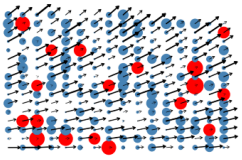


Outbreak

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 your analysis of data about a medical emergency in a city to an end user who must make a rapid decision using the data.

Assume your end user is a time critical decision maker, such as a gold command police officer, an army commander or a politician in a government COBR meeting. They are trying to decide where to deploy limited tactical resources (medics and medical supplies) based on your visualization. Keep in mind they need to see **predicted impact *and* level of uncertainty** of that impact to make a decision.

There are several constraints in these and similar situations:

- You will typically not be there to explain the visualization, it must be entirely 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.

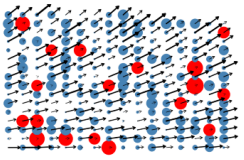
The data
The data you have been given are a set of outputs from a DSTL/PHE supercomputer simulation of an airborne infectious disease outbreak over Manchester. Each simulation output (data file) simulates how the epidemic might spread given a certain set of environmental conditions with varying wind 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. The infected number given for each cell is the number of expected infections in that cell for the simulation conditions in that run. Note that any cell in the simulation that has a zero output for all four simulations will have been excluded from the data file.

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 uncertainty 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 (60% of the mark for this assignment)

For this part of the assignment you must work only with `datafile_014.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 uncertainty of that impact. Your aim is 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 consist just one PowerBI report page 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.

Part 2 : Visualizing variation across multiple scenarios (30% of the mark for this assignment)

For this additional part of the assignment you can work with any number of the datafiles and all 1000 simulation outputs (250 files x 4 outputs per file).

Your task here is to demonstrate the impact and uncertainty 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 is possible you will need an overview visualization of some kind.

Your part 2 visualization should consist of no more than one PowerBI report page.

What to submit

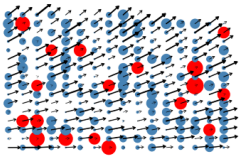
a) Please fill in the description sheet sections for both parts of the assignment. This can be in note form and should in total be between one and no more than three pages long. You should add a list of references to articles or other sources of information you have used to produce the visualization. Please submit this as a pdf file.

b) Your PowerBI visualization(s) as a standalone pbix file (or files).

You will need to create one zip archive file of all the files you want to submit, please make sure the file name includes your name and/or student number.

This submission should be done via NESS the online submission system linked to from Canvas.

The deadline for all submissions is 16:00 on Friday 22nd October



Use of software tools

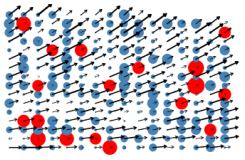
You must use PowerBI as the visualization tool reading input data from one or more spreadsheets in CSV or Excel format. You may use any data pre-processing tools that you find helpful (for example R or Python) but the visualization(s) you submit must be created in standalone PowerBI reading from CSV/Excel 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	/9	
Gestalt design principles	/9	
Use of colour	/9	
Use of interaction	/9	
Use of language and text	/9	
Technical aspects: reliability of operation, fit on desktop screen.	/5	
Total for part 1	**/60**	
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.	/20	
Total for part 2	**/30**	
Report		
Logical content structure, range and quality of references used.	/10	
Total for assignment	**/100**	