Cartogram Mapping and its Application to Cancer Data Visualisation

Stephanie Kobakian, Jessie Roberts and Dianne Cook

Abstract

Utilising maps to present statistics has been widely used for centuries. Connecting the data to the geographical representation of areas that are already familiar. It is not enough to be recognisable, if the value of the statistic cannot be seen. Cancer outcomes directly relate to the people living within a geographical area. In situations where the people are of interest, not the land they live on, it is reasonable to explore views that enhance the communication of the cancer outcomes. This has spurred innovations over the previous centuries to enhance the shapes and maps presented to effectively communicate cancer outcomes.

Disease mapping

Choropleth map (Steff)

A choropleth map is used to show the spatial variability of measurements, with usage dating back to the 1800s. They are constructed by drawing administrative boundaries, and filling the polygon with colour to represent values of a measured variable. Early versions used symbols or patterns instead of colour. Bell et al. (2006) discusses the use of choropleth maps for visualising cancer data, and Walter (2001) gives an overview of the development of the use of choropleth maps for displaying disease data. Figure 1 shows a choropleth map on XXX. What can be seen. A choropleth map is a true map, constructed for visual inspection of spatial patterns across a familiar geographic form, that might show trends in disease occurrence, or even localised outbreaks.

< Figure 1 here: choropleth map >

The familiar shapes help viewers to visually infer spatial relationships, and hence intuitive to read (Brewster and Subramanian 2010). However, the different sizes of administrative areas can lead viewers to place more attention on the shades of the largest areas. The inferences derived, and the hypotheses developed can suffer from a large area "bias".

In epidemiology, choropleths are often used as a mapping tool in the study of the spatial distribution. Almost 100 years of cancer maps have explored the US and the UK with increased effectiveness in presentation of unbiased rates. Howe (1989) considered maps a research tool, dots maps use spots, choropleths and thematic maps shade areas and isopleths use isolines. These techniques were commonly used before the advent of computer-assisted cartography. The data collection methods of cancer mortality rates across regions, and administrative control within regions lends itself choropleth visualisation. Mortality rates are now often presented as relative rates of risk across the population, and age adjusted to correct for the proportion of old people. Howe (1989) describes Stocks development of the standardised mortality ratios through the 1930s. The choropleth maps presented levels of cancer via hatchings on a black and white scale.

Walter (2001) cites Cruickshank's (1947) discussion of using visuals as 'formal statistical assessment of the spatial pattern' as a major advancement. Kraak (2017) considers cartographic data analysis in practice. Presenting a clear guide for preparing cartograms for use with both qualitative and quantitative information. The authors recognise that the creation of cartograms was largely the work of professional cartographers until the innovation of geographical information systems that welcomed map users as map creators. These systems are utilised depending on 'the effectiveness, efficiency, and satisfaction of the map products (Nielsen 1994)'. Exeter (2017) reports the increase in disease mapping development due to the availability of geographic information system softwares.

Choropleths can inhibit visual inference when presenting human related statistics as the display may draw attention from the 'potentially more important results in the more populous communities' (Exeter 2017).

Examples of atlases: Kraak 1998, Kraak and Ormeling 2011 Bertin 1967

Moore and Carpenter (1999) suggests it is the "investigators' objectives" that drive the "representation of diseases on maps". There will be situations like infectious diseases that require chоropleth methods

Public face (Jessie)

what has been provided to the public, norms, common practices, all of Jessie's lit review

Cartograms (Steff)

Cartograms are a thematic map, presenting statistical and geographical information like choropleths. Dorling (2011) suggests the difference is the intentional and desirable augmentation of the size, shape or distance of geographical areas. World projections reflect the frequent distortions seen from altering perspectives. Choropleth maps will always be distorted if they cover enough of the globe, as will photographs of the globe from space. Cartogram creation requires choosing a favourable distortion of the geography for presenting the set of spatial information. This concept was originally implemented to allow a map base that represented the proportion of the population within each region. Dorling (2011) considers this design a 'more socially just form of mapping' by giving all members of the population 'equitable representation'. This is known as an iso-demographic map, and the design inherently lends itself to epidemiolgy. Howe (1989) argues that 'cancer occurs in people, not in geographical areas' and the map base should reflect this to avoid allocating 'undue prominence' to rural areas. Walter (2001) reduce the visual impact of large areas with small populations, the distortion of spatial areas is proportional to the denominator variable.

Alternatively, the area of the map space can represent the value. There have been many algorithms presented, Nusrat and Kobourov (2016) provided a framework to investigate implementations and the "statistical accuracy, geographical accuracy, and topological accuracy".

Howe (1989) shows the introduction of electronic computer-assisted techniques created a flood of disease atlases like Howe's National Atlas of Disease Mortality in the United Kingdom (Howe 1963) which upgraded to a demographic base "map" (Howe 1970).

The presentation of small areas requires more thought during the implementation of maps, Jahan et al. (2018) encourage their use to uncover local-level inequalities frequently masked by health estimates from large areas.

< Figure 2 here - basic cartogram - easy to see >

Contiguous

Dorling (2011) presents the three methods for creating contiguous cartograms. John Hunter and Johnathan Young (1968), and Durham used phsyical accretion models, arranging wooden tiles by hand. Tobler (1973) used a computer programs. Skoda and Robertson (1972) developed a mechanical model utilising steel ball bearings.

Cartograms keep spatial relationships of neighbours intact by preserving borders when adjusting sizes. Tobler's Conformal mapping means to preserve angles locally so that the shapes of very small areas on a traditional map and a cartogram would be similar From a computer graphics perspective it is a problem of 'map deformation' to account for the value assigned to each area. Min Ouyang and Revesz (2000) discuss their implementation of three methods for creating value-by-area cartograms.

Their intention is to allow the map space to highlight the distribution of the variable. However a reader may have to know the difference between initial geography and new layout given by a cartogram, to be able to recognise the significant changes.

Australia (McGlashan 1977),

Dorling (2011) reiterates: 'There is no "best" cartogram or method of creating cartograms just as there is no "best" map' (Monmonier and Schnell, 1988). There are many alternatives to consider, the intended audience of the map, and its purpose are key points in cartogram use and creation.

< Figure 3 here>

Non-Contiguous

Dorling (2011) puts forward a simple question: >If, for instance, it is desirable that areas on a map have boundaries which are as simple as possible, why not draw the areas as simple shapes in the first place?

He answers this with his implementation of maps created with 'the simplest of all shapes'. While contiguous cartograms may be a 'more sophisticated' method, they produce 'very complex shapes'. Circular cartograms use the same shape for every region represented, and size them according to the population, or statistic represented. To produce a compelling map, a gravity model is applied to avoid overlaps, and keep spatial relationships with neighbouring areas over many iterations. This implementation can work for up to 'one hundred thousand' areas.

Keim et al. (2002) also present the problem of maintaining shape presevation in non-contiguous cartograms.

Dorling (2011) suggests 'population distribution is often extremely uneven in former British colonies'.

'In Australia the urban federal constituencies occupy only a tenth of the land, but contain nine tenths of the people. It would be almost unthinkable to show election results for that country on a conventional equal land area map.' This 1966 cartogram uses mostly straight lines, and the result looks very little like the geographical shape of Australia.

‘Given the increasingly uneven population distribution of the United States and the growing social divides between the populations of neighbourhoods living at different densities, the need for cartograms like this is greater now than ever.’

< Figure 3 here>

Used in displays of the UK by Howe in 1986 cited by Howe (1989)

Tobler’s method and the many implementations that ‘elaborated’ on it are derived from ‘numerical approximations to a pair of equations’(Dorling 2011). They all operate through incremental adjustments, and can produce wildly different outcomes from small changes in the inputs.

Tobler (2004) Value-Area Cartograms. In these cartograms a region,country, or continent is subdivided into small regions, each ofwhich is represented by a rectangle. This rectangle is proportionate in area to the value which it represents in certain statistical distributions. The regions are grouped in approximately the same positions as they are on the map.

Computer generated map examples: Howe (1989) (Hopps et al. 1968; Armstrong 1972). There has followed a flood of disease atlases, mainly concentrating on the modem problems of cancer and degenerative diseases from countries as scattered as the United States (Burbank 1971; Mason et al. 1975, 1976; Pickle et al. 1987), the Soviet Union (Levin 1980), Japan (Shigematsu 1977), the Federal Republic of Germany

Cano et al. (2015) define the term ‘mosaic cartograms’. Compare amount of tiles to contrast population of regions. ‘Cartograms show a data value per input region by scaling each region such that its area is proportional to its data value. Mosaic cartograms show data in multiples of tiles, hence the input data must consist of, or be cast into, small integer units.’

Centroid displays

Dot plot: one dot for each region, coloured, and placed at centroid.

Plotting centroids on top of geographies. (Size is kept constant) Hex maps

If the goal of the map is the statistic.

If the distribution is the focus, the display should reflect that.

Replies on the idea that every area is important, no matter the size.

Recent methods have every area represented with the same map space.

This gives equal emphasis to every area, allows distributions and relationships between neighbours to become more clear.

A critique choropleth and cartograms

> designing a map tailored to precise goals [is] easier than forcing a single map to accommodate diverse objectives - Bell et al. (2006)

With that in mind, the intented user and message to communicate should drive map selection.

‘Where control of the message is important, static maps will continue to be the most effective, although good tables, graphs, and explanatory text are still needed in order to ensure that different people will see the same thing in the maps’ Bell et al. (2006)

(only worth including if it is possible for us to implement) Tabular form comparing and contrasting - Relationship to geography

- Show using cancer examples

Animation

lends to temporal pattern exploration

Keim et al. (2002) ?? Highlight the value of animating contigous to see changes over time, US can be recognisable but animation aides interpretation

Acknowledgements

What software did we use, for eg

References

Bell, B Sue, Richard E Hoskins, Linda Williams Pickle, and Daniel Wartenberg. 2006. “Current Practices in Spatial Analysis of Cancer Data: Mapping Health Statistics to Inform Policymakers and the Public.” *International Journal of Health Geographics* 5: 49. https://doi.org/10.1186/1476-072X-5-49.

Brewster, Mark B., and S. V. Subramanian. 2010. “Cartographic Insights into the Burden of Mortality in the United Kingdom: A Review of ‘The Grim Reaper’s Road Map’.” Journal Article. *International Journal of Epidemiology* 39 (4): 1120–2. https://doi.org/10.1093/ije/dyp395.

Cano, R. G., K. Buchin, T. Castermans, A. Pieterse, W. Sonke, and B. Speckmann. 2015. “Mosaic Drawings and Cartograms.” *Computer Graphics Forum* 34 (3): 361–70.

Dorling, Daniel. 2011. “Area Cartograms: Their Use and Creation.” In *Concepts and Techniques in Modern Geography (CATMOG)*, 59:252–60. https://doi.org/10.1002/9780470979587.ch33.

Exeter, Daniel J. 2017. “Spatial Epidemiology.” Journal Article, 1–4. https://doi.org/10.1002/9781118786352.wbieg0283.

Howe, G. M. 1989. “Historical Evolution of Disease Mapping in General and Specifically of Cancer Mapping.” In *Cancer Mapping*, edited by Peter Boyle, Calum S. Muir, and Ekkehard Grundmann, 1–21. Berlin, Heidelberg: Springer Berlin Heidelberg.

Jahan, Farzana, Earl Duncan, Susanna Cramb, Peter Baade, and Kerrie Mengersen. 2018. “Making More of Spatial Maps: A Bayesian Meta-Analysis Approach.” In.

Keim, D.A, S.C North, C Panse, and J Schneidewind. 2002. “Efficient Cartogram Generation: A Comparison.” In *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002*, 2002-:33–36. IEEE.

Kraak, M. J. 2017. "Cartographic Design." In *The International Encyclopedia of Geography : People, the Earth, Environment, and Technology*, edited by D. Richardson, 1–16. United States: Wiley.

Min Ouyang, and P. Revesz. 2000. "Algorithms for Cartogram Animation." In *Proceedings 2000 International Database Engineering and Applications Symposium (Cat. No.PR00789)*, 231–35. https://doi.org/10.1109/IDEAS.2000.880581.

Moore, Dale A., and Tim E. Carpenter. 1999. "Spatial Analytical Methods and Geographic Information Systems: Use in Health Research and Epidemiology." *Epidemiologic Reviews* 21 (2): 143–61. https://doi.org/10.1093/oxfordjournals.epirev.a017993.

Nusrat, Sabrina, and Stephen G. Kobourov. 2016. "The State of the Art in Cartograms." *CoRR* abs/1605.08485. http://arxiv.org/abs/1605.08485.

Tobler, Waldo. 2004. "Thirty Five Years of Computer Cartograms." *Annals of the Association of American Geographers* 94 (1): 58–73.

Walter, S. D. 2001. *Disease Mapping: A Historical Perspective.* Book. Oxford University Press: Oxford.