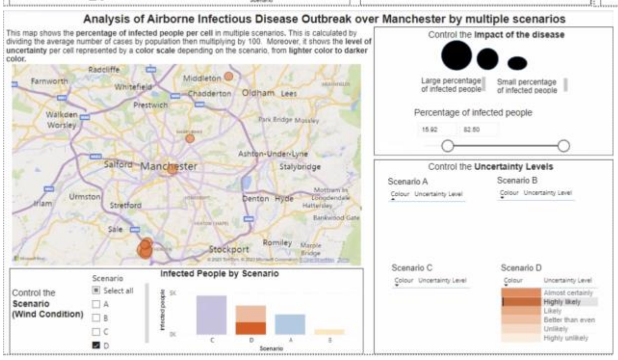




<u>Part 2</u>**:**







References

Munzner. (2015). *Visualization analysis & design* (Maguire, Ed.; p. 1 online resource (xxiii, 404 pages) ;). CRC Press.

Colin. Ware. (2013). *Information visualization : perception for design* (Third edition..).

Yau, Nathan. (2018, January 8). *Visualizing the uncertainty in data.* Flowingdata. https://flowingdata.com/2018/01/08/visualizing-the-uncertainty-in-data/

Kovesi, P. (205). *Good Colour Maps: How to Design Them.* Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia Crawley, Western Australia. https://s3.eu-central-1.wasabisys.com/binocularity/Teaching/DV2020_01/b1_teaching_day_two/05_VPP_Colour/CET_ColourMaps.pdf