

Cartogram Mapping and its Application to Cancer Data Visualization

Stephanie Kobakian, Jessie Roberts and Dianne Cook

Contents

1	Introduction	2
2	Map displays for disease data	3
3	Cancer atlases	3
3.1	Overview of publicly available atlases	5
3.1.1	The Environment and Health Atlas of England and Wales	5
3.1.2	Globocan 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide	6
3.1.3	Atlas of Cancer in Queensland	6
3.1.4	Bowel Cancer Australia Atlas	6
3.1.5	United States Cancer Statistics: An Interactive Cancer Statistics Website	6
3.1.6	Map of Cancer Mortality Rates in Spain	6
3.1.7	Atlas of Childhood Cancer in Ontario	6
3.2	Additional considerations	6
3.2.1	Geographic hierarchies	7
3.2.2	Population distribution	7
3.2.3	Statistical uncertainty	7
3.2.4	Demographics	7
3.2.5	Socio-economic indicators	7
4	Alternatives to choropleths	8
4.1	Cartograms	8
4.1.1	Contiguous	10
4.1.2	Non-contiguous	10
4.1.3	Dorling	12
4.2	Tile map	12
4.3	Geofaceting	13
4.4	Multivariate displays	13
5	Comparison of maps and cartograms	14
5.1	Neither choropleths or cartograms perform well for Australia	15
6	User interaction	17
6.1	Interaction and animation in publicly available atlases	17
6.2	Under-utilised techniques	20
7	Conclusions	20
8	Acknowledgements	20

Abstract

Choropleth maps are widely used in cancer atlases, and traditionally communicate cancer statistics over geographic domains. The type of statistic displayed varies immensely, and effective choices can assist in the understanding of impact of cancer on a population. Atlases have moved from print media to web format, and the availability of graphical user interfaces, and user interaction and animation, greatly enhance the ability to communicate more information and positively engage readers. More use of alternatives to choropleth maps, such as cartograms, could be employed in online atlases, for better representation of spatial distributions, and show statistics on small areas. This is especially true for the geography of Australia, where the population is concentrated in small areas along the coastlines.

1 Introduction

Cancer statistics are usually delivered as an aggregated value for a geopolitical area. Presenting these statistics requires transforming individual observations into aggregations of communities as geographical units, in large part for privacy protection or political and policy purposes. The information could be as simple as counts per area (e.g. state, province, local government area, post/zip code). Counts alone are not sufficient to compare areas, because the populations of areas are all different. In this case, the counts data needs to be merged with population data to appropriately calibrate it to incidence, for example, rate per 100000 people. This type of data is collected on a routine basis for public health purposes, and may be made available to the general public as a service to the community. The task, then, is to examine what are the usual ways to communicate cancer statistics, to the public, are there alternative approaches, and what are the pros and cons of these choices.

A common approach to communicate cancer statistics, is to display statistics on a map. Using a choropleth map: the statistic is mapped to color and the geographic region is filled with this color. The viewer would then be able to examine the spatial distribution of the disease incidence, where there is a trend in longitude or latitude, or rural vs urban, or coastal vs inland, or even specific hot spots of the disease. Visualizing diseases on maps is often the first step in exploratory spatial data analysis and effectively helps in the formulation of hypotheses. Disease maps help to present geographic patterns that may be overlooked in a table, obscuring the geospatially related statistics [1]. By providing a visual representation of cancer outcomes, geographic patterns of disease may be identified and effectively addressed with public health policy and actions. [2] recognizes one of the key challenges with mapping spatial patterns of disease is the design of visualizations. This paper addresses the visualization techniques and their applications to cancer statistics. Highlighting the differences and historic use of these displays.

The paper is organized as follows. The next section describes the choropleth map which is the common approach to disease maps. Section 3 surveys atlases in use today. Section 4 describes alternative displays, including the cartogram which is useful when the map has heterogeneously sized geographic units. The pros and cons of these approaches is discussed in Section 5. Disease

maps are more useful when made interactive, and common options are described in Section 6, along with a discussion of benefits and disadvantages. The last section summarizes the paper.

2 Map displays for disease data

A choropleth map is used to display differences in the geographical distribution of data by spatial unit by shading areas of a map [3]. The geography is faithfully rendered, and the color rendering is designed to reveal spatial patterns among data values. Identifying and explaining spatial structures, patterns, and processes involves considering the individuals in communities and organizing communities into representable units [1]. Early versions of choropleth maps used symbols or patterns instead of color. [4] discuss the use of choropleths to visualize cancer data, and [5] gives an overview of the development of these maps for displaying disease data. In epidemiology, choropleths are often used as a tool to study the spatial distribution of cancer incidence and mortality.

Figure 2 shows various geographic projections of choropleth maps of age-adjusted rate (per 100,000 people) of new cases of lung and bronchus in the USA, averaged over 2012 through 2016. The data was extracted from the official federal statistics by [6] on cancer incidence and deaths, produced by the Centers for Disease Control and Prevention (CDC) and the National Cancer Institute (NCI). A west to east spatial trend of increasing rates, can be seen. There is also a spatial outlier – Utah has a noticeably lower rate than its neighbors. Also Kentucky has a noticeably high rate, and Maine also has a higher rate than its neighbors. There is something of a cluster of higher rates around the tobacco states.

Utilizing the state boundaries can make a map familiar to read [7], and allows viewers to visually infer the spatial relationships in the data. The familiarity of the geography is a worthy consideration when presenting results of spatial analysis, to consumers that extend beyond researchers, to the public, policymakers and the media [4].

While the areas are recognizable shapes, they are often politically driven boundaries with individual areas being of non-uniform size, containing different population densities and subject to change over time. The different population and geographical sizes of administrative areas can attract attention to the shades of the underpopulated but large areas [3], Skowronnek [8] calls this an area-size bias. Choropleths can inhibit visual inference when presenting human-related statistics as the display may draw attention away from the ‘potentially more important results in the more populous communities’ that are geographically smaller [2].

3 Cancer atlases

A cancer atlas is a map, or collection of maps, representing cancer incidence and mortality for a country, or group of countries. Atlases are key to developing hypotheses regarding areas with unusually high rates, geographic correlations, work related exposures, and high risk diets [9]. The data collection of statistics across regions, and the administrative control within regions, lends itself to choropleth visualization. Atlases date to Haviland’s maps in 1875, with UK cancer atlases appearing in 1961 [10], and early work in US cancer atlases appearing in 1971 [11]. The presentation of cancer statistics has increased with greater access to computational power, and the availability of geographic information systems software [2].

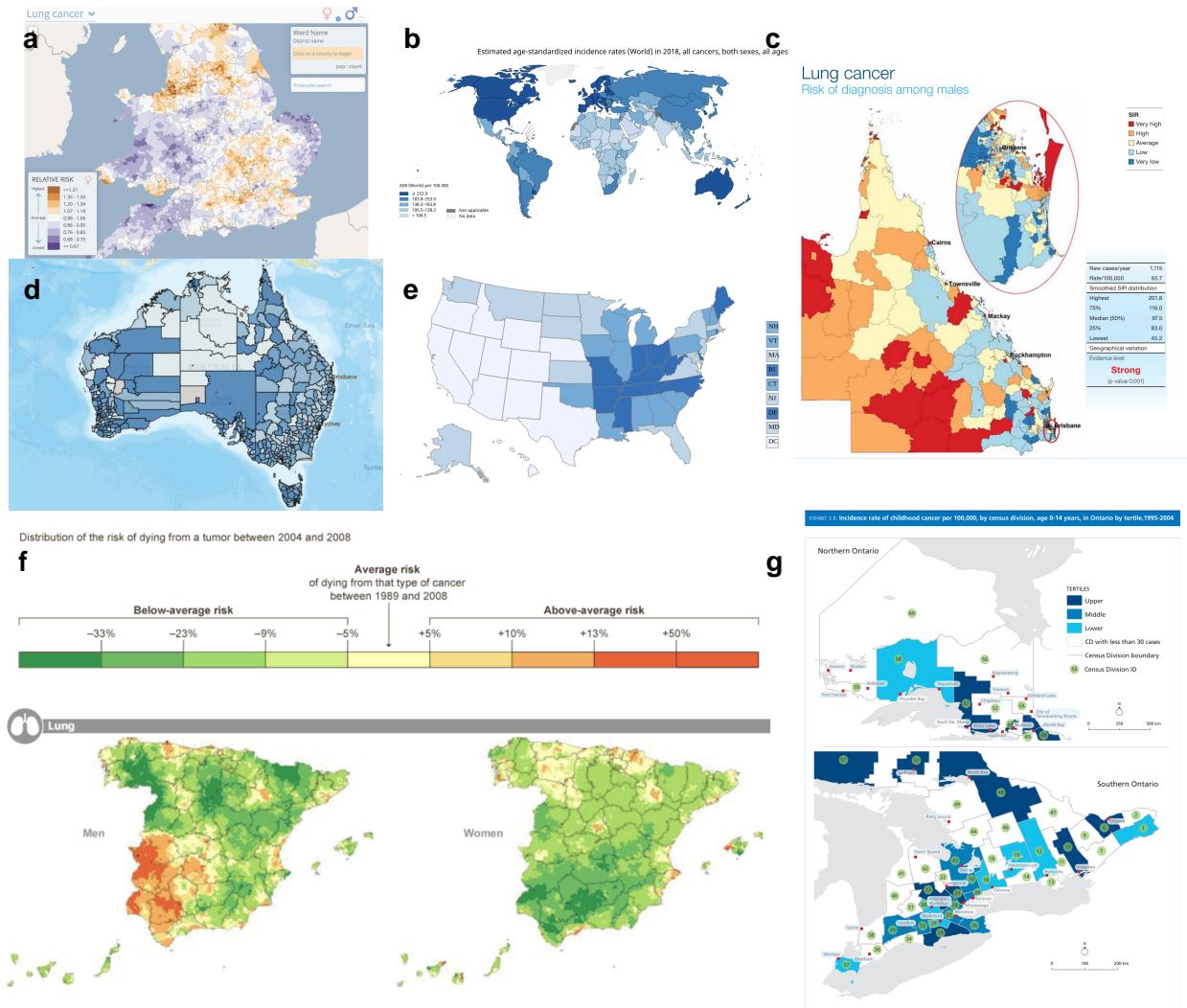


Figure 1: Publicly available choropleth cancer maps published between 2010 and 2015.

Mortality rates are now often presented as relative rates of risk across the population, and age adjusted to correct for the higher prevalence of cancers in older populations. Howe [10] describes Stock's development of the standardized mortality ratios through the 1930s. Table 1 summarizes the measures commonly presented in published cancer atlases.

Table 1: Commonly used measures for reporting cancer information.

Measure	Details
1. Count	Crude cancer counts
2. Rate per 100,000	Cancer incidence per 100,000 population
3. 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
4. Age Adjusted Rate per 100,000	Standardized by age structure or region
5. Age Adjusted Relative Risk	Standardized by age structure in each region i
6. SIR (Standardized Incidence Ratio)	Standardized by incidence in each region i
7. Below or above Expected	Alternative expression of the SIR
8. 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'

Cancer maps are effective tools for communicating incidence, survival and mortality to wide range of audiences, including the general public and others not trained in statistical analyses. These visualizations enable non-expert audiences to interpret the outputs of sophisticated statistical analyses. Cruickshank (1947) as cited by [5], discusses using visuals as a 'formal statistical assessment of the spatial pattern'. Overwhelmingly, the cancer maps utilized to communicate to the public and other non-expert audiences are choropleths.

3.1 Overview of publicly available atlases

Roberts [12] identified 33 publicly available cancer atlases, published between January 2010 and November 2015, from different countries around the globe. All of these use choropleth maps. All but one of these were published by non-commercial organizations, including not-for-profits, government, research organizations, advocacy groups or government funded partnerships. Figure 1 displays a subset of maps from these atlases.

3.1.1 The Environment and Health Atlas of England and Wales

Figure 1a shows the relative risk for women developing lung cancer in England and Wales in 2010 [13]. The cancer data was from Office for National Statistics (ONS) (England) and from the Welsh Cancer Intelligence and Surveillance Unit (WCISU).

3.1.2 Globocan 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide

The map seen in Figure 1b shows age standardized incidence rates (per 100,000) for all invasive cancers for both men and women, aggregated at a national level for 2018 [14]. It is published by the World Health Organization's International Agency for Research on Cancer. Data was sourced from cancer registries of each country.

3.1.3 Atlas of Cancer in Queensland

Figure 1c shows the relative incidence ratio of lung cancer in males in the state of QLD within Australia based on data from 1998 to 2007 [15]. It is published by the Queensland Cancer Council, using data from the *Queensland Cancer Registry*.

3.1.4 Bowel Cancer Australia Atlas

Figure 1d shows the percentage of Australian males between 50 - 54 years of age diagnosed with bowel cancer in 2016 in Australia [16]. It is published by *Bowel Cancer Australia*.

3.1.5 United States Cancer Statistics: An Interactive Cancer Statistics Website

The *United States Cancer Statistics: An Interactive Cancer Statistics Website* [6] can be seen in Figure 1e. This map contains the incidence rate per 100,000, of all cancer types for men and women in the United States in 2016, aggregated at the state level. The map was published by the *Centers for Disease Control and Prevention*, with data from state cancer registries.

3.1.6 Map of Cancer Mortality Rates in Spain

Figure 1f shows side by side maps of relative risk of lung cancer for men vs women for 2004 to 2008 [17].

3.1.7 Atlas of Childhood Cancer in Ontario

Figure 1g displays the incidence rate of childhood cancers per 100,000 (by census division) for children aged 0-14, in Ontario from 1995 to 2004 [18].

3.2 Additional considerations

Maps cannot tell the entire story. Supplementary graphs and plots are often included to add more depth and information. Additional materials such as tables, graphs, and text explanations, support understanding and inference derived from maps, ensuring the message communicated will be consistent across a range of viewers [4]. The many displays of statistical summaries, including dot plots, bar plots, box plots, cumulative distribution plots, scatter plots, normal probability plots, can provide alternative views of the cancer statistics. These can also display supporting statistics

such as error, confidence intervals, distributions, sample or population sizes, standard deviation. When presenting cancer maps, [9] believes the intuition derived from maps must be ‘validated by rigorous statistical analyses’, and the display of the supplementary statistics helps in this regard.

3.2.1 Geographic hierarchies

While atlases are often used to describe differences between areas, statistics may be displayed at different levels of aggregation. Global health statistics can be aggregated to administrative and arbitrarily defined regions, such as those used by the World Health Organization and the United Nations [19]. World atlases can allow for displays of data aggregated into continents, countries, states, provinces and congressional districts [6].

3.2.2 Population distribution

It is extremely likely that each population area will have a different number of people. The distribution of the population residing in all areas may also be communicated in a table or histogram display [20]. Atlases can connect the population to the land available to them by communicating population density. Alternatives to choropleth to address this are described in Section 4 .

3.2.3 Statistical uncertainty

With any number comes some uncertainty about its value. This may occur for a number of reasons: numbers are drawn from samples, errors occur in the disease reporting, values are simulated from a model to maintain privacy. Close to half of the atlases identified (42%, n=14) included some measure of uncertainty. Uncertainty is often recorded as sample or population size, standard deviation, confidence intervals, statistical significance, quantiles, credible intervals. The most common measure used to represent uncertainty were credible or confidence intervals (CIs).

Displaying the uncertainty associated with reported statistics is a vital feature of a cancer map, but it can be difficult to display effectively. The map focuses on displaying the statistic, and lacks additional space to represent the uncertainty. Providing an adjacent map or overlaying maps with symbols [21] are two approaches.

3.2.4 Demographics

Demographics include information regarding the age and sex distribution of the areas presented. Digital atlases allow for users to control the display, to select subsets such as males, females or those aged over 65.

3.2.5 Socio-economic indicators

Socio-economic indicators, such as unemployment rates, poverty rates, remoteness, and education levels, can explain how the experience of cancer prevalence varies for various members of a society. Few atlases provide this level of detail. Data is available from the United Nations Human Development Index and financial indicators are available from the World Bank [19].

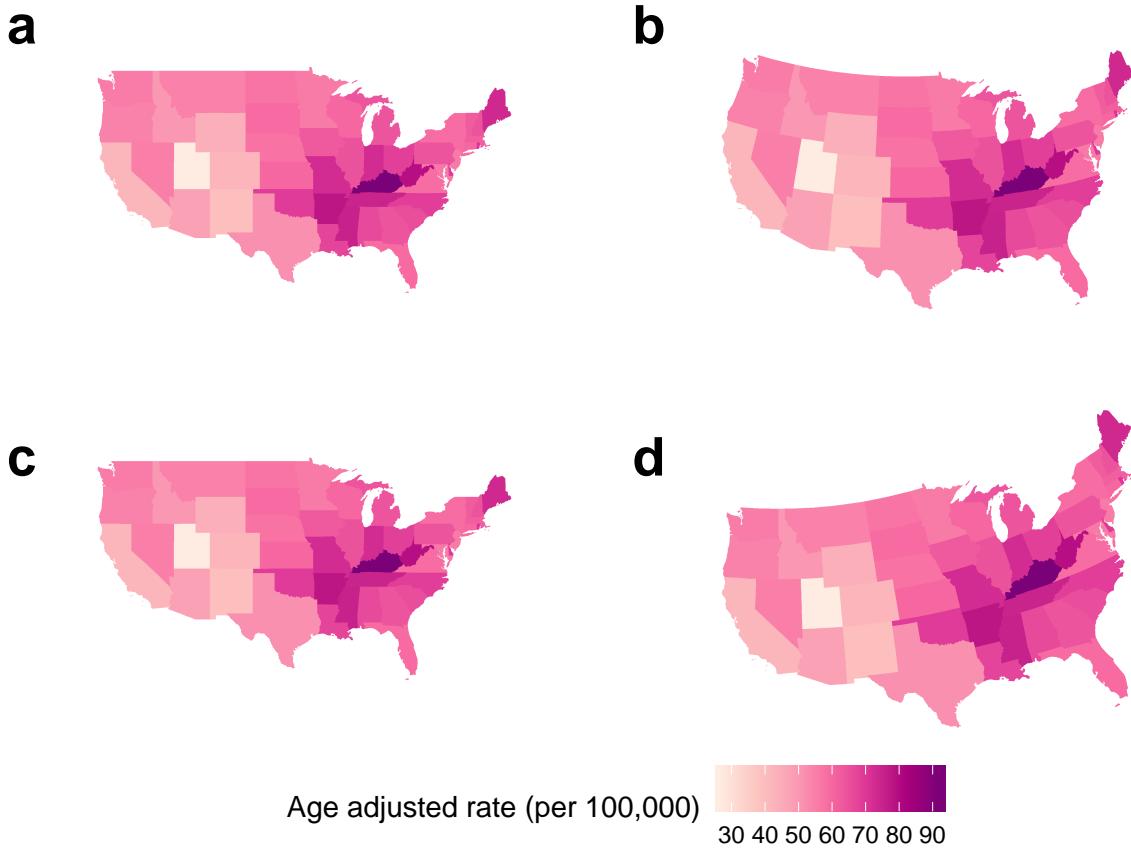


Figure 2: Four choropleth maps average age-adjusted rate of incidence for lung and bronchus in the United States averaged over sex and years 2012-2016, using four different map projections. The map projections alter the shapes and angles of the boundaries of each state.

4 Alternatives to choropleths

4.1 Cartograms

Choropleths imply uniformity of data across the geographic space however population densities are extremely unlikely to be uniform [8]. The cartogram was developed to address this, by transforming the map to better indicate population [22]. The resulting display more accurately communicates the impact of the disease across the population, as recorded by the statistic, at the sacrifice of geographic accuracy.

The purposeful distortion of the map space, transformed according to population density, is beneficial when a uniform density of the map base is desired. Population then becomes a uniformly distributed background for the statistic presented [23]. “population distribution is often extremely uneven in former British colonies” [24], and this makes the distortion necessary [25]. When implementing a distortion of the geographical shape according to population, the resulting display is an area cartogram [26], or population-by-area cartogram [27]. Event cartograms change the area of regions on a map depending on the amount of disease related events, but this does not consider the effects of land area and population [21].

Cartograms provide an alternative visualization method for statistical and geographical information. The key difference between a choropleth and a cartogram is the desirable augmentation of the size, shape or distance of geographical areas [24]. Monmonier [28] suggests that white lies may be employed to create useful displays and map creators have the ability to draw lines that may distort the geometry and suppress features and it is easy for the average person to disregard the impact of transformations used to create cartograms. Cartograms may be seen as an extension of map transformations and projections. The favorable distortion is proportional to a value other than actual earth size area [26]. A disadvantage of the conventional map is that sparsely populated rural areas may be emphasized, whereas the areas representing cities are very small, making interpretation of spatial patterns very difficult. The distortion of a cartogram accounts for the population density, preventing it from obscuring the spatial patterns [27]. The spatial transformation of map regions relative to the data emphasizes the data distribution instead of land size [29]. When visualizing population statistics, Dorling [24] considers this equitable representation design ‘more socially just’, or honest [30], giving due attention to all members of the population and reducing the visual impact of large areas with small populations [5]. Howe [10] suggests that ‘cancer occurs in people, not in geographical areas’ and [25] believe that spatial socio-economic data, like cancer rates, are best presented on a cartogram for urban areas as the population map base avoids allocating ‘undue prominence’ to rural areas.

The creation of cartograms was historically in the hands of professional cartographers [31]. Early approaches including John Hunter and Jonathan Young (1968) and Durham’s wooden tile method, Skoda and Robertson’s (1972) steel ball bearing approach and Tobler’s (1973) computer programs [24]. Geographical information systems allowed map users, and researchers to implement their own cartograms, but these systems are utilized depending on ‘the effectiveness, efficiency, and satisfaction of the map products (Nielsen 1994), [31]. Howe [10] discusses the impact of electronic computer-assisted techniques.

There are many alternatives to consider, the intended audience of the map, and its purpose are key points in cartogram use and creation. ‘There is no “best” cartogram or method of creating cartograms just as there is no “best” map’ (Monmonier and Schnell, 1988) [24]. Nusrat and Kobourov [32] provided a framework to investigate implementations of the many algorithms presented, and the “statistical accuracy, geographical accuracy, and topological accuracy”.

Table 2: Maps used to present statistics for the United States of America. Each state has been colored according to the average age-adjusted rate of incidence for lung and bronchus for females and males in the United States 2012-2016.

Map display	Details
a. Contiguous	Each stated shape has been distorted according to the population of the state in 2015. The state of California has become much larger due to its large population density. This draws attention to the densely populated North East region, and detracts from the less populated Mid West.

Map display	Details
b. Non-contiguous	The geographic shape of the states has been maintained, but the size has altered according to the population of the state in 2015. The state of California has remained closer to its original size than its surrounding states. The North East states have remained closer to their geographical size, in the case of Massachusetts and Connecticut. This draws attention to the densely populated North East region, and the sparse Mid West.
c. Dorling	The states have been represented by a circle, but the size was determined by the population of the state in 2015. The North East states remain closer to their neighbors, and may be displaced from their geographic location. The sparsity of the population in the Mid West is highlighted by the distance between the circles, located at the geographic centroids.
d. Hexagon Tessellation	Each state is represented by a hexagon of equal size. The neighbouring states are easily contrasted, however the north east regions have been displaced from their geographic location. The sparsity of the population in the Mid West is highlighted by the light yellow color, the age adjusted rate in Kentucky is the darkest and its neighbors are similar.

Figure 3 shows four different cartograms for the same data displayed in Figure 2. Table {tab:usa} summarizes information that can be observed in the four types of cartograms.

4.1.1 Contiguous

A contiguous cartogram maintains connectivity of the map regions while areas are altered according to a statistic. This transformation often occurs at the expense of the shape of areas [29]. From a computer graphics perspective, [33] explain the application of ‘map deformation’ to account for the value assigned to each area, they provide three methods for creating value-by-area cartograms. Examples include Tobler’s Pseudo-Cartogram Method, Dorling’s Cellular Automaton Method [24], Radial Expansion Method of Selvin et al., Rubber Sheet Method of Dougenik et al., Gusein-Zade and Tikunov’s Line Integral Method, Constraint-Based Method (Kocmoud and House) [29].

4.1.2 Non-contiguous

Non-contiguous cartograms preserve shape of the area over connectivity. Each area stays in a similar position to their location on a choropleth map. The choropleth map base is often also plotted as a comparison point to highlight the change in area. The addition is the gap between areas, created as each individual area shrinks or grows according to the associated value of the statistic. Olson [26] discusses creation of these maps, the significance of the empty areas left between the geographic boundaries and the new shape, and the ‘degree of difference from the original map that is the real message’ of these displays.

As the trade-off regarding boundaries approaches simplicity, the distortion of region shapes on the contiguous cartogram presents an additional hurdle to visual recognition and this hurdle is not

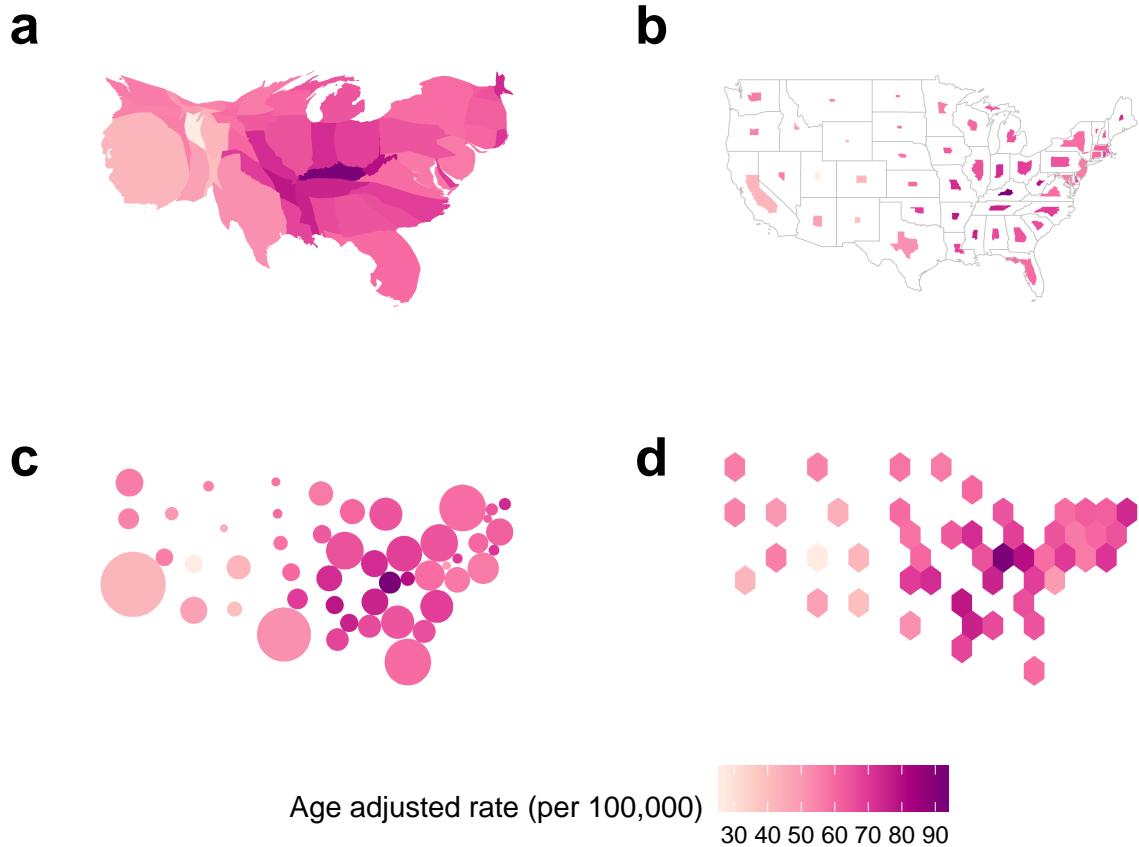


Figure 3: Common alternatives to maps, showing the same information as in Figure efchoroCRS for the United States of America: (a) contiguous cartogram, (b) non-contiguous, shape-preserved cartogram, (c) Dorling cartogram (non-contiguous), (d) hexmap (non-contiguous). In (a) - (c) the state has been resized, and reshaped, to match the 2015 population of the state. This provides a better sense of the extent of disease relative to the population in the country, and can help alleviate losing information about physically small but population dense states. In the hexmap (d) each state is given equal size, and thus equal emphasis.

only eliminated on the non-contiguous cartogram but is replaced by the meaningful empty-space property [34].

4.1.3 Dorling

Daniel Dorling presented an alternative display engineers to highlight the spatial distribution and neighbourhood relationships without complex distortions of borders and boundaries. This approach opposes preserving the intricate shape details and is founded in the simple question put forward by Daniel Dorling [24]:

“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 acknowledged the sophistication of contiguous cartograms but critiqued their ‘very complex shapes,’ he answers this with his implementation of maps created using ‘the simplest of all shapes’. Circular cartograms use the same simple shapes for every region represented, and resizes the shapes according to the statistic represented or the population for a base map. This familiar shape may be more effective for understanding the spatial distribution than contiguous cartograms, as the ‘nonsense’ shapes used have ‘no meaning’ after distortions are applied [30]. 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.

The groundwork for this approach had been laid in the mid 1930’s by Raisz, and rectangular cartograms provide dramatic comparisons and are especially useful for correcting misconceptions communicated by geographic maps. [35] quotes the official definition of Value-Area Cartograms, the simplistic displays which represent each area as a single rectangle, sized according to the value of the statistic. This rectangular display also allows for tiling, where geographic neighbors placed in suitable relative positions also share borders, however contiguity may be sacrificed [36]. Rectangular cartograms allow for bi-variate displays, population can be effectively communicated by the size of each rectangular, and a second variable can be communicated using color [37].

A similar method, where each geographic area is represented using a square, tessellated to create a square grid. This method has been used broadly in the media, e.g. FiveThirtyEight, Bloomberg Business, The Guardian, The Washington Post, The New York Times and NPR. Each area is represented by a square of the same dimensions.

4.2 Tile map

A tile map provides a tessellated display of consistent shapes. Each tile is usually one unit of measurement, this could be geographic regions such as states, or population based where one tile is used for a consistent measure of population.

A simple tile map presents the areas in a tessellated grid display, where geographic neighbors are found next to each other, with some necessary displacement employed for regions with more than four neighbors. These tiles may be labelled or colored to represent a value. Tile maps may be difficult to create, they are best created manually, with additional time and care required as the amount of geographic areas to include increases. [38] define the term ‘mosaic cartograms’ for hexagonal tile displays, where the number of tiles for each area can be used to communicate the

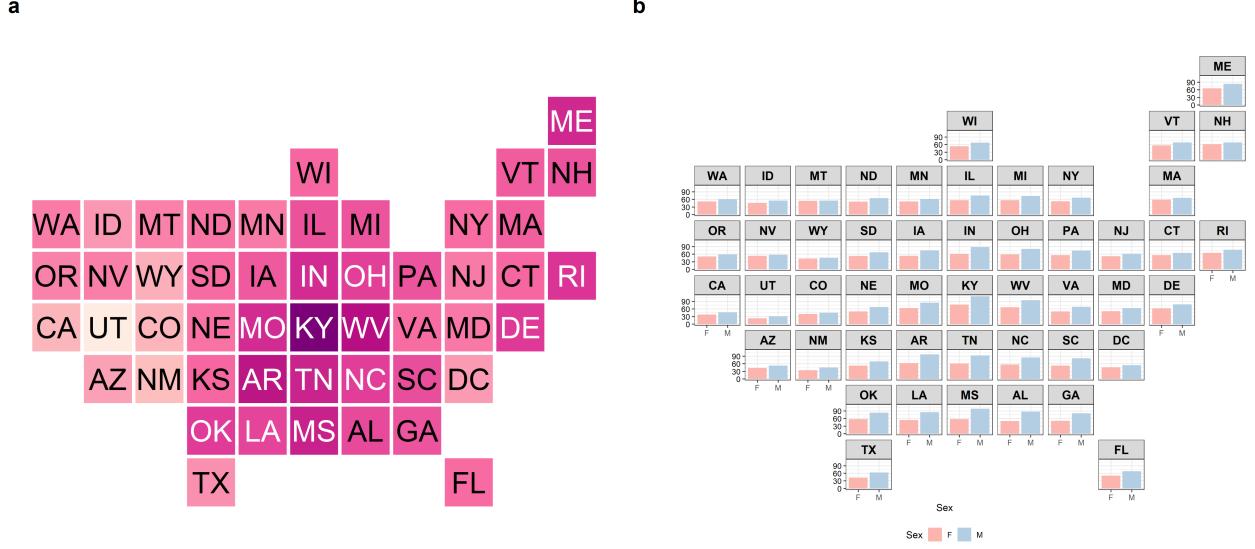


Figure 4: Two more alternative displays, tile map (left) and geofaceted map (right), showing state age-adjusted rate of incidence for lung and bronchus in the USA. In the tile map, the layout approximates spatial location, with each state being an equal box filled with colour representing cancer incidence. The geofaceted map shows bar charts laid out in a grid approximating the spatial location of the state. The age-adjusted rates for males and females is shown. This type of display allows multiple variables to be displayed for each geographic area.

statistic of regions. The complexity of the boundaries can be adjusted in the resulting display, as the size of the tiles used allows a trade-off be made between boundary complexity and simplicity.

4.3 Geofaceting

Tile maps can be extended to allow more than just simple coloring of tiles to present data. [39] formalizes the term geofaceting to describe a grid display, the arrangement of tiles to create a grid that mimics the geographic topology of the set of areas. Like tile maps, the arrangement of the areas can be reused for other spatial distribution visualizations. Geofaceting have the functionality of facets, often used to replicate visualizations for each subset of the data, in this case the data subsets are geographic. The amount of information able to be communicated has increased from one value per region in a tile map to one visualization, this is a more flexible display. Virtually any plot display can be fit into the tile representing the areas, allowing displays of multiple variables or values per geographic entity. As the amount of areas increase care must be taken to ensure legibility of the displays. [40] acknowledges that this display suffers when the areas have very irregular shapes and large size disparities, as manual creation of the grid is still required.

4.4 Multivariate displays

[41] present linked micromap plots to visually link geographic and statistical data, this serves as a solution to multi-dimensionality issues. This design is widely used in online atlas displays. Areas

are grouped based on their value for one variable, and additional columns provide displays that contrast the areas in each group by other variables. The display juxtaposes of choropleth maps and statistical plots; it shows one map per group of the key separating variable, in a row with each additional statistical plot. Linked micromaps predominantly utilize the choropleth map for displays of spatial relationships, that are seen when spatial neighbours are allotted to the same group. It is one of several alternative displays that allow maps to become bivariate displays, commonly used to present both an estimate and the associated uncertainty.

Bivariate choropleth maps [42] blend colour schemes to convey the intersection of categorised levels of an estimate and the associated uncertainty for each spatial area. [42] also suggests map pixelation, each region is broken into small pixels, the individual pixels are allocated values that reflect the uncertainty around the area's estimate by creating texture. This display can also be animated, with each frame produced by resampling the pixels. Areas with uncertain values will flicker more dramatically than areas with more certain values.

Solutions have been provided for bi-variate displays for movement data, containing Origin - Destination matrices and flow map representations [43]. This method makes the maps and matrix connections clear, and easy to link geographic areas with link lines not crossing. Bar charts can also be added to these to show group totals.

5 Comparison of maps and cartograms

To choose an appropriate map display, the map creator must consider the intended user, and message the map will communicate. Choropleth displays utilize more traditional cartographic methods, they are usually true to the topography, displaying familiar boundaries of countries, states or administrative areas. Highlighting the geographic distribution, and differences between equal spatial units. When there is a relationship between the variable being mapped and geographical features, a choropleth map would pair this information visually. Choropleth methods require a decision to be made about the projection of the display, in this case aim to find a map projection that gives the least distorted representation of the geography [44]. However a choropleth is limited to a uni-variate display, presenting a spatial distribution by filling each area with a color. These displays favor geographically large regions over small regions, though inset displays are often used to zoom in on a geographic area containing many small regions, this is seen in Figure 1 c of the state of Queensland, where the capital city Brisbane is highlighted by the inset map. Users can identify the areas relevant to them by the familiar boundaries, and can easily contrast the experience of geographic neighbors. The map is reusable and fit for many purposes.

Cartographers were historically the creators of geographic displays, and epidemiologists utilized the maps that had been prepared in advance. Cartograms with the population variable used to distortion the geographic map regions and boundaries are density equalising maps, that reduce the visual prominence of large, but low-population areas [21]. Cartogram displays incorporate the statistics and population into the design of the display, shifting and sometimes sacrificing familiar boundaries to draw attention to large outliers in the data space or population density. The unusual shapes are not longer familiar and may be difficult to compare. This can prevent effective comparison of the statistic. Cartograms also favor extreme statistics in the positive direction. Small extreme values are often lost due to the resulting small sized areas the do not draw user's attention. The change in the map based on one variable provides the opportunity for bi-variate displays, using the

change in area for one variable, and color for another. The difference between the familiar map display helps highlight the impact of the disease on communities.

Alternative maps may replace the unexpected boundaries and shapes with familiar, simple shapes. This makes the spatial distribution the primary concept communicated by the display. Areas that are geographically close will maintain connectedness in some way, but the population and the statistic will dominate. Tile map displays may only allow one variable to be visualized if each region is given consistent map space, this occurs in geofaceting. Dorlings provide bi-variate displays as they use the size of the shape for one variable, and color for another. These alternatives are especially helpful for data aggregations where administrative boundaries break populations into groups. The experience of each community may be worth considering, as the experiences shared by the population within them may be similar due to the services and facilities they share. This display allows a more equitable view of each community, and does not minimise those that operate on a smaller geographic scale.

Faceting provides an opportunity to contrast the overall distribution of variables across a geographic space or collection of geographic areas. This display makes it difficult to compare specific regions.

Table 3: Summary of features and constraints on common mapping methods used to display cancer statistics (Y=Yes, N=No, S=Sometimes).

	Choropleth	Contiguous	Non-contig	Dorling	Tile maps	Geofacets
Spatial distortion	N	Y	Y	Y	Y	Y
Preserves neighbors	Y	Y	Y	S	S	S
Conceals small areas	Y	S	N	N	N	N
Uniform shape	N	N	N	Y	Y	Y
Univariate only	Y	Y	Y	S	S	N
Manual construction	N	N	N	N	Y	Y

Creating maps of diseases now involves more decisions to be made by map makers, rather than cartographers. Technology has played an enormous role in increasing the opportunities for map makers. The computation and graphics power have made creation, alteration and interactivity possible. as these options have expanded and it is the objectives of the investigator that will drive the choices. [4] and [1] have provided suggestions and comments to help map creators best communicate their health data and spatial analyses.

5.1 Neither choropleths or cartograms perform well for Australia

Australia presents an extreme case of an urban rural divide. The land mass occupied by urban electoral districts is only 10% of Australia, yet 90% of the population live in these urban areas. To present election results on a choropleth map should be ‘unthinkable’, as it means diminishing the visual impact of majority of the electorates. A 1966 cartogram presented an alternative where boundary lines were largely straight line, and the result looked very little like the geographical shape of Australia. This issue is felt in any nation which experiences a spatially heterogeneous population distribution. As this feature of population distributions continues to intensify, the need for cartograms as an alternative to a choropleth map should only increase.

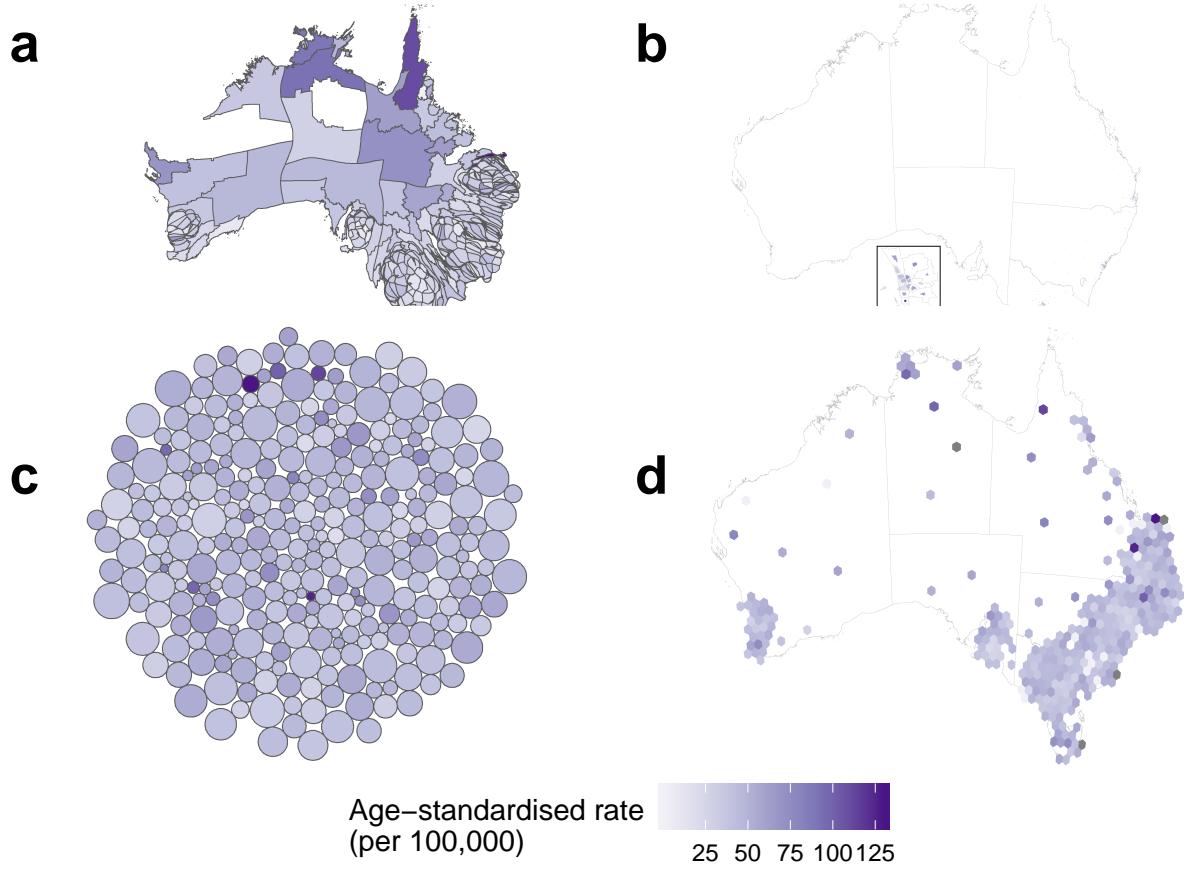


Figure 5: Cartograms showing melanoma incidence in Australia: (a) contiguous, partially population transformed, (b) non-contiguous shape preserved, (c) Dorling, (d) hexmap. The contiguous cartogram has expanded the highly populated areas, while preserving the full shapes of rural areas. If it accurately sized areas by population the country would be unrecognisable. The shape-preserved is unreadable. Dorling has lost all geographic context. The hexmap provides a reasonable spatial distribution despite having isolated hexagons in the outback areas.

6 User interaction

One of the concerns of adding too much information to a map is the fear of cognitive overload [45] in which the user reaches an information threshold, beyond which will not be able to make sense of the information. It can be a juggling act for a diverse audience, with experts probably preferring more detail [46]. Interactivity is a design feature within modern mapping methods that can be used to incorporate additional information and complexity without overloading the user. The effect of user-centered interactive actions needs to produce rapid, incremental, and reversible changes to the display [47].

[28] says that interactivity can be used to allow users to explore the map for more information and provides flexibility for the display. The user can toggle between different variables, map views or even multiple realizations of future scenarios [48]. This provides additional mechanisms for the users to digest the uncertainty of the available information [49]. Where the needs of the audience is changeable and is the priority, the map creator can allow interactivity for map users to explore a data set through dynamic interactions that allow inspection of the data from many views [51]. [52] says user interaction with maps helps them to understand and interpret the spatial distribution of disease, to validate, explain or explore the presented statistics and their relationships to each other.

Interactivity enables supplementary information to be incorporated into online atlases without cluttering the display. Interactive design features, found in online cancer maps, include tool tips, drop down menus, data selection, zooming and panning allow users to explore the map as they want more information and allow flexibility in the display [28]. The use of these supports were found in various online cancer maps identified by [12].

Animation, in contrast to interactivity, usually involves pre-computing views, and showing these in a sequence. [53] provided an overview of animation in association with the R package `ggridge` [54], with some focus on maps, explaining transitions between frames in a linear, to provide a fixed narrative to passive users. Animations are used to communicate a message by capturing and directing users attention. It is most often employed to show changes over time. The controls for basic animation are usually placed outside of the plot space [53], the map image is updated/replaced as the animation progresses.

[4] provide weather maps as a thoroughly developed example of animation of spatial displays to communicate information to the general public. The movement of a weather system will follow a forecasted path, all map users can follow the animated path of the weather system across the geography over a specified period of time.

The Australian Cancer Atlas [55] provides tours which change the display to draw user's attention to areas on the map that are relevant to the story. This implementation of animation gives users tools to plan their own exploration.

6.1 Interaction and animation in publicly available atlases

Many publicly available cancer atlases employ interactivity and/or animation. Figure ?? shows the interactive GUI tools provided with the public atlases discussed in Section 3.1. Mostly these are menus, and toggles, for choosing indicators, and subsets of the population.

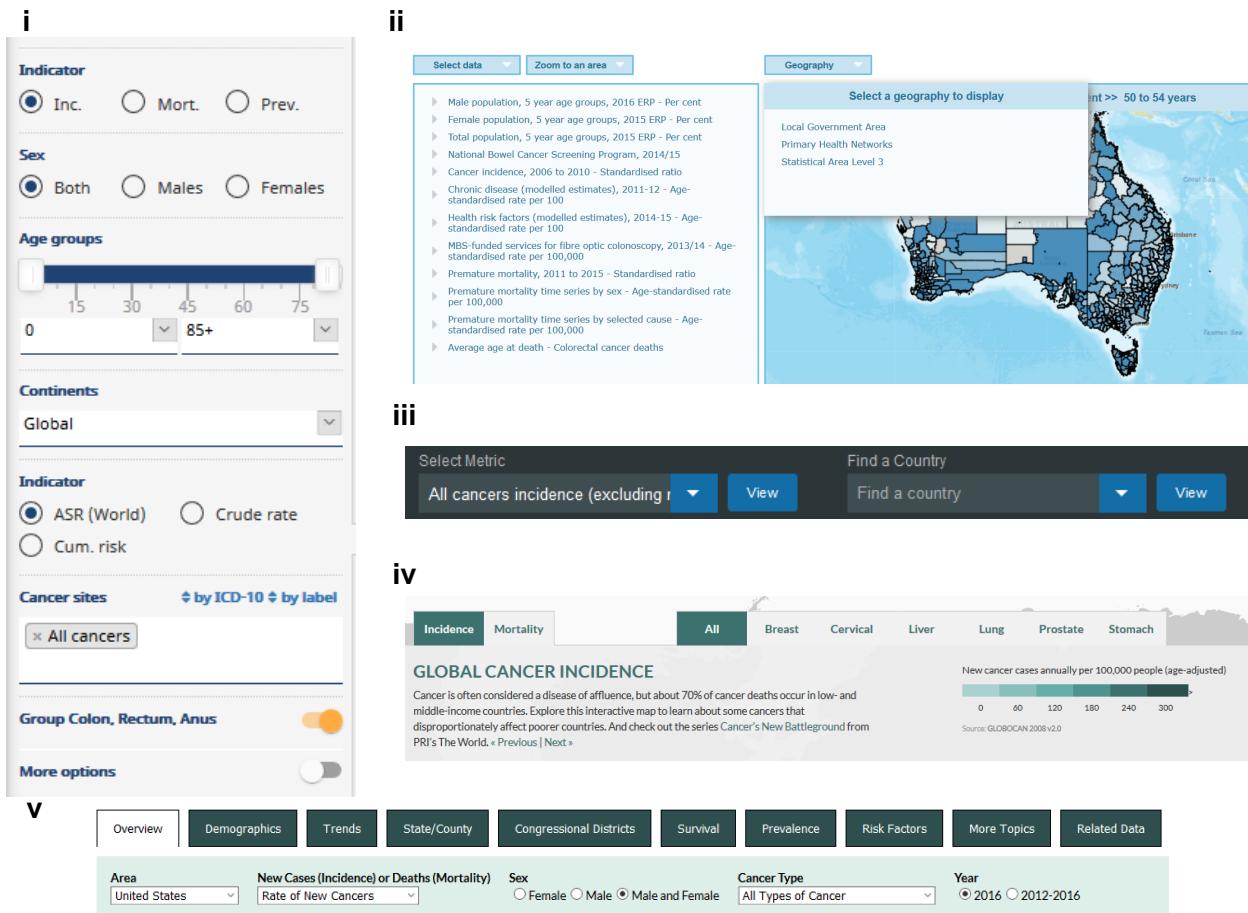


Figure 6: Interactive controls of displays in publicly available choropleth cancer maps: (i) GUI controls for statistic, sex, age groups, continents, and cancer types for Globocan 2018: Cancer Today, (ii) Menus for variable selection and zooming on Bowel Cancer Australia Atlas, (iii) Menus for choosing variables and countries in The Cancer Atlas, (iv) Tabs for different indicators and cancer types in Global Cancer Map, (v) Menus and toggles for variable and subset selection in United States Cancer Statistics: Data Visualizations.

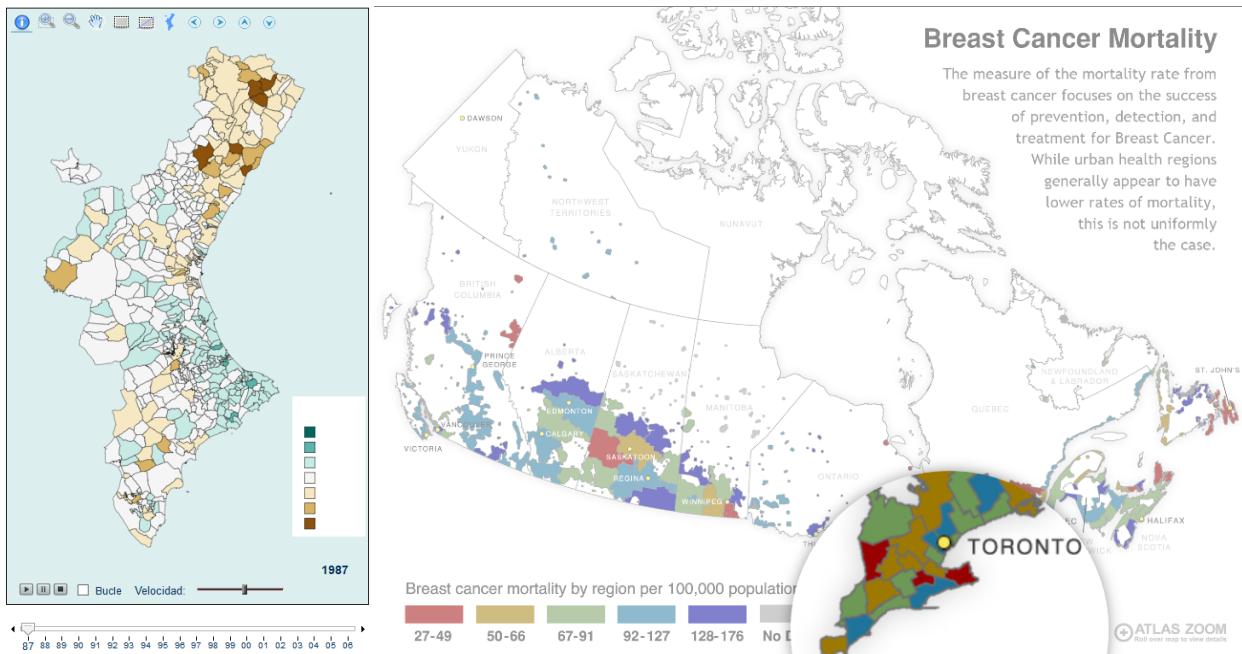


Figure 7: Two examples of advanced interactivity (and animation) in publicly available choropleth cancer maps: a. Linked maps and time series line plots, with temporal animation in Map of Cancer Mortality Rates in Spain, b. A highly responsive magnifying glass on a map of Breast Cancer Mortality in Canada.

6.2 Under-utilised techniques

Figure ?? shows two examples of more sophisticated interactive maps. The Spanish Cancer map (left) contains a linked display between a choropleth map, and time series plots of cancer change. In linked plots, changing values in one display will trigger changes of corresponding elements in another display. Here, the temporal change in the choropleth map can be played out as an animation. Mousing over the time series plots will highlight the line for a particular region. The Canadian Breast Cancer Mortality map (right) has a magnifying glass which allows the user to zoom into small areas. It is easy to control and shows precise details in the small areas.

7 Conclusions

This paper provides an overview of mapping practices as used for cancer atlases, and new approaches that could be adopted. The conventional approach is the choropleth map, and it is widely used. When there are small areas, as occurs in Australia where the population is concentrated on the coast, the information about cancer can be lost, and alternatives are needed. Making an inset can clarify congested regions but this breaks the viewers attention, because they need to shift focus from the map to the inset, and if there are many congested areas, many insets would be needed. The map alternatives, like cartograms and their variations, can be useful to allow the spatial distribution of cancer data to be digested.

There are many different statistics that are commonly used for display. The most basic is incidence rate, which is easy to understand. It is common to see relative rates in many maps which measure how far a region is above or below the average. The purpose of using a relative rate is, perhaps the desire to pinpoint the areas that need attention because they have higher than expected rates. However, we lose the incidence rate information and thus interpretability. A region might be much higher than average, but it may not be close to a health concern, because all regions have low incidence.

Interaction with maps is an important component to public atlases. A purpose is to provide access to more information than is possible to display in a single map, without overwhelming the viewer. Too many choices though can similarly overwhelm a viewer, and thus decisions do need to be made about content to provide, or hide, for accurate and comprehensive communication of information. Similarly, providing ways for users to interact with the display encourages engagement, and creative, efficient, elegant, interactive tools elicits curiosity about the data.

8 Acknowledgements

The authors would like thank Dr Earl Duncan, Professor Kerrie Mengersen, Dr Susanna Cramb and Dr Peter Baade for conversations on the content of this article.

The following R [56] packages were used to produce this paper: tidyverse [57], RColorBrewer [58], ggthemes [59], png [60], cowplot [61], sf [62], spData [63], cartogram [64], sugarbag [65], knitr [66], rmarkdown [67].

Files to reproduce the paper, and code to reproduce the plots, are available at <https://github.com/srkobakian/review>.

9 References

- [1] Moore DA, Carpenter TE. Spatial Analytical Methods and Geographic Information Systems: Use in Health Research and Epidemiology. *Epidemiologic Reviews* 1999;21:143–61. doi:10.1093/oxfordjournals.epirev.a017993.
- [2] Exeter DJ. Spatial Epidemiology. International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology 2016:1–4.
- [3] Tufte ER. Envisioning Information. Graphics Press; 1990.
- [4] Bell BS, Hoskins RE, Pickle LW, Wartenberg D. Current Practices in Spatial Analysis of Cancer Data: Mapping Health Statistics to Inform Policymakers and the Public. *International Journal of Health Geographics* 2006;5:49. doi:10.1186/1476-072X-5-49.
- [5] Walter SD. Disease Mapping: A Historical Perspective. Oxford University Press; 2001. doi:<https://dx.doi.org/10.1093/acprof:oso/9780198515326.003.0012>.
- [6] U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute - Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool (data 1999-2016) 2019. <http://www.cdc.gov/cancer/dataviz> (accessed September 26, 2019).
- [7] Brewster MB, Subramanian SV. Cartographic Insights into the Burden of Mortality in the United Kingdom: A Review of “The Grim Reaper’s Road Map”. *International Journal of Epidemiology* 2010;39:1120–2. doi:10.1093/ije/dyp395.
- [8] Skowronnek A. Beyond Choropleth Maps – A Review of Techniques to Visualize Quantitative Areal Geodata. Infovis Reading Group WS 2015/16 2016. https://alsino.io/static/papers/BeyondChoropleths_AlsinoSkowronnek.pdf.
- [9] d’Onofrio A, Mazzetta C, Robertson C, Smans M, Boyle P, Boniol M. Maps and Atlases of Cancer Mortality: A Review of a Useful Tool to Trigger New Questions. *Ecancermedicalscience* 2016;10:670–0. doi:10.3332/ecancer.2016.670.
- [10] Howe G. Historical Evolution of Disease Mapping in General and Specifically of Cancer Mapping. In: Cancer mapping, Springer; 1989, pp. 1–21.
- [11] Burbank F. Patterns in Cancer Mortality in the United States 1950-67. NCI, Washington DC: National Cancer Institute Monograph Vol. 33; 1971.
- [12] Roberts J. Communication of Statistical Uncertainty to Non-expert Audiences. Master’s thesis. Queensland University of Technology, 2019. doi:10.5204/thesis.eprints.130786.
- [13] Emperial College London - Small Area Health Statistics Unit. The environmental and health atlas of england and wales: National male lung cancer rate 2010. <http://www.envhealthatlas.co.uk/eha/Breast/> (accessed September 26, 2019).
- [14] World Health Organization’s International Agency for Research on Cancer. Globocan 2012: Estimated cancer incidence, mortality and prevalence 2017. <http://globocan.iarc.fr/Pages/Map.aspx> (accessed September 26, 2019).
- [15] Queensland Cancer Registry. The Atlas of Cancer in Queensland (1998 - 2007) 2011. <https://>

cancerqld.org.au/research/queensland-cancer-statistics/queensland-cancer-atlas/ (accessed September 26, 2019).

[16] Bowel Cancer Australia. Bowel Cancer Australia Atlas 2016. <http://www.bowelcanceratlas.org/> (accessed September 26, 2019).

[17] El Pais. Map of Cancer Mortality Rates in Spain 2014. http://elpais.com/elpais/2014/10/06/media/1412612722_141933.html (accessed September 26, 2019).

[18] Pediatric Oncology Group of Ontario. Incidence Rate of Childhood Cancers, Atlas of Childhood Cancer in Ontario (1985-2004) 2015. https://www.pogo.ca/wp-content/uploads/2015/02/POGO_CC-Atlas-3-Incidence_Feb-2015.pdf (accessed September 26, 2019).

[19] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. Global Cancer Observatory: Cancer Today 2018. <https://gco.iarc.fr/today>.

[20] Northern Ireland Cancer Registry. All-Ireland Cancer Atlas (1995-2007) 2011. <http://www.ncri.ie/publications/cancer-atlases>.

[21] Kronenfeld BJ, Wong DWS. Visualizing Statistical Significance of Disease Clusters Using Cartograms. International Journal of Health Geographics 2017;16:19. doi:10.1186/s12942-017-0093-9.

[22] Dougenik JA, Chrisman NR, Niemeyer DR. An Algorithm to Construct Continuous Area Cartograms. The Professional Geographer 1985;37:75–81. doi:10.1111/j.0033-0124.1985.00075.x.

[23] Berry BJL, Morrill RL, Tobler WR. Geographic Ordering of Information: New Opportunities. The Professional Geographer 1964;16:39–44. doi:10.1111/j.0033-0124.1964.039_q.x.

[24] Dorling D. Area Cartograms: Their Use and Creation. In: Concepts and techniques in modern geography (catmog), vol. 59, 2011, pp. 252–60. doi:10.1002/9780470979587.ch33.

[25] Griffin TL. Cartographic Transformation of the Thematic Map Base. Cartography 1980;11:163–74. doi:10.1080/00690805.1980.10438102.

[26] Olson JM. Noncontiguous Area Cartograms. The Professional Geographer 1976;28:371–80. doi:10.1111/j.0033-0124.1976.00371.x.

[27] Levison ME, Haddon Jr W. The Area Adjusted Map. An Epidemiologic Device. Public Health Reports 1965;80:55–9.

[28] Monmonier M. How to Lie with Maps (Third Edition). University of Chicago Press; 2018. doi:10.1191/0309132505ph540pr.

[29] Kocmoud C, House D. A Constraint-based Approach to Constructing Continuous Cartograms. In: Proc. Symp. Spatial data handling, 1998, pp. 236–46.

[30] Dent BD. A Note on the Importance of Shape in Cartogram Communication. Journal of Geography 1972;71:393–401. doi:10.1080/00221347208981697.

[31] Kraak MJ. Cartographic Design. In: The International Encyclopedia of Geography: People, the Earth, Environment, and Technology, United States: Wiley; 2017, pp. 1–16.

[32] Nusrat S, Kobourov SG. The State of the Art in Cartograms. Computer Graphics Forum 2016;35:619–42. doi:10.1111/cgf.12932.

- [33] Min Ouyang, Revesz P. Algorithms for Cartogram Animation. In:. Proceedings 2000 International Database Engineering and Applications Symposium (Cat. No.PR00789), 2000, pp. 231–5. doi:10.1109/IDEAS.2000.880581.
- [34] Keim DA, North SC, Panse C, Schneidewind J. Efficient Cartogram Generation: A Comparison. In:. IEEE Symposium on Information Visualization, 2002. INFOVIS 2002, vol. 2002, IEEE; 2002, pp. 33–6.
- [35] Tobler W. Thirty Five Years of Computer Cartograms. *Annals of the Association of American Geographers* 2004;94:58–73. doi:10.1111/j.1467-8306.2004.09401004.x.
- [36] Monmonier M. Cartography: Distortions, World-views and Creative Solutions. *Progress in Human Geography* 2005;29:217–24. doi:10.1191/0309132505ph540pr.
- [37] Kreveld M van, Speckmann B. On rectangular cartograms. *Computational Geometry* 2007;37:175–87. doi:10.1016/j.comgeo.2006.06.002.
- [38] Cano RG, Buchin K, Castermans T, Pieterse A, Sonke W, Speckmann B. Mosaic Drawings and Cartograms. In:. *Computer graphics forum*, vol. 34, Wiley Online Library; 2015, pp. 361–70.
- [39] Hafen R. Introducing geofacet. Ryanhafen 2018.
- [40] Xie Y, Hofmann H, Cheng X. Reactive Programming for Interactive Graphics. *Statistical Science* 2014;29:201–13. doi:10.1214/14-STS4.
- [41] W. PL, Carr DB, Pearson JB. micromapST: Exploring and Communicating Geospatial Patterns in US State Data. *Journal of Statistical Software* 2015;63:1–25. doi:10.18637/jss.v063.i03.
- [42] Lucchesi LR, K. WC. Visualizing Uncertainty in Areal Data with Bivariate Choropleth Maps, Map Pixelation and Glyph Rotation. *Stat* 2017. doi:10.1002/sta4.150.
- [43] Yang Y, Dwyer T, Goodwin S, Marriott K. Many-to-Many Geographically-Embedded Flow Visualisation: An Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 2017;23:411–20. doi:10.1109/TVCG.2016.2598885.
- [44] Tobler W. Unusual Map Projections. In:. *Cartographic Perspectives*, Association of American Geographers; 1999.
- [45] McGranaghan M. A Cartographic View of Spatial Data Quality. *Cartographica: The International Journal for Geographic Information and Geovisualization* 1993;30:8–19. doi:10.3138/310V-0067-7570-6566.
- [46] Cliburn DC, Feddema JJ, Miller JR, Slocum TA. Design and Evaluation of a Decision Support System in a Water Balance Application. *Computers & Graphics* 2002;26:931–49. doi:10.1016/S0097-8493(02)00181-4.
- [47] Perin C. Direct Manipulation for Information Visualization. Theses. Université Paris Sud - Paris XI, 2014.
- [48] Goodchild M, Buttenfield B, Wood J. On Introduction to Visualizing Data Validity. *Visualization in Geographical Information Systems* 1994:141–9.
- [49] MacEachren AM. Visualizing Uncertain Information. *Cartographic Perspectives* 1992:10–9. doi:<https://doi.org/10.14714/CP13.1000>.

- [50] Van der Wel FJ, Hootsmans RM, Ormeling F. Visualization of Data Quality. In: Modern cartography series, vol. 2, Elsevier; 1994, pp. 313–31. doi:1473871616629516.
- [51] Dang G, North C, Shneiderman B. Dynamic Queries and Brushing on Choropleth Maps. In: Proceedings Fifth International Conference on Information Visualisation, 2001, pp. 757–64. doi:10.1109/IV.2001.942141.
- [52] Carr DB, Wallin JF, Carr DA. Two New Templates for Epidemiology Applications: Linked Micromap Plots and Conditioned Choropleth Maps. Statistics in Medicine 2000;19:2521–38.
- [53] Pedersen TL. The Grammar of Animation 2018. <https://youtu.be/21ZWDrTukEs> (accessed November 16, 2018).
- [54] Pedersen TL, Robinson D. ganimate: A Grammar of Animated Graphics. 2019.
- [55] Cancer Council Queensland, Queensland University of Technology, and Cooperative Research Centre for Spatial Information. Australian Cancer Atlas 2018. <https://atlas.cancer.org.au>.
- [56] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
- [57] Wickham H. tidyverse: R packages for data science 2017. <https://CRAN.R-project.org/package=tidyverse>.
- [58] Neuwirth E. RColorBrewer: ColorBrewer palettes 2014. <https://CRAN.R-project.org/package=RColorBrewer>.
- [59] Arnold JB. ggthemes: Extra Themes, Scales and Geoms for 'ggplot2' 2019. <https://CRAN.R-project.org/package=ggthemes>.
- [60] Urbanek S. png: Read and write PNG images 2013. <https://CRAN.R-project.org/package=png>.
- [61] Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2' 2019. <https://CRAN.R-project.org/package=cowplot>.
- [62] Pebesma E. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 2018;10:439–46. doi:10.32614/RJ-2018-009.
- [63] Bivand R, Nowosad J, Lovelace R. spData: Datasets for Spatial Analysis 2019. <https://CRAN.R-project.org/package=spData>.
- [64] Jeworutzki S. cartogram: Create Cartograms with R 2018. <https://CRAN.R-project.org/package=cartogram>.
- [65] Kobakian S, Cook D. sugarbag: Create Tessellated Hexagon Maps 2019. <https://CRAN.R-project.org/package=sugarbag>.
- [66] Xie Y. knitr: A General-Purpose Package for Dynamic Report Generation in R. 2019.
- [67] Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, et al. rmarkdown: Dynamic Documents for R. 2019.