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.

Introduction

Disease mapping methods

Current best practice map displays for cancer data

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 where each state of the United States of America is coloured by the average annual rate of new cases of lung and bronchus cancer from years 2012 to 2016. A choropleth map is a true map of the topology, constructed for visual inspection of spatial patterns across a familiar geographic form, that might show trends in disease occurrence, or even localised outbreaks.

The familiar state 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 tool to study the spatial distribution of cancer incidence and mortality. Almost 100 years of cancer maps have explored the United States and the United Kingdom with increased effectiveness in presentation of unbiased rates. The increasing development and use of disease maps can be attributed to the availability of geographic information system softwares (Exeter 2017). 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

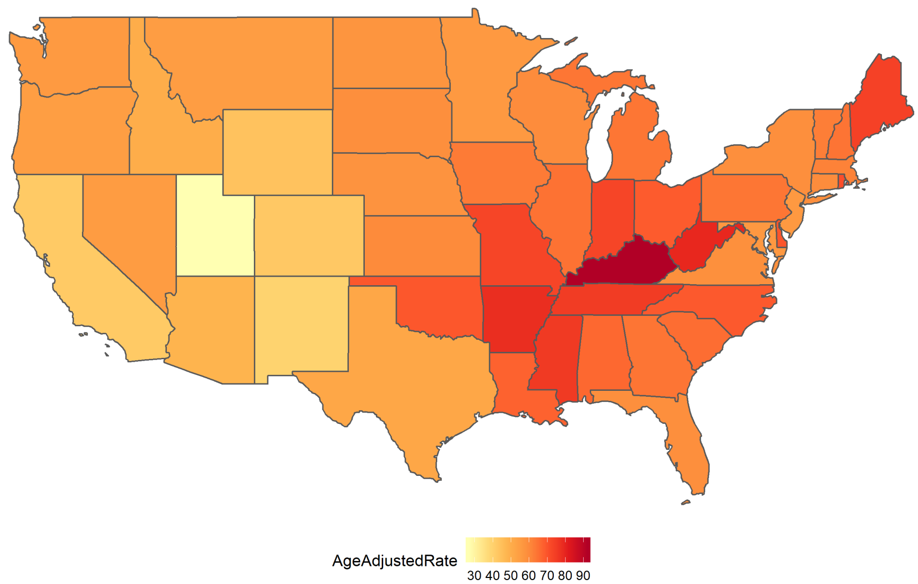

Figure 1: "A choropleth map of the United States of America."

ratios through the 1930s. The choropleth maps used presented levels of cancer via hatchings on a black and white scale.

Bell et al. (2006) The potential audience for the results of a spatial analysis of health data is no longer limited to scientists but now also includes the public, policymakers, the media, and a host of others.

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)'.

Using the recognisable geography, that is familiar, helps communicate the information. –>

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 choropleth methods

Public face (Jessie)

A grey literature review conducted by (ref) identified 33 Cancer Atlases publically available cancer atlases available on the internet and published between 01/01/2010 to 11/11/2015.

All the maps identified in this review were chloropleth maps, in which the representation of the map geometry stayed true to the underlying geography.

Publishers

The majority of atlases were published by non-commercial organisations, including not-for-profits (NFPs), government, research organisations, advocacy groups or a partnership between an NFP & government. Only one map was published by a commercial entity (Maps of Cancer Mortality Rates in Spain), in this case a media organisation El Pais[1].

Reported measures

The majority of maps identified in this study reported age adjusted rates of either incidence, mortality or survival. The list below summarises these measures, and provides a definition of each measure.

Geographical coverage

Identified cancer atlases covered geographies from all around the world: 4 were global, 3 from Australia (AUS), 11 from the United States (US), 2 from Canada (CAN), 7 from the United Kingdom (UK), 2 from Spain, 1 from Switzerland, 1 from Germany, 1 from Norway, and 1 covering the European Union. Not all maps had a national focus and 10 covered a region or state rather than an entire nation. The states or counties/regions covered were South Australia (AUS), Queensland (AUS), Ontario (CAN), Valencia (Spain), Pennsylvania county Massachusetts (US), New Hampshire (US), Cape Cod (US), Missouri (US), Florida (US), New York State (US) and Arizona (US).

[1] https://elpais.com/

Table 1: (#tab:measures) Measures used to report cancer statistics

Measure	Details
1. IR (Incidence Ratio)	$(IR)_i = \frac{(Incidence\ Rate)_i}{Average\ Incidence\ Rate}$, Cancer incidence rate in region i over the average cancer incidence rate for the total region
2. SIR (Standardised Incidence Ratio)	IR standardised by age structure in each region i
3. RER (Relative Excess Risk)	$RER = \frac{(Cancer\ related\ mortality)_i}{Average\ cancer\ related\ mortality}$ Represents the estimate of cancer related mortality within five years of diagnosis Also referred to as 'excess hazard ratio'
4. Age Adjusted Relative Risk	RR standardised by age structure in each region i
5. Rate per 100,000	Cancer incidence per 100,000 population
6. Age Adjusted Rate per 100,000	#5 standardised by age structure or region
7. New cancer cases per 100,000	Specific methods could not be found
8. Count	Crude cancer counts
9. Below or above Expected	Alternative expression of the SIR

Statistical uncertainty

Cancer atlases were considered to report uncertainty to the non-expert user if they included a measure of statistical uncertainty either within or alongside the map. Maps that only reported this information within the supplementary material were not considered to have directly attempted to report uncertainty.

The review did not reveal any novel uncertainty visualisation approaches or visualisations. Maps used standard and well known measures including credible intervals and standard deviation, statistical significance, boxplots and distributions. These maps ranged from static pdfs or infographics to interactive online resources. The interactivity of the more modern maps enabled uncertainty information to be incorporated without cluttering the screen, such as in a tool tip feature.

Close to half of the atlases identified (42%, n=14) included some measure of uncertainty. The most common measure used to represent uncertainty were credible or confidence intervals (CIs). CIs were either visualised by including their bounds in a scatterplot or graph of estimates vs region (see Figures @ref(fig:ci-viz2)[2], and @ref(fig:ci-viz3)[3] positioned next to the map, or reported numerically through the CI upper and lower bounds listed in a data table (see Figure 3.4[4]). Of those that

[2] Alberta Health IHDA Geographic. (2012). Age-Standardised Incidence Rate of COPD, 2011. Retrieved from: http://www.health.alberta.ca/health-info/IHDA-geographic/COPD/incidence-agestandard/atlas.html?epik=0GJSpE_IW34lx

[3] Centres for Disease Control and Prevention (CDC). (n.d). United States Cancer Statistics: An Interactive Cancer Atlas (InCA). Retrieved from:https://nccd.cdc.gov/DCPC_INCA/

[4] Pennsylvania Cancer Atlas. (n.d). Retrieved from: https://www.geovista.psu.edu/grants/CDC/?epik=0lJSpE_IW34lx

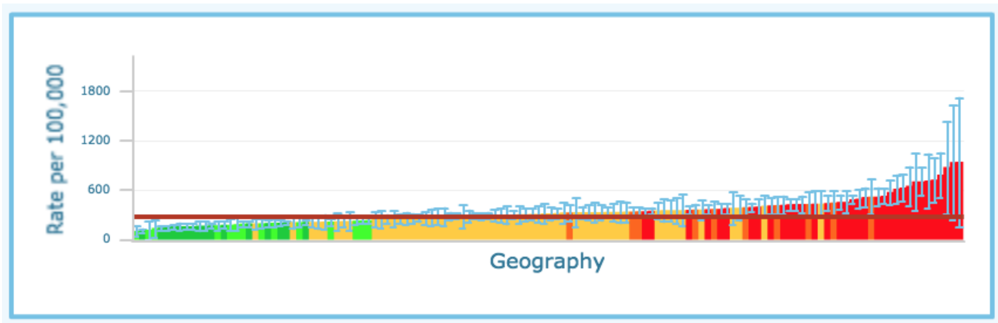

Figure 2: Example of CI visualisation for uncertainty representation in cancer mapping (1/3). Source: Alberta Health IHDA Geographic. (2012)

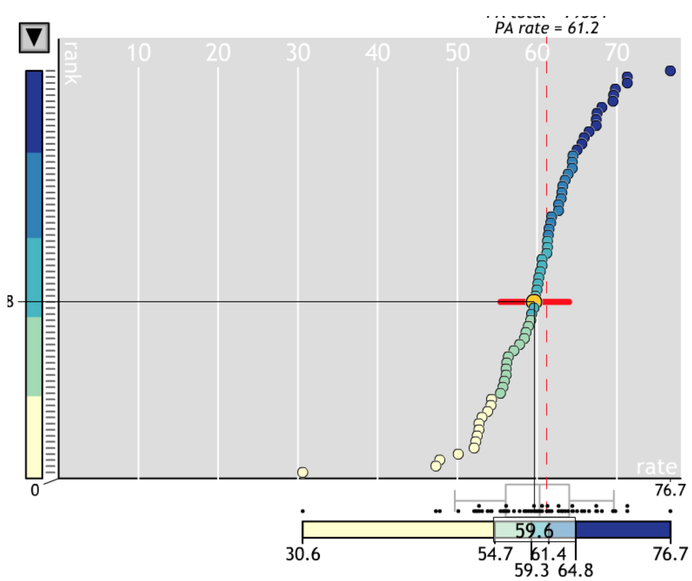

Figure 3: Example of CI visualisation for uncertainty representation in cancer mapping (2/3). Source: Pensylvania Cancer Atlas

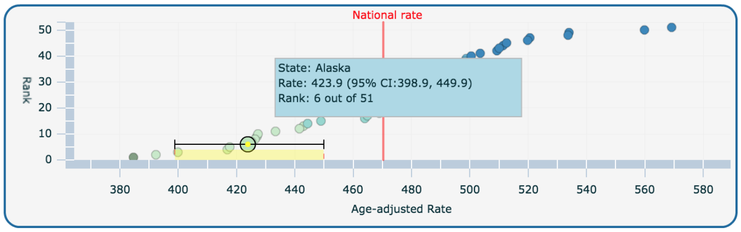

Figure 4: Example of CI visualisation for uncertainty representation in cancer mapping (3/3). Source: Centres for Disease Control and Prevention (CDC). United States Cancer Statistics: An Interactive Cancer Atlas (InCA)

Data Table | Glossary | Further Info

County	Rate	Lower CI	Upper CI	Count	Population
Adams	60.4	55.3	65.9	518	801,717
Allegheny	61.4	60.1	62.7	9,390	11,676,746
Armstrong	58.9	53.8	64.4	506	656,638
Beaver	56.4	53.3	59.7	1,247	1,654,797
Bedford	52.3	46.3	58.8	287	446,092
Berks	60.1	57.6	62.6	2,287	3,306,811
Blair	67.4	63.2	71.7	1,012	1,171,282
Bradford	54.3	48.8	60.3	363	563,934
Bucks	61.3	59.1	63.4	3,174	5,298,961
Butler	61.5	57.8	65.4	1,042	1,537,647
Cambria	65.9	62.3	69.7	1,291	1,400,022
Cameron	76.7	57.7	102	57	54,504
Carbon	65.0	59.2	71.4	472	530,218
Centre	52.1	47.6	56.9	502	1,214,956
Chester	56.2	53.8	58.7	2,015	3,797,562
Clarion	64.5	57	72.8	273	376,923

Figure 5: Example of an interactive data table with CI upper and lower bounds used in cancer mapping. Source: Pennsylvania Cancer Atlas

visualised the CIs, 30% (n=10) embedded the visualisation within a tool tip function which visualised the CI when the mouse hovered over the relevant area (see Figure 3.5[5]).

Methods for representing sources of uncertainty information can be visualised or communicated in different ways, examples identified through this grey literature review are listed below.

Table 2: (#tab:method-exp) Implicit and explicit measures of uncertainty.

Measure	Example
CI Interval	Figures 3.1, 3.2, 3.3
Statistical Significance	Textured overlay on top of coloured regions used to indicate statistical significance
Distribution	Figure 3.4
Boxplots	Figures 3.4, 3.5
Sample Size	Textured overlay or lack of colour on a region, was used to show regions with small sample size

[5] International Agency for Research on Cancer. (2017). Atlas of Cancer Mortality in the European Union and European Economic Area 1993-1997, Annex 4 - Cancer mortality maps by site. Retrieved from: http://www.iarc.fr/en/publications/pdfs-online/epi/sp159/

Measure	Example
Standard deviation	Figure 3.6 - the second map in the bottom right corner shows standard deviation

Cartograms (Steff)

Cartograms aand choropleths both present statistical and geographical information. Dorling (2011) suggests the difference is the intentional and desirable augmentation of the size, shape or distance of geographical areas. Waldo Tobler (2012) explores many graphical techniques, and suggests there are particular methods for particular purposes.

Some consider choropleths true topological maps, however, if the land mass presented covers enough of the globe, there must be a transformation to visualise it in 2D. The 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. Choropleth creation requires choosing a favourable distortion of the geography for presenting the set of spatial information.

Cartograms may be seen as an extension of this concept, implementing a distorion of the geographical shape to allow a map base that represents 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'. Walter (2001) reduce the visual impact of large areas with small populations, the distortion of spatial areas is proportional to the denominator variable. 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.

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.

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

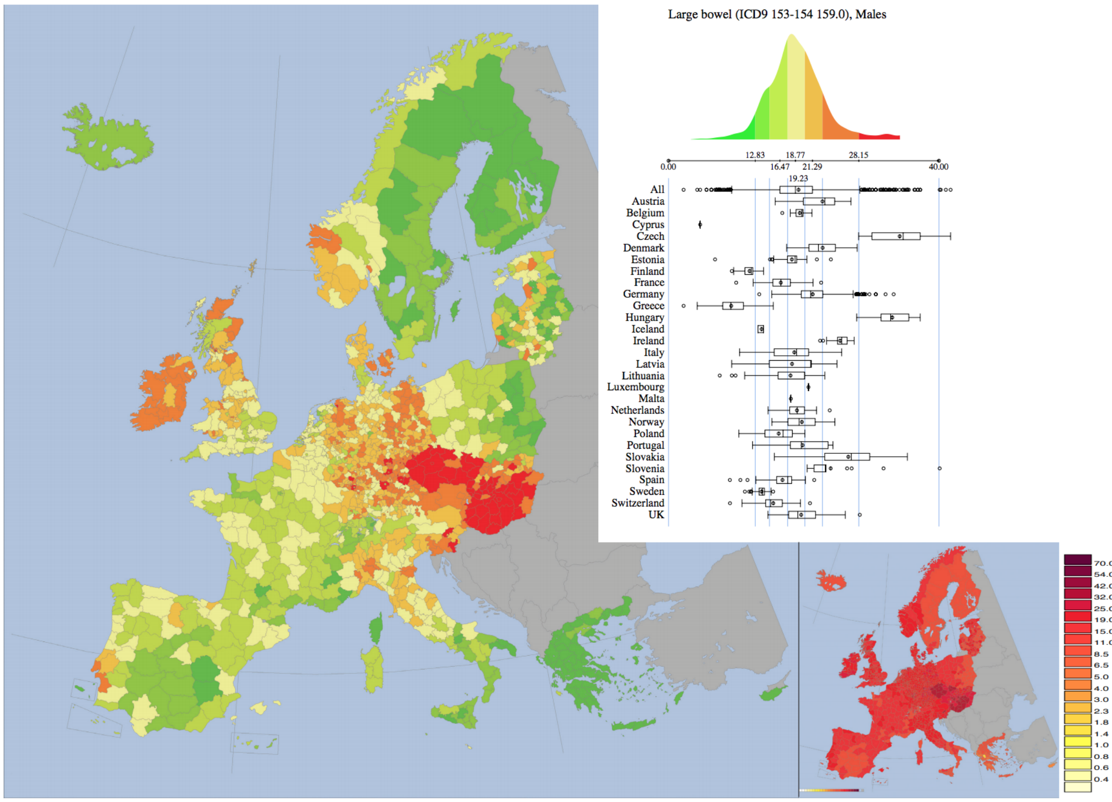

Figure 6: Example of standard deviation visualised in cancer mapping

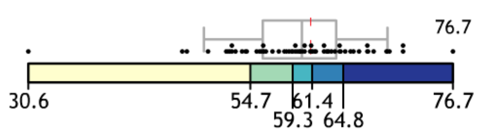

Figure 7: Example of boxplot used in cancer mapping. Source: Source: Pensylvania Cancer Atlas

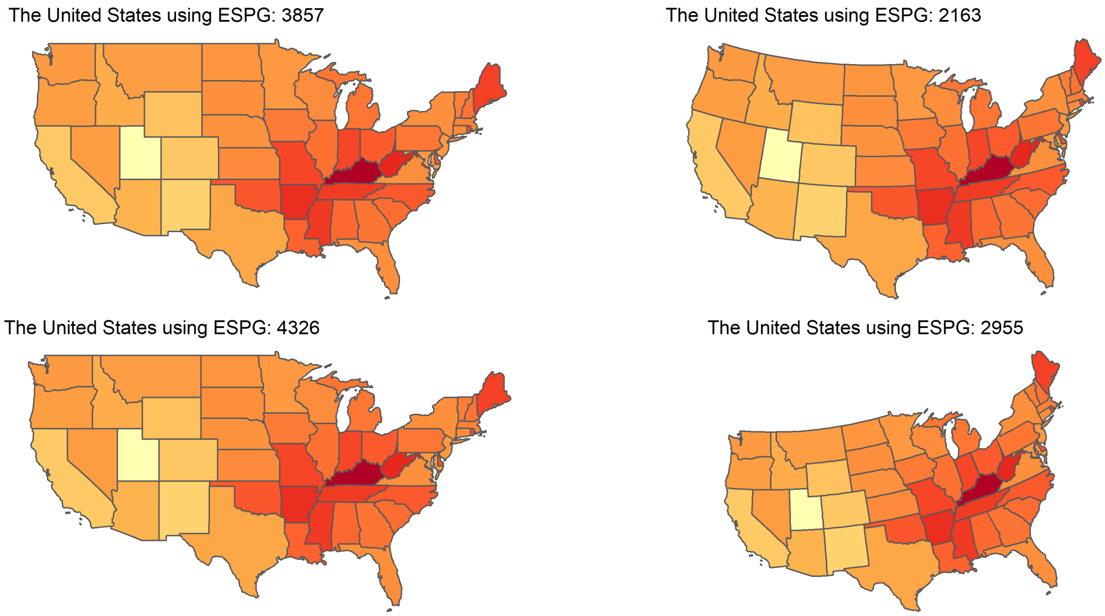

Figure 8: “Choropleth maps of the United States of America using four coordinate reference systems.”

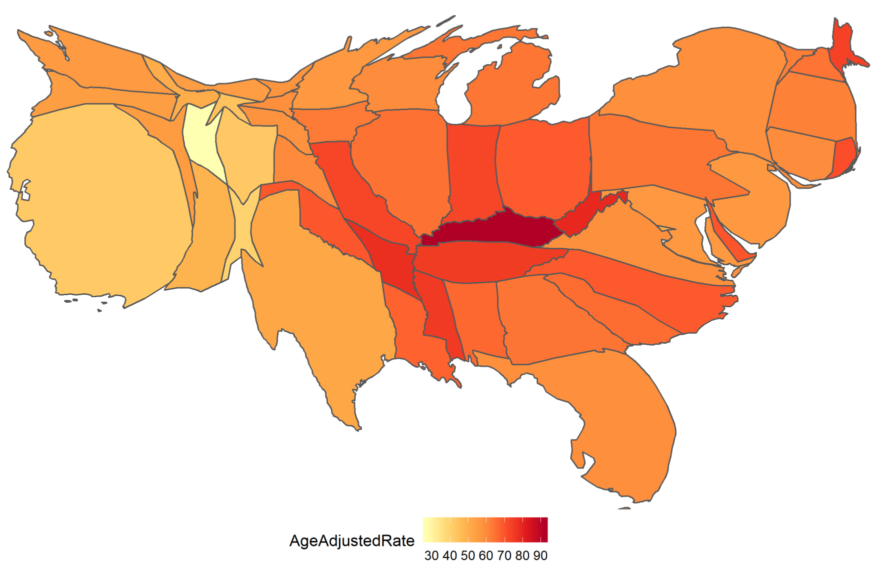

Figure 9: “A contiguous cartogram of the Unites States of America”

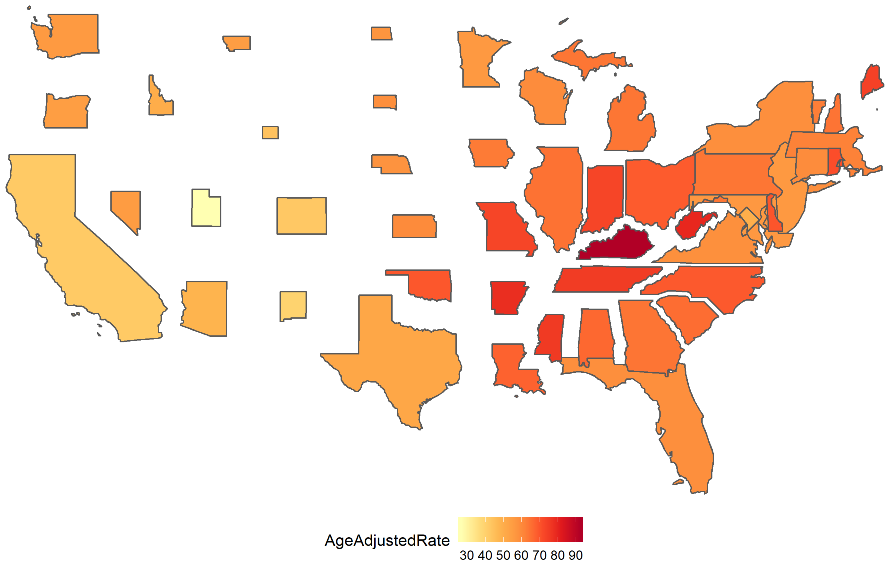

Figure 10: "A Non - contiguous cartogram of the Unites States of America"

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.

Non-Contiguous

Dorling 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.

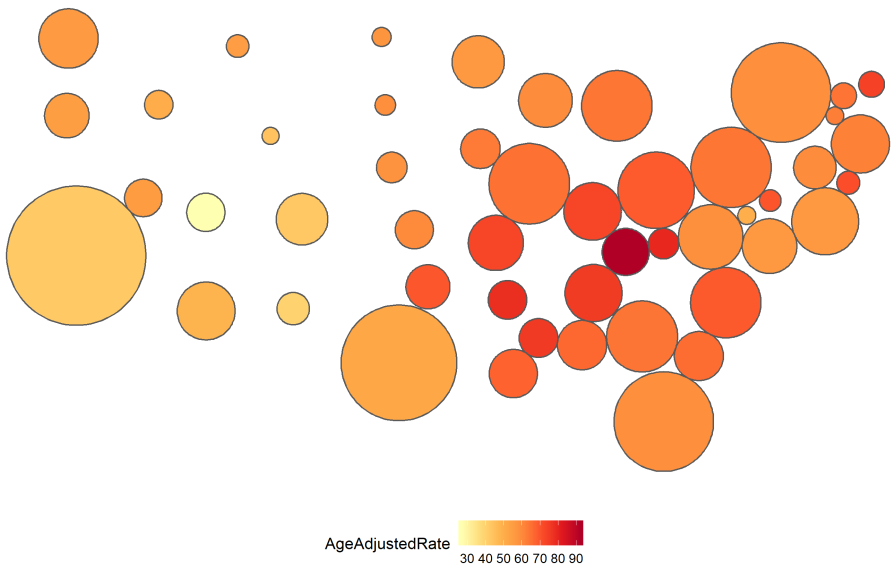

Figure 11: "A dorling cartogram of the Unites States of America"

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. doi: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. doi: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. doi:10.1002/9780470979587.ch33.

———. 2012. *The Visualisation of Spatial Social Structure.* John Wiley & Sons Ltd.

Exeter, Daniel J. 2017. "Spatial Epidemiology." Journal Article, 1–4. doi: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. doi: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. doi: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). Taylor & Francis Group: 58–73.

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