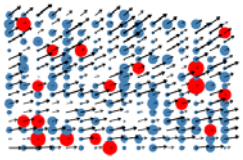


CSC8626 Data Visualization Summative Assignment

Preamble

Visualization has become a tool both for the exploration of raw data and for the presentation of analysed data to end users. In this assignment you are asked to represent data primarily to an end user who must make a decision using the data, but you may, as you develop your solution, use visualization to explore the data sets you have been given and you can report on this if you do.

You are asked to produce a visualization for a time critical decision maker, such as gold command police officers, an army commander or for politicians in a government COBR meeting. They need to be able to see clearly where an outbreak of an infection is estimated to have the most likely impact so that they can deploy limited tactical resources (medics and medical supplies) to best effect. Bear in mind that both level of impact and the level certainty influence the resource allocation decision.

There are several constraints in these and similar situations:

- You will typically not be there to explain the visualization, it must be standalone.
- Most, if not all, the decision makers will not understand statistical methods or mathematics.
- It may be important to print or fax your visualization, it should work on screen and on paper.

The data

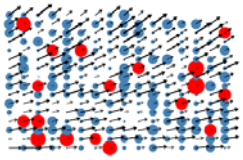
The data you have been given are outputs from a DSTL/PHE supercomputer simulation of an airborne epidemic outbreak over Manchester. Each simulation output (data file) simulates how the epidemic might spread given a certain set of environmental conditions, particularly wind direction and speed. The prevailing wind in all the simulations is coming from a direction roughly between west and south west, as it often does in Manchester.

Each data file holds outputs from groups of four simulation runs, each group of four has some similarities in initial environmental conditions. The simulation computes the infection outcome on a regular grid of cells across the city. Note that any cell in the simulation that has a zero output for all four simulations will have been excluded from the data file. The numbers given for the remaining cells are the number of expected infections in that cell for the simulation conditions in that run.

For each cell in each file you have:

- Longitude and latitude which is common to all simulation outputs the cell is included in.
- A unique cell ID which is common across all simulation outputs the cell is included in.
- The population of the cell.
- Four estimates of the number of infected people from each of the four simulation runs.
- The following derived statistics per cell across the four runs: mean, variance, standard deviation, index of dispersion and coefficient of variation. The latter two are standardised ways to help compare variability between cells that have different ranges of values.

The origin point of the epidemic for all the simulations is:
longitude: -2.2807386, latitude: 53.4034207



The assignment

Part 1 : Visualizing uncertainty and impact from a single file (70% of the mark for this assignment)

For this part of the assignment you must work only with `datafile_007.csv`

Your task is to create an interactive visualization that allows a decision maker to compare different areas of the city by the impact of the outbreak and by the (un)certainty of that impact. You are aiming to enable the viewer to understand the situation and then decide which areas of the city to target first with medical aid.

Your visualization should be designed to:

- Associate visual channels consistently with data variables.
- Only allow the interactions you intend to be allowed to happen.
- Apply Gestalt design principles, for example as expressed in the PARC guidelines.
- Consider the perceptual experience of colour scales and colour-blind viewers.

Your report should include a written description of the design and operation of your visualization and can include up to ten images of your visualization in the appendix, one of which should be black and white as if the visualization had been faxed to a field commander. References must be cited in a consistent and standard way.

Part 2 : Visualizing variation across multiple scenarios (20% of the mark for this assignment)

For this additional part of the assignment you can work with any number of the datafiles.

Your task here is to demonstrate the impact and (un)certainty of the scenarios across multiple simulation runs. Again, you are aiming to help the decision maker decide which areas of the city it is most important to target with aid first. But now there are up to one thousand different variations in the wind speed and direction (four per data file). You might be able to do this with your solution to part 1 but it seems likely you will need an overview visualization of some kind.

As above you should write up the design and operation of your visualization and you can include up to ten images of your visualization in the appendix, one of which should be black and white as if the visualization had been faxed to a field commander.

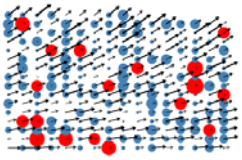
What to submit

A two part report of approximately 2000 words and no more than ten images of each of your visualizations (this word limit does not include your references and diagrams). In pdf or word format.
Your PowerBI visualizations as a standalone pbix file (or files).
Please also upload your visualization to the online PowerBI group CSC8626_2018_SA1 so we can view it easily during the viva.

The deadline for all submissions is 16:00 on Friday 26th October.

Viva arrangements

There will be individual 20-minute viva on Thursday 25th October in the IDT. A sign up poll will be sent out on the 22nd, if you are a part time student and cannot make this day please contact me.



Use of software tools

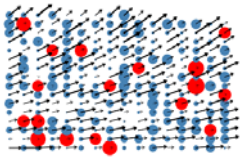
You must use PowerBI as the visualization tool reading input data from one more spreadsheets in CSV or Excel format. You may use any data pre-processing tools that you find helpful (for example R or Python) but the visualization(s) you submit must be created in standalone PowerBI reading from CSV/Excel files. If you download and use any PowerBI extensions, e.g. from the community marketplace, these must be noted and referenced.

Sources of further information beyond that already covered in the course

https://flowingdata.com/2018/01/08/visualizing-the-uncertainty-in-data/

https://flowingdata.com/tag/uncertainty/

https://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-11-37



The marking scheme

In order to gain marks you must demonstrate in the report your application of visualization skills and techniques against the following marking scheme.

Part One	**Mark**	**Feedback & how to improve.**
Fit to task: does the visualization allow the identification of areas most and least in need of aid.	/10	
Use of visual channels	/10	
Gestalt design principles	/10	
Use of colour scheme	/10	
Use of interaction	/10	
Use of language and text	/10	
Technical aspects: reliability of operation, screen size fit.	/10	
Total for part 1	**/70**	
Part Two		
Fit to task: does the visualization allow the identification of areas most and least in need of aid.	/10	
Effective visual representation of variation over multiple runs.	/10	
Total for part 2	**/20**	
Report		
Logical structure, quality of technical writing, labelling and relevance of figures, range and quality of referencing.	/10	
Total for assignment	**/100**	