Cartogram Mapping and its Application to Cancer Data Visualization

Stephanie Kobakian, Jessie Roberts and Dianne Cook

Abstract

Cancer atlases communicate cancer statistics over geographic domains. These domains are subdivided into administrative regions such as countries, states or suburbs. When communicating human-related statistics in Australia, the geographic map base draws attention to the sparsely populated rural areas. The smaller geographic areas may not draw the attention of readers but they are important to consider if they are more densely populated. Alternative map displays can reveal patterns that are not obvious when the same data is shown on a choropleth map. Alternative displays have proven to be successful on online news sites, especially for the communication of election data, where votes are aggregated at state level.

1 Introduction

Cancer statistics are often delivered as aggregated values for geopolitical areas. Presenting these statistics requires aggregating individual observations into statistics for the geographical units, especially for privacy protection but also for political and policy purposes. Counts per area require minimal transformation (e.g. state, province, local government area, post/zip code). However, counts alone are not sufficient for the comparison areas, because the populations of areas are all different. The count data needs to be merged with population data to present incidence rates, for example, the 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.

Map displays are commonly required for the communication of cancer statistics over geographic domains. When using a choropleth map: the statistic is mapped to color and the geographic region is filled with this color. Disease maps help to present geographic patterns that may be overlooked in a table, obscuring the geospatially related statistics [1]. Providing a visual representation of cancer outcomes allows geographic patterns of the disease to be identified and effectively addressed with public health policy and actions. The spatial distribution of the disease incidence can be examined using a choropleth and may reveal 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. Daniel Exeter [2] recognizes one of the key challenges with mapping spatial patterns of disease is the design of visualizations. This paper considers the current visualization techniques to communicate statistics to the public, and their applications to cancer statistics. Alternative approaches are also considered, and the limitations of the visualization methods, 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 and presents examples of atlases in use today. The limitations of the choropleth map are discussed in Section 3. Section 4 describes alternative displays, including the cartogram which is useful when the map has heterogeneously sized geographic units. Section 5 presents the limitaions of producing and using alternative displays. Disease maps are more useful when made interactive, and common options are described in Section 6, along with a discussion of benefits and disadvantages.

2 Traditional approaches for cancer map displays

A choropleth map displays the geographical distribution of data over a set of spatial units by shading areas of a map [3]. The geography of a choropleth is faithfully rendered, and the color rendering is designed to reveal spatial patterns among data values. Identifying and explaining spatial structures, patterns, and processes involve considering the individuals and organizing them into representable units of communities [1]. Early versions of choropleth maps used symbols or patterns instead of color. Choropleths can be used for displaying disease data [4], including cancer data [5]. 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 from 2012 to 2016. The data was extracted from the official federal statistics on cancer incidence and deaths [6], produced by the Centers for Disease Control and Prevention (CDC) and the National Cancer Institute (NCI). The rates exhibit a spatial trend, increasing from west to east. There is also a spatial outlier – Utah has a noticeably lower rate than its neighbors. Also, Kentucky and Maine have higher rates than their neighbors. There is possibly a cluster of higher rates around the tobacco states.

Displaying familiar state boundaries can make a map simpler to read [7], and allow viewers to visually infer the spatial relationships in the data using their mental model of the geography. Consumers of disease displays may include researchers, the general public, policymakers and the media [5], the familiarity of the geography is a worthy consideration when presenting results of spatial analysis.

The familiar shapes 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].

2.1 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 methods across regions and the administrative control within regions lends itself to choropleth visualization. Cancer maps and atlases date back to Haviland's maps in 1875, and early work in US cancer atlases appearing in 1971 [10]. The presentation of cancer statistics has increased with greater access to computational power, and the availability of geographic information systems software [2].

Cancer maps are effective tools for communicating incidence, survival, and mortality to a wide range of audiences, including the general public and others not trained in statistical analyses. These

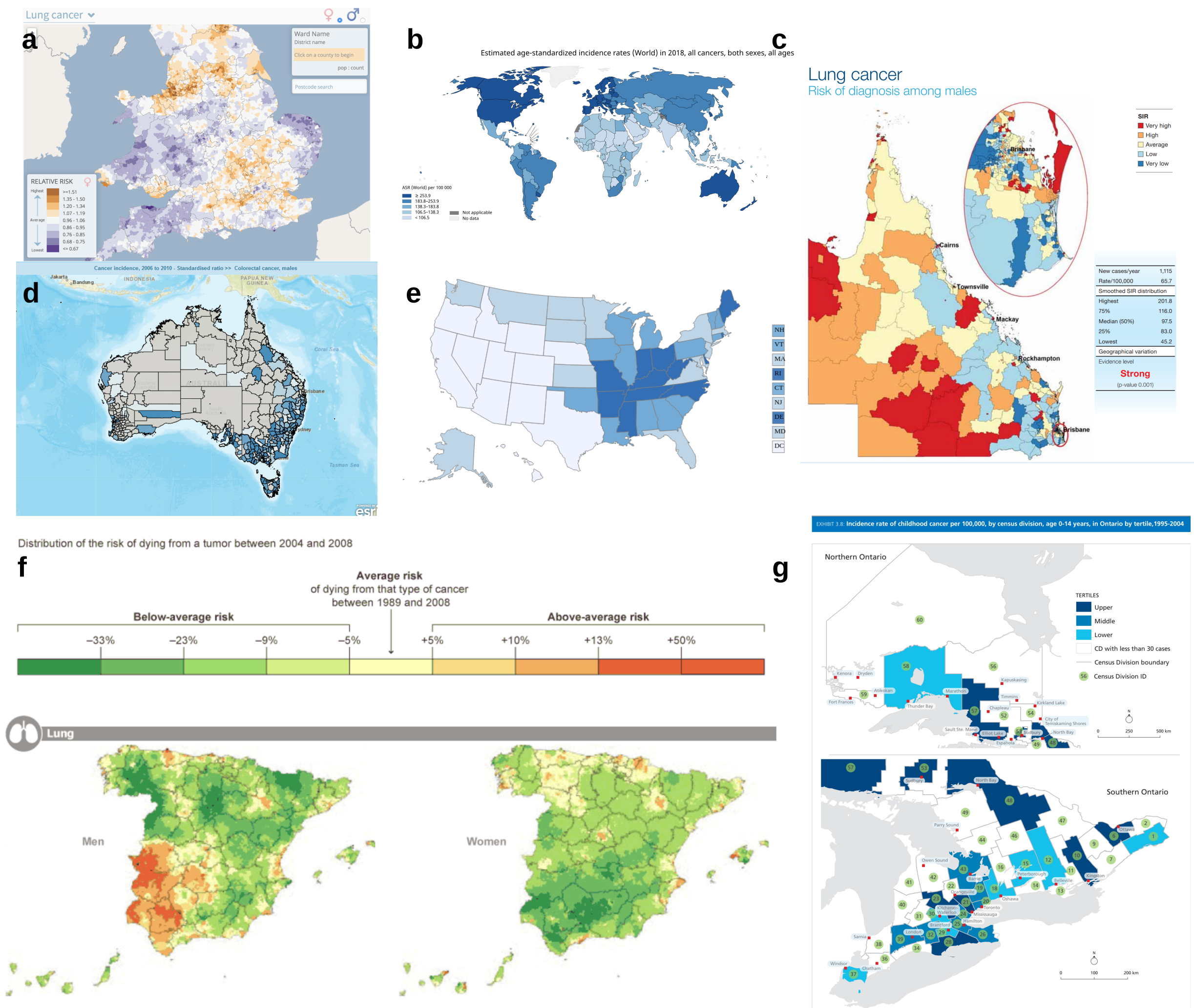

Figure 1: Publicly available choropleth cancer maps published between 2010 and 2015.

visualizations enable non-expert audiences to interpret the outputs of sophisticated statistical analyses. Cruickshank (1947) as cited by S. D. Walter [4], 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.

The statistics presented have been developed over time. 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 [11] 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)	$(IR)_i = \frac{(Incidence\ Rate)_i}{Average\ Incidence\ Rate}$, The cancer incidence rate in region i over the average cancer incidence rate for all of the regions
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	An 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'

Roberts [12] identified 33 publicly available cancer atlases, published between 2010 and 2018, from different countries around the globe. All of these online atlases used 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.

2.1.1 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 the Office for National Statistics (ONS) (England) and the Welsh Cancer Intelligence and Surveillance Unit (WCISU).

2.1.2 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.

2.1.3 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*.

2.1.4 Bowel Cancer Australia Atlas

Figure 1d shows the average Standardised Incidence Ratio of colorectal cancer for Australian males from 2006 to 2010 in Australia [16]. It is published by *Bowel Cancer Australia*.

2.1.5 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.

2.1.6 Map of Cancer Mortality Rates in Spain

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

2.1.7 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].

2.2 Additional considerations

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 [5]. 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, the intuition derived from maps must be 'validated by rigorous statistical analyses' [9], and the display of the supplementary statistics helps in this regard.

2.2.1 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].

2.2.2 Population distribution

Each population area will likely 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 .

2.2.3 Statistical uncertainty

Uncertainty about the value of a statistic is often communicated in cancer atlases. Uncertainty may occur for several reasons: numbers are drawn from samples, errors occur in the disease reporting, or values are simulated from a model to maintain privacy. 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.

2.2.4 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.

2.2.5 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 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].

3 Limitations of choropleth displays

A choropleth display may be limited by the administrative boundaries used to define regions. This issue is often discussed in the context of election mapping, where each electorate decides the outcome of one parliamentary seat in the house or representatives. While all states are visible in a map display of Australia, smaller regions such as electorates present an extreme case of an urban-rural divide. The landmass 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 the majority of the electorates. Using a map of Australia as a visual statistic would communicate an incorrect distrbution of votes across the country. The issue of spatially heterogeneous population distribution across a country's landmass is common to many countries, especially British colonies [22].

Choropleths provide a familiar display. This may be less threatening to an audience that could be frightened by statistics, but they suffer primarily from unfaithfully representing the disease information by hiding small geographic areas. Inset maps like in Brisbane city in Figure 1c of the

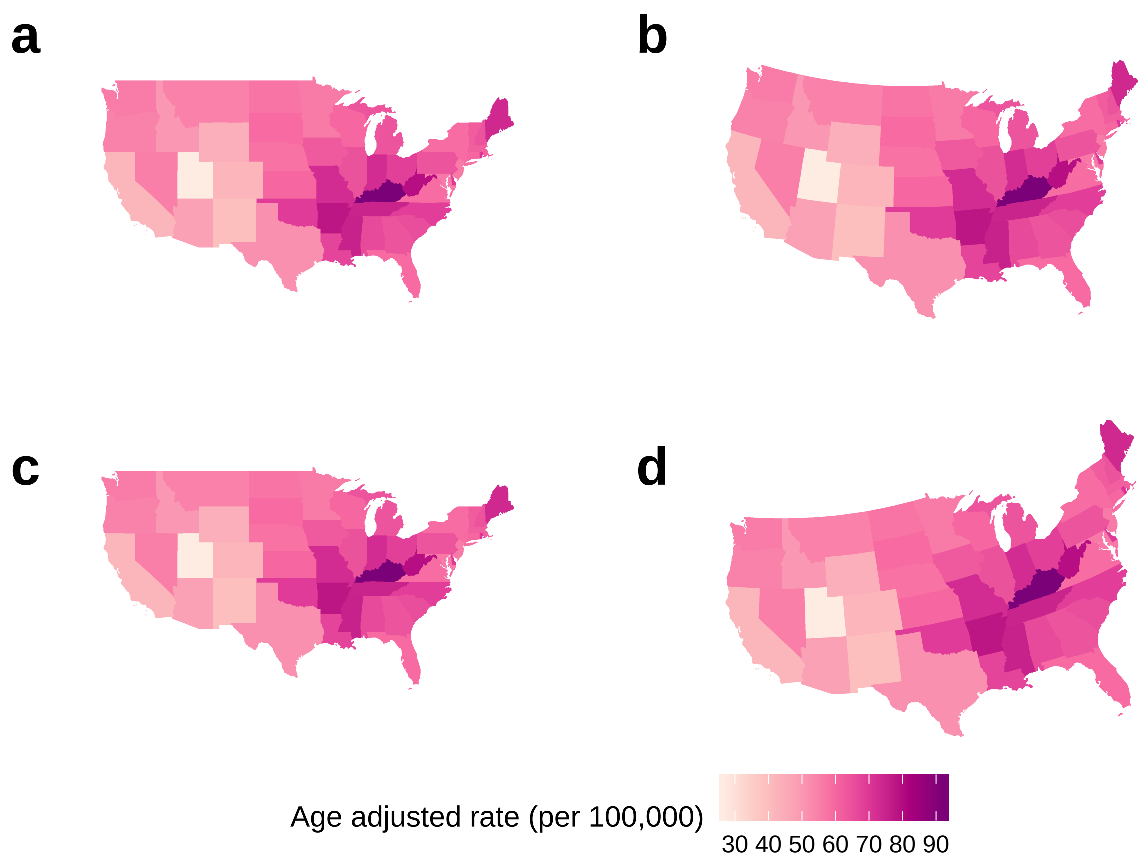

Figure 2: Four choropleth maps average age-adjusted rate of incidence for lung and bronchus in the United States averaged from sex and years 2012-2016, using four different map projections. The map projections alter the shapes and angles of the boundaries of each state.

state of Queensland, are commonly used to reduce distorted interpretations, but it is a bandaid remedy.

As this feature of population distributions continues to intensify, the need for cartograms as an alternative to a choropleth map should only increase.

4 Contemporary 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 [23]. 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. The population then becomes a uniformly distributed background for the statistic presented [24], "population distribution is often extremely uneven in former British colonies" [22], 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 [22]. Monmonier [28] suggests that white lies may be employed to create useful displays and map creators can 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 valueother than the 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, interpreting 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 [22] 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 [4]. Howe [11] suggests that 'cancer occurs in people, not in geographical areas' and 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 [25].

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 [22]. Geographical information systems allowed map users, and researchers to create cartograms, but these systems are utilized depending on 'the effectiveness, efficiency, and satisfaction of the map products (Nielsen 1994), [31]. Howe [11] 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) [22]. 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 it's 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 could 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 neighboring 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 highlighteded 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]. Three methods are provided for creating value-by-area cartograms [33]. They implement 'map deformation' to account for the value assigned to each area. Other methods include Tobler's Pseudo-Cartogram Method, Dorling's Cellular Automaton Method [22], 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 prioritize the shapes of the areas instead of connectivity. Each area stays in a similar position to its location on a choropleth map. The choropleth map base is often also displayed to allow comparisons to be made regarding the change in the area displayed. The addition is the gap between areas, created as each area shrinks or grows according to the associated value of the statistic. Olson [26] discusses the 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 only eliminated on the non-contiguous cartogram but is replaced by the meaningful empty-space property [34].

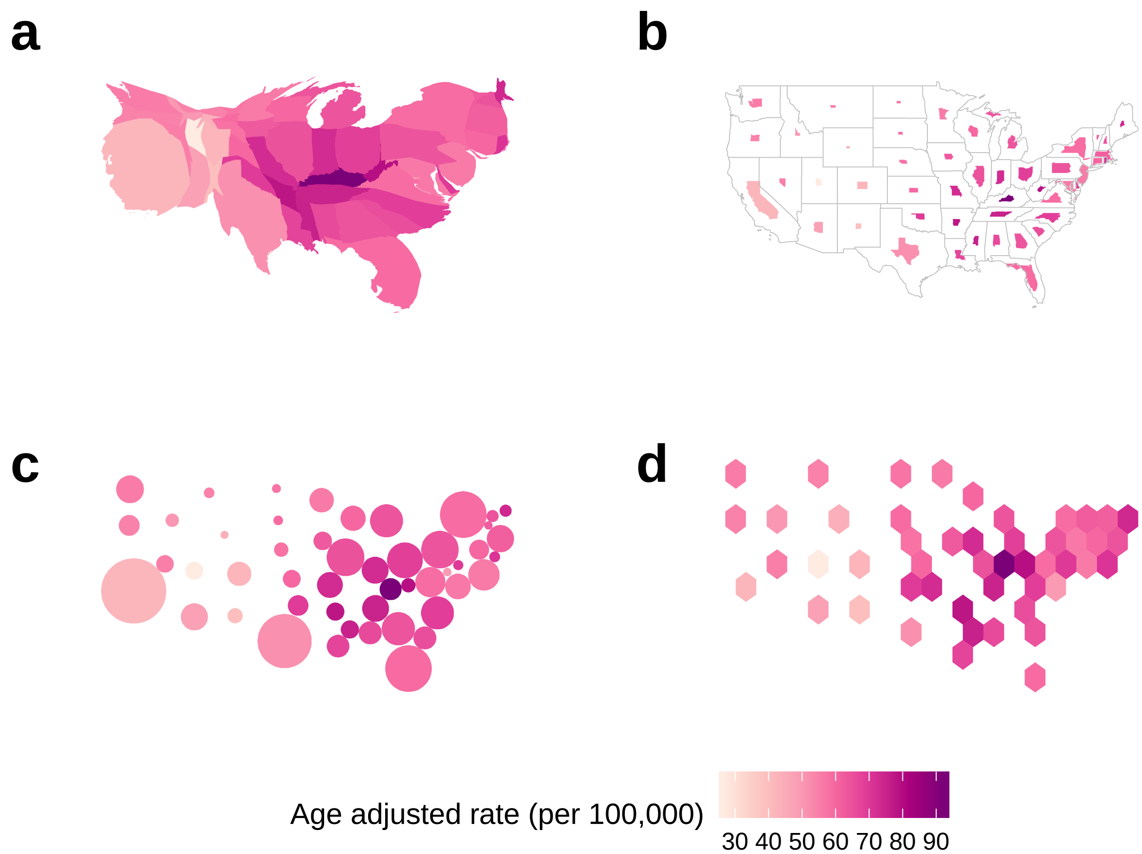

Figure 3: Common alternatives to maps, showing the same information as in Figure efchoroCRS for the United States of America: (a) contiguous cartogram, (b) non-contiguous, shape-preserved cartogram, (c) Dorling cartogram (non-contiguous), (d) hexagon tilemap (non-contiguous). In (a) - (c) the state has 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 hexagon tilemap (d) each state is given equal size and thus equal emphasis.

4.1.3 Dorling

Daniel Dorling presents an alternative display engineered to highlight the spatial distribution and neighborhood 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 [22]:

> “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 resize 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 shown in contiguous cartograms 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 neighboring 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-1930s 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 as the 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 tilemap

A tilemap 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 tilemap 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 labeled or colored to represent a value. tilemaps may be difficult to create, they are best created manually, with additional time and care required as the number of geographic areas to include increases. Cano and others [38] define the term ‘mosaic cartograms’ for hexagonal tile displays, where the number of tiles for each area can be used to communicate the 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 to be made between boundary complexity and simplicity.

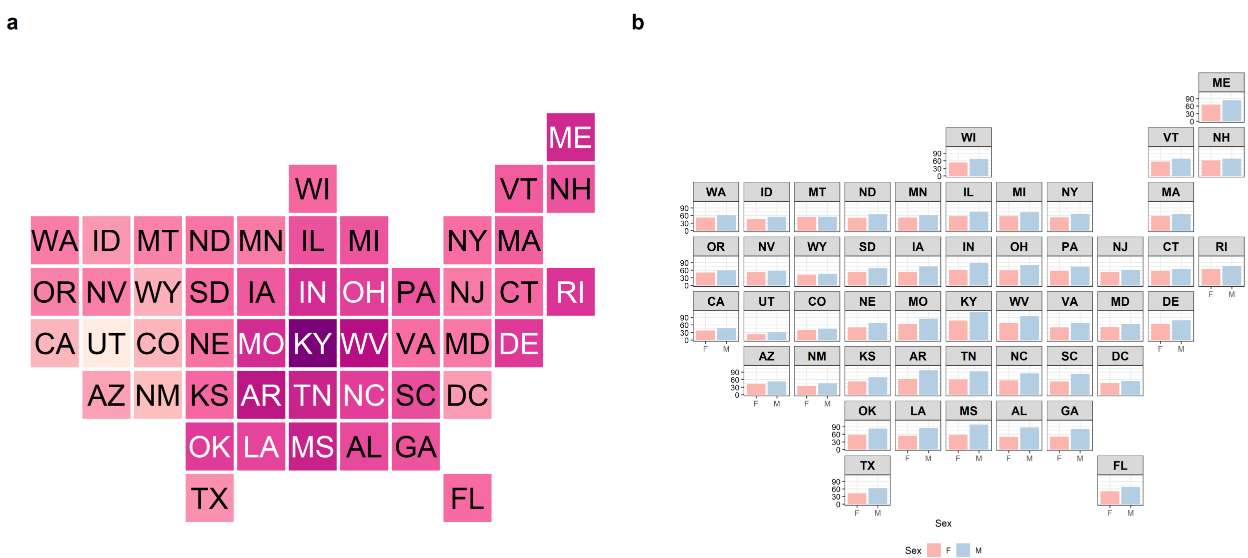

Figure 4: Two more alternative displays, tilemap (left) and geofaceted map (right), showing state age-adjusted rate of incidence for lung and bronchus in the USA. In the tilemap, the layout approximates spatial location, with each state being an equal box filled with color representing cancer incidence. 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 are shown. This type of display allows multiple variables to be displayed for each geographic area.

4.3 Geofacet

Hafen [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 tilemaps, the arrangement of the areas can be reused for other spatial distribution visualizations. Geofaceting has 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 tilemap 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. Manual creation of the grid is required.

4.4 Multivariate displays

Pickle and others [40] present linked micromap plots to visually link geographic and statistical data, this serves as a solution to multi-dimensionality issues. This design has been used in online atlas displays of US states. 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 neighbors are alotted 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 [41] blend color schemes to convey the intersection of categorized levels of an estimate and the associated uncertainty for each spatial area. Lucchesi and Wikle [41] 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.

5 Limitations of alternative displays

5.1 Neither choropleths or cartograms perform well for Australia

Figure 5 shows four main types of cartograms displaying melanoma incidence on SA3 areas in Australia. The version of a contiguous cartogram (a) has expanded the highly populated areas while preserving the full shapes of rural areas. It has not fully resolved the population transformation of areas, and if it had accurately sized areas by population the country would be unrecognizable. The shape-preserved cartogram is unreadable, and all areas have been reduced to tiny spots on the map. Zooming in on a high-resolution output shows that it does indeed preserve the areas though. The Dorling cartogram has lost all geographic context. The hexagon tilemap provides a reasonable spatial distribution despite having too much white-space in the outback areas.

5.2 Pros and cons of different choices

Cartograms provide the spatial distortion to more accurately convey the statistical distribution, focuses on the human impact of the disease. However, if the population density is highly dissonant with geographic density the cartogram will lose all spatial context, and thus futile. Dorling [22] has

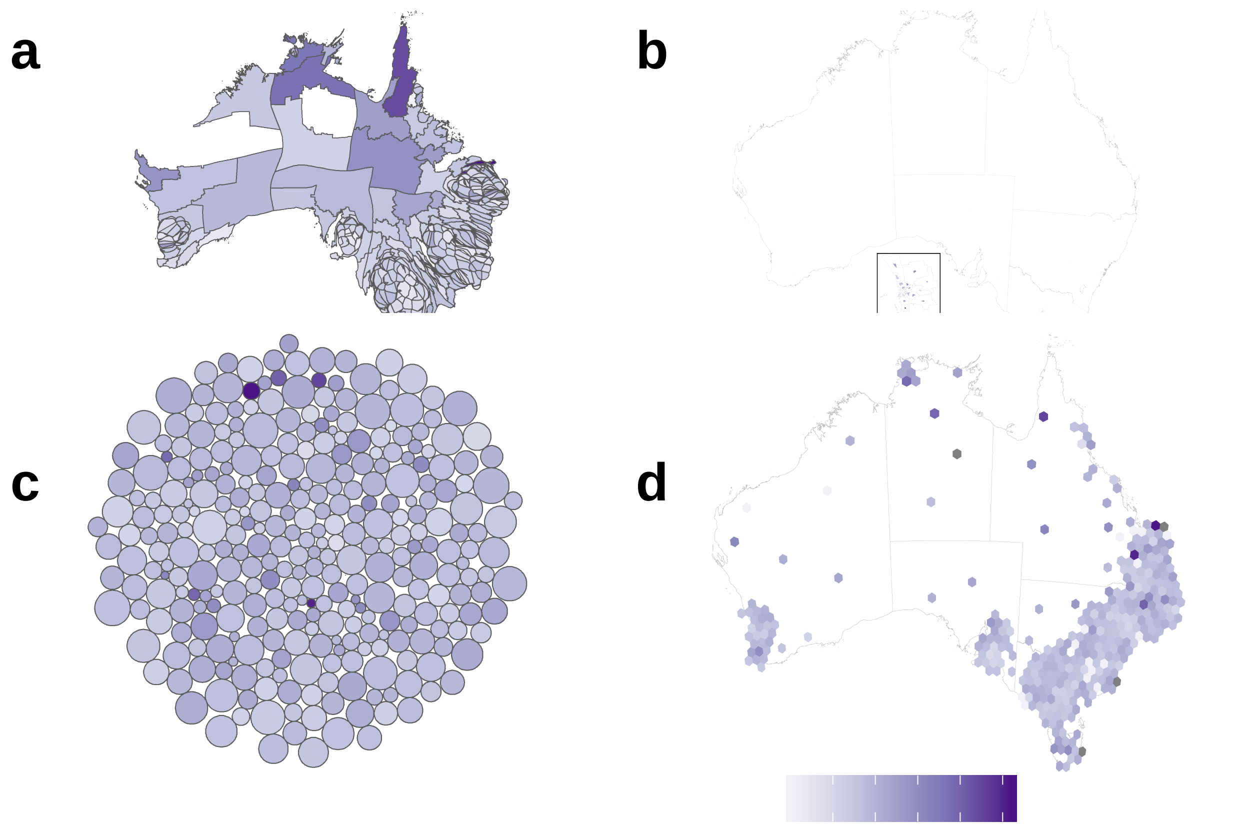

Figure 5: Cartograms showing melanoma incidence in Australia: (a) contiguous, partially population transformed, (b) non-contiguous shape preserved, (c) Dorling, (d) hexagon tilemap. The contiguous cartogram has expanded the highly populated areas while preserving the full shapes of rural areas. If it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

a cartogram showing the 1966 general election results, which looked very little like the geographical shape of Australia.

Some mix of tiling, faceting or even micromaps, which allow some spatial continuity while also zooming into small areas, are good solutions for difficult geographies. Table 3 summarizes the key criteria for evaluating maps and alternative displays. Suggestions and comments to help map creators best communicate their health data and spatial analyses are provided by Moore and Carpenter [1], Bell et al. [5].

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

	Choropleth	Contiguous	Non-contig	Dorling	tilemaps	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

6 User interaction

One of the concerns of adding too much information to a map is the fear of cognitive overload [42] 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 [43]. Interactivity is a design feature within modern mapping methods that can be used to incorporate additional information and complexity without overloading the user. Effective user-centered interactive actions produce rapid, incremental, and reversible changes to the display [44].

Monmonier [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 [45]. This provides additional mechanisms for the users to digest the uncertainty of the available information [46]. When the needs of the audience are changeable and are also the priority, the map creator can allow interactivity for map users to explore a data set through dynamic interactions. This can allow inspection of the data from many views [48]. User interaction with maps helps to understand and interpret the spatial distribution of disease, to validate, explain or explore the presented statistics and their relationships to each other [49].

Interactivity enables supplementary information to be incorporated into online atlases without cluttering the display. Interactive design features, found in online cancer maps, include tooltips, 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 can be found in various online cancer maps [12].

Animation, in contrast to interactivity, usually involves pre-computing views and showing these in a sequence. An overview of animation in association with the R package `gganimate` [50], with some focus on maps, explaining transitions between frames in a linear, to provide a fixed narrative to passive users [51]. 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 [51], the map image is updated/replaced as the animation progresses.

Weather maps are a thoroughly developed example of animation of spatial displays to communicate information to the general public [5]. 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.

The Australian Cancer Atlas [52] provides tours that change the display to the draw users' attention to areas on the map that are relevant to the story. This implementation of animation gives users tools to plan their exploration.

6.1 Interaction and animation in publicly available atlases

Many publicly available cancer atlases employ interactivity and/or animation. Figure 6 shows the interactive GUI tools provided with the public atlases. Mostly these are menus, and toggles, for choosing indicators, and subsets of the population.

6.2 Under-utilized techniques

Figure 7 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 that allows the user to zoom into small areas. It is easy to control and shows precise details in 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.

Many different statistics are commonly used for cancer displays. The most basic is the 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 a low incidence.

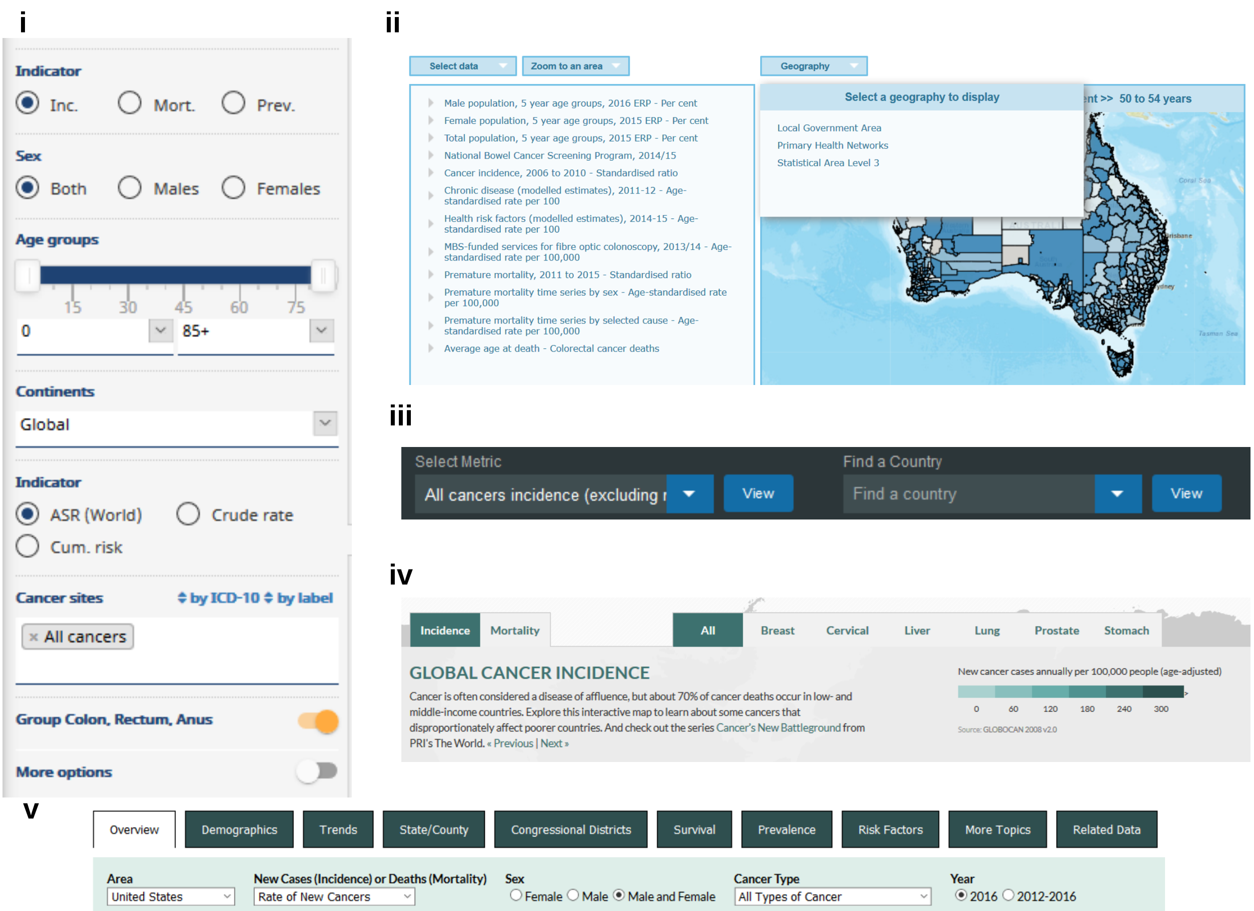

Figure 6: Interactive controls of displays in publicly available choropleth cancer maps: (i) GUI controls for statistic, sex, age groups, continents, and cancer types for Globocan 2018: Cancer Today, (ii) Menus for variable selection and zooming on Bowel Cancer Australia Atlas, (iii) Menus for choosing 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.

I II

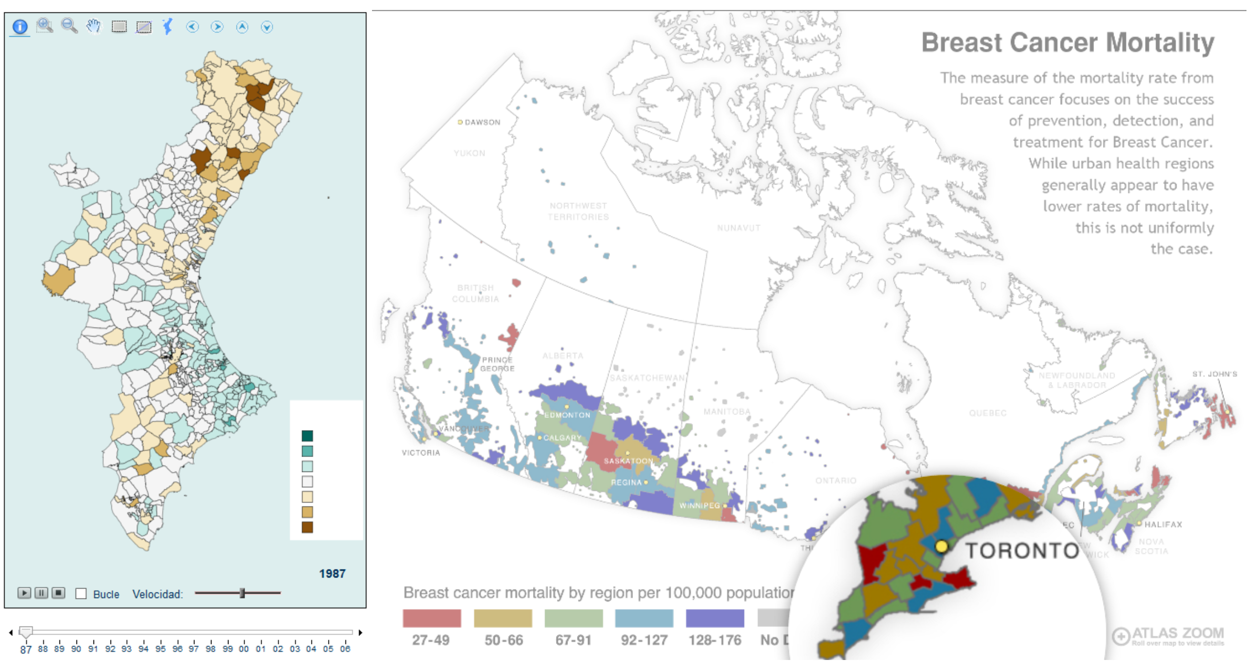

Figure 7: Two examples of advanced interactivity (and animation) in publicly available choropleth cancer maps: a. Linked maps and time-series line plots, with temporal animation in Map of Cancer Mortality Rates in Spain, b. A highly responsive magnifying glass on a map of Breast Cancer Mortality in Canada.

Interaction with maps is an important component of 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 elicit curiosity about the data.

8 Acknowledgements

The authors would like to 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 [53] packages were used to produce this paper: tidyverse [54], RColorBrewer [55], ggthemes [56], png [57], cowplot [58], sf [59], spData [60], cartogram [61], sugarbag [62], knitr [63], rmarkdown [64].

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] Walter SD. Disease Mapping: A Historical Perspective. Oxford University Press; 2001. doi:https://dx.doi.org/10.1093/acprof:oso/9780198515326.003.0012.

[5] 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.

[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] Burbank F. Patterns in Cancer Mortality in the United States 1950-67. NCI, Washington DC: National Cancer Institute Monograph Vol. 33; 1971.

[11] Howe G. Historical Evolution of Disease Mapping in General and Specifically of Cancer Mapping. In:. Cancer mapping, Springer; 1989, pp. 1–21.

[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] 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.

[23] 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.

[24] 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.

[25] Griffin T. 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 D, North S, 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] 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.

[41] Lucchesi L, C.K. W. Visualizing Uncertainty in Areal Data with Bivariate Choropleth Maps, Map Pixelation and Glyph Rotation. Stat 2017. doi:10.1002/sta4.150.

[42] 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.

[43] 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.

[44] Perin C. Direct Manipulation for Information Visualization. Theses. Université Paris Sud - Paris XI, 2014.

[45] Goodchild M, Buttenfield B, Wood J. On Introduction to Visualizing Data Validity. Visualization in Geographical Information Systems 1994:141–9.

[46] MacEachren AM. Visualizing Uncertain Information. Cartographic Perspectives 1992:10–9. doi:https://doi.org/10.14714/CP13.1000.

[47] 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.

[48] 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.

[49] 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.

[50] Pedersen TL, Robinson D. gganimate: A Grammar of Animated Graphics. 2019.

[51] Pedersen TL. The Grammar of Animation 2018. https://youtu.be/21ZWDrTukEs (accessed November 16, 2018).

[52] Cancer Council Queensland, Queensland University of Technology, and Cooperative Research Centre for Spatial Information. Australian Cancer Atlas 2018. https://atlas.cancer.org.au.

[53] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.

[54] Wickham H. tidyverse: R packages for data science 2017. https://CRAN.R-project.org/package=tidyverse.

[55] Neuwirth E. RColorBrewer: ColorBrewer palettes 2014. https://CRAN.R-project.org/package=RColorBrewer.

[56] Arnold JB. ggthemes: Extra Themes, Scales and Geoms for 'ggplot2' 2019. https://CRAN.R-project.org/package=ggthemes.

[57] Urbanek S. png: Read and write PNG images 2013. https://CRAN.R-project.org/package=png.

[58] Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2' 2019. https://CRAN.R-project.org/package=cowplot.

[59] 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.

[60] Bivand R, Nowosad J, Lovelace R. spData: Datasets for Spatial Analysis 2019. https://CRAN.R-project.org/package=spData.

[61] Jeworutzki S. cartogram: Create Cartograms with R 2018. https://CRAN.R-project.org/package=cartogram.

[62] Kobakian S, Cook D. sugarbag: Create Tessellated Hexagon Maps 2019. https://CRAN.R-project.org/package=sugarbag.

[63] Xie Y. knitr: A General-Purpose Package for Dynamic Report Generation in R. 2019.

[64] Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, et al. rmarkdown: Dynamic Documents for R. 2019.