Cancer Applications of Choropleth maps, and the Potential of Cartograms and Alternative Map Displays

Stephanie Kobakian, Jessie Roberts and Dianne Cook

Abstract

Cancer atlases communicate cancer statistics over geographic domains. They subdivide these domains into administrative regions such as countries, states, or suburbs. When communicating human-related statistics in Australia, the geographic map base draws attention to 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 be more effective than a choropleth map at revealing spatial patterns. Alternative displays are effective for the communication of election data aggregated at the state or electorate level on online news sites.

1 Introduction

Researchers, health authorities, governments and not-for-profits are common communicators of cancer statistics. They often present statistics to the public as aggregated values for geopolitical areas. Presenting these statistics requires aggregating individual observations into the statistics for the geographical units, especially for privacy protection, but also for political and policy purposes. Data collection or aggregation regions for the amount of cancer diagnoses or excess deaths per area include states, provinces, local government areas, and post/zip codes. To provide the amount, or count, of the diagnoses or excess deaths with these areas, it needs minimal transformation. However, the amount of people living in each of the areas in a set will be different. Those who wish to calculate rates can use the amount of people living in an area as a denominator and the diagnoses can become a numerator to create rates per amount of people. Health registries routinely collect data for public health reasons and may make this available to the public as a service to the community.

Choropleth map displays are commonly used for the communication of cancer statistics over geographic domains. In this display, map creators map a statistic to color and apply the colour to the polygons representing geographic regions. Map users may overlook geographic patterns of disease when obscured in a table [1]. Visualizing diseases on maps is often the first step in exploratory spatial data analysis and helps in the formulation of hypotheses. Providing a visual representation of cancer outcomes allows identification, and public health policy and actions can address these patterns. choropleth maps allow the examination of the spatial distribution of the disease incidence, this may reveal a trend in longitude or latitude, differences between rural areas and urban, or coastal and inland areas, or even specific hot spots of the disease. Daniel Exeter [2] recognizes that 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 members of the public and their applications to cancer statistics. It also considers alternative approaches and compares the visualization methods, highlighting the differences and historic use of these displays.

The paper is structured 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. Section 3 discusses the limitations of the choropleth map. Section 4 describes alternative displays, including the cartogram, which is useful when the map has heterogeneously sized geographic units. Section 5 presents the limitations of the production and use of alternative displays. Disease maps are more useful when made interactive, Section 6 describes common options, along with a discussion of benefits and disadvantages.

2 Traditional approaches for cancer map displays {#ch:choropleth maps}

A choropleth is traditionally used to make a cancer map. It is used to show the geographical distribution of data values by shading areas of a map [3]. A choropleth map involves faithfully rendering the geography, and the design of the color scheme and rendering 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. choropleth maps can be used for displaying disease data [4], including cancer data [5]. In epidemiology, choropleth maps 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 official federal statistics provide cancer incidence and deaths [5], 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. Utah is a spatial outlier, it 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 easier to read [6] and allow viewers to infer the spatial relationships visually in the data using their mental model of the geography. Consumers of disease displays may include researchers, members of the public, policymakers, and the media [5]. For these consumers, the familiarity of the geography is a worthy consideration when presenting results of spatial analysis.

Politically driven boundaries produced the now familiar shapes or regions, this has created individual areas of non-uniform size. These areas have different population densities and are subject to change. The different population and geographical sizes of administrative areas can attract attention to the shades of the underpopulated but large areas [3], Skowronnek [7] calls this an area-size bias. choropleth maps 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 [8]. The data collection methods across regions and the administrative control within regions lends itself to choropleth visualization. Cancer maps and atlases dating back to Haviland's maps in 1875, and early work in US cancer atlases appearing in 1971 [9]. The presentation of cancer statistics has

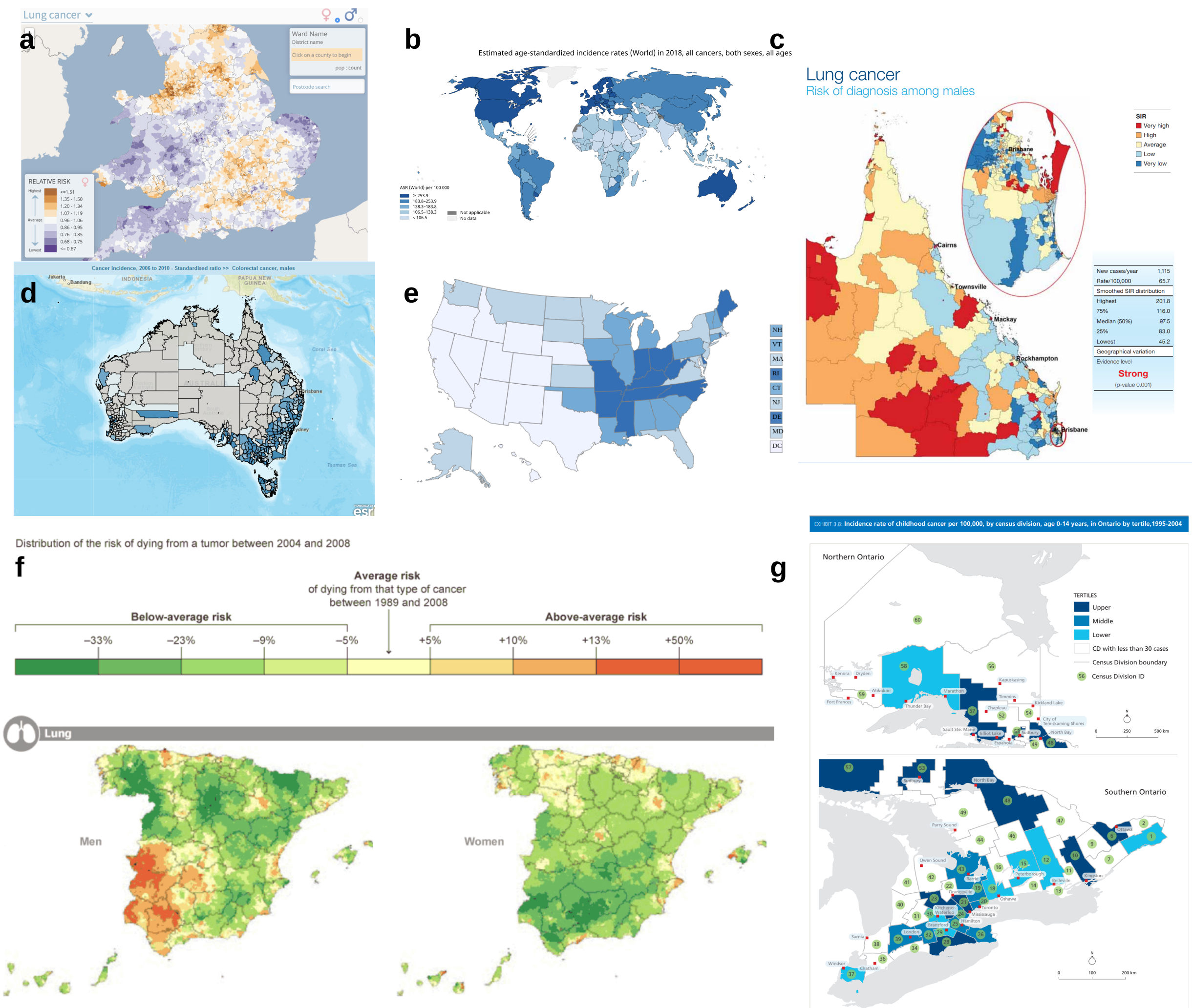

Figure 1: Examples of publicly available choropleth cancer maps.

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 analysis. These visualizations enable non-expert audiences to interpret the outputs of sophisticated statistical analysis. 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 used to communicate to members of the public and other non-expert audiences are choropleth maps.

Epidemiologists and statisticians have developed these statistics over several decades. Mortality rates commonly 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: 1 Common 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 [11] identified 33 publicly available cancer atlases, published between 2010 and 2018, from different countries around the globe. These online atlases used choropleth maps. Non-commercial organizations, including not-for-profits, government, research organizations, advocacy groups or government-funded partnerships published all but one of these. Figure 1 displays a subset of maps from these atlases.

The selection varies in the geographies explored, Figure 1b, Globocan 2012 [12] explores Estimated Cancer Incidence, Mortality and Prevalence Worldwide utilizing data sourced from cancer registries of each country. The Bowel Cancer Australia Atlas presents an example of a cancer specific atlas, Figure 1d shows the average Standardized Incidence Ratio of colorectal cancer for Australian males from 2006 to 2010 in Australia [13]. Like many of the atlases examined, there is a choice of gender displayed in the Bowel Cancer Atlas. An alternative approach used in the 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 [14].

There are vast differences in the geographies considered by online atlases. Figure 1b shows global age-standardized incidence rates (per 100,000) for all invasive cancers for both men and women, aggregated at a national level for 2018. The United States Cancer Statistics [15] presented in Figure 1e 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 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 [16] at a neighborhood (small-area) scale. The Atlas of Cancer in Queensland, Figure 1c shows the relative incidence ratio of lung cancer in males for each Statistical Area at Level 2 in the state of Queensland within Australia [17].

Age-specific atlases are less common, Figure 1g displays Atlas of Childhood Cancer in Ontario, this communicates the incidence rate of childhood cancers per 100,000 (by census division) for children aged 0-14, in Ontario from 1995 to 2004 [18]. This atlas communicates information regarding a tiny geographic region.

2.2 Additional materials

Including supplementary graphs and plots add more depth and information. Additional materials such as tables, graphs, and text explanations, support the 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 display of the supplementary statistics helps. The intuition derived from maps must be 'validated by rigorous statistical analyses' [8].

The statistics communicated in atlases are often used to describe differences between areas. This can occur at different levels of aggregation. Aggregation of global health statistics occurs within 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 [15]. Each population area will probably have a different number of people. Cancer atlases may also communicate the distribution of the population living in all areas in a table or histogram display [20]. Atlases can connect the population to the land available to them by communicating population density. Section 4 considers alternatives to choropleth that introduce population into the visualizations.

Map users can define the population further using the demographics of the individuals within each area. Demographics that can influence cancer rates include information regarding the age and sex distribution of the areas presented. Some digital atlases surveyed allow users to control the display to select subsets such as males, females or those aged over 65. Similarly, socioeconomic 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. Introducing population and demographic information helps to interpret the rates in areas effectively, but there will still be uncertainty around the rates. To address this, cancer atlases often communicate uncertainty about the value of a statistic. Cancer atlas creators may introduce uncertainty when using samples, or there are errors occurring in the disease reporting, or models are used to simulate values to maintain privacy. The statistics that report on uncertainty include sample or population size, standard deviation, confidence intervals, statistical significance, quantiles, and credible intervals. The most common measures 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 is 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 Limitations of choropleth displays

The administrative boundaries used to define regions may limit a choropleth display. Those who critique choropleth maps often dicuss election mapping, where each electorate decides the outcome of one parliamentary seat in the house of representatives. While all states are visible in a map display of Australia, smaller regions such as electorates present an urban-rural divide, where rural regions are much larger than metropolitan areas of similar or greater populations. 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',XXX as it means diminishing the visual impact of most of the electorates. Using a map of Australia as a visual statistic would communicate an incorrect distribution of votes across the country.

Spatially heterogeneous population distribution across a country's landmass is an issue common to many countries, especially British colonies [22]. However, the boundaries shown in choropleth maps provide a familiar display. This may be less threatening to an audience that could be unfamiliar was interpreting spatial statistics, even though 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 state of Queensland are commonly used to reduce distorted interpretations, but it is a band aid 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 choropleth maps

4.1 Cartograms

choropleth maps imply uniformity of data across the geographic space but population densities are unlikely to be uniform [7]. Cartographers developed the cartogram to draw the attention to the population by transforming the map [23]. The resulting display can communicate the impact of the disease more accurately across the population, as recorded by the statistic, at the sacrifice of geographic accuracy.

When a map creator desires a uniform population density of the map base, the purposeful distortion of the map space is beneficial. The "population distribution is often extremely uneven in former British colonies" [22], and this makes the distortion necessary, [24] population will then become a uniformly distributed background for the statistic presented [25]. An area cartogram [26], or population-by-area cartogram [27] occurs after the distortion of the geographical shape according to population. 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 map creators can use

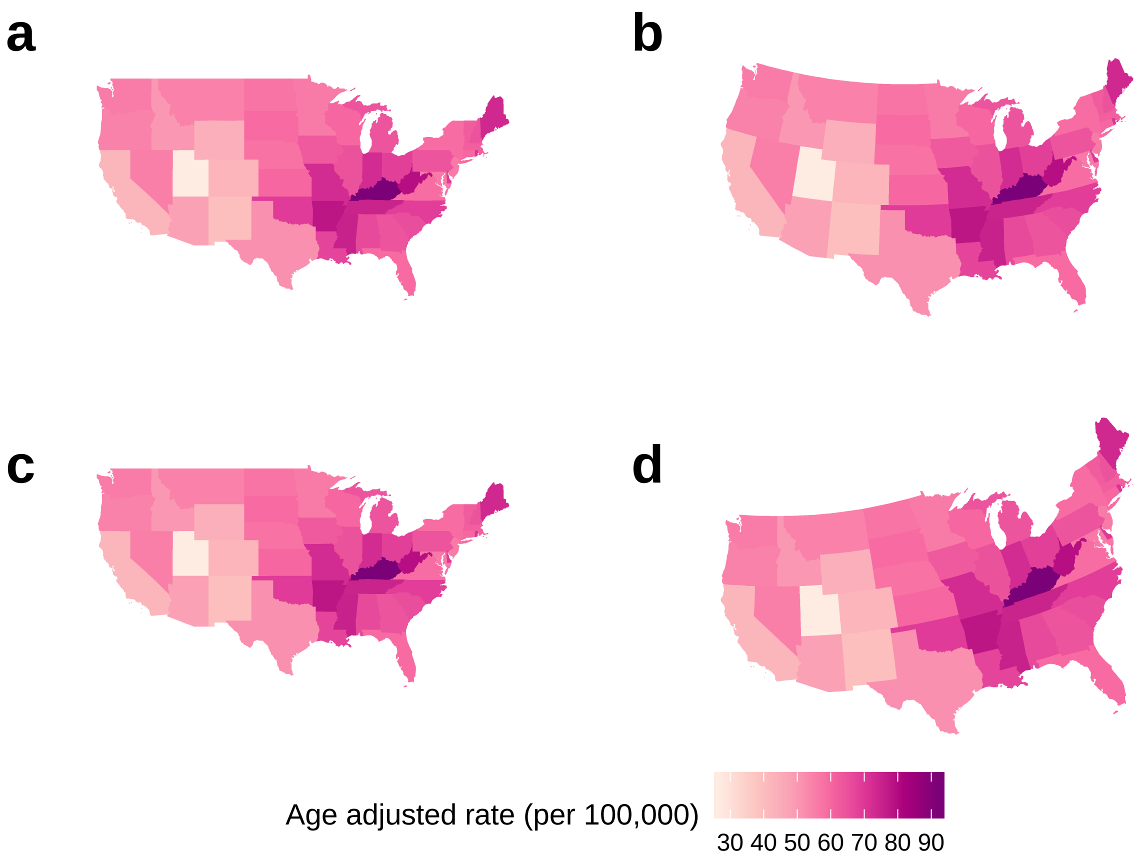

Figure 2: Four choropleth maps average age-adjusted rate of incidence for lung and bronchus in the United States averaged from sex and years 2012-2016, using four different map projections. The map projections alter the shapes and angles of the boundaries of each state.

white lies to create useful displays by distorting the geometry and suppressing features. It is easy for the average person to disregard the impact of transformations used to create cartograms. One disadvantage of the conventional map is interpreting spatial patterns is difficult because of the sparsely populated rural areas being emphasized in a choropleth map, and the inner city areas are geographically small. The distortion of each area when making a cartogram is proportional to a value other than the actual earth size area [26]. The distortion in an area 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 design 'more socially just', or honest [30], giving equitable representation and attention to all members of the population and reducing the visual impact of large areas with small populations [4]. Howe [10] 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 [24].

The creation of cartograms was historically in the hands of professional cartographers [31]. Early approaches by 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]. Howe [10] discusses the impact of electronic computer-assisted techniques. Geographical information systems allow map creators to produce cartograms and they use these systems depending on 'the effectiveness, efficiency, and satisfaction of the map products' (Nielsen 1994) [31].

There are many issues to consider when creating alternative map displays: the intended audience of the map, and its purpose are key points in cartogram use and creation as '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: 2 Maps used to present statistics for the United States of America. The colour of each state communicates 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	It has distorted each state's shape according to the population of the state in 2015. The state of California has become much larger because of the large population density. This draws attention to the densely populated North-East region and detracts from the less populated Mid West.
b. Non-contiguous	It maintains the geographic shape of the states, 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, for Massachusetts and Connecticut. This draws attention to the densely populated North-East region and the sparse Mid West.

Map display	Details
c. Dorling	Circles are used to represent each state, but the population of the state determines the size in 2015. The North-East states remain closer to their neighbors and are slghtly displaced from their geographic location. It highlights the sparsity of the population in the Mid West by the distance between the circles at the geographic centroids.
d. Hexagon Tessellation	A hexagon of equal size represents each state. It is easy to contrast the neighboring states however the North-East regions have been displaced from their geographic location. It highlights the sparsity of the population in the Mid West 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. The information in Table 2 summarizes what can be observed in the four types of cartograms.

4.1.1 Contiguous

A contiguous cartogram alters the choropleth according to a statistic and maintains connectivity of the map regions. This transformation often occurs at the expense of the shape of areas [29]. Min Ouyang and Revesz [33] present three algorithms for creating value-by-area cartograms. 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].

a shows a population contiguous cartogram of the United States. All states are visible and the shape of the United States overall is still recognizable. In contrast, Figure 5 a shows an Australian contiguous cartogram also based on population. The south east is completely distorted, areas with low population are still large very the map, as their initial area was used to determine the area allocated in the contiguous cartogram.

To be able to recognize the significant changes, a reader will usually have to know the initial geography to find the differences in the new cartogram layout [26]. Tobler's Conformal mapping means to preserve angles locally so that the shapes of small areas on a traditional map and a cartogram would be similar. [29] presents this issue as conflicting tasks or aims, to adjust region sizes and retain region shapes. –>

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. Displaying the choropleth map base allows map users to make comparisons regarding the change in the area. 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 'difference from the original map that is the real message' of these displays.

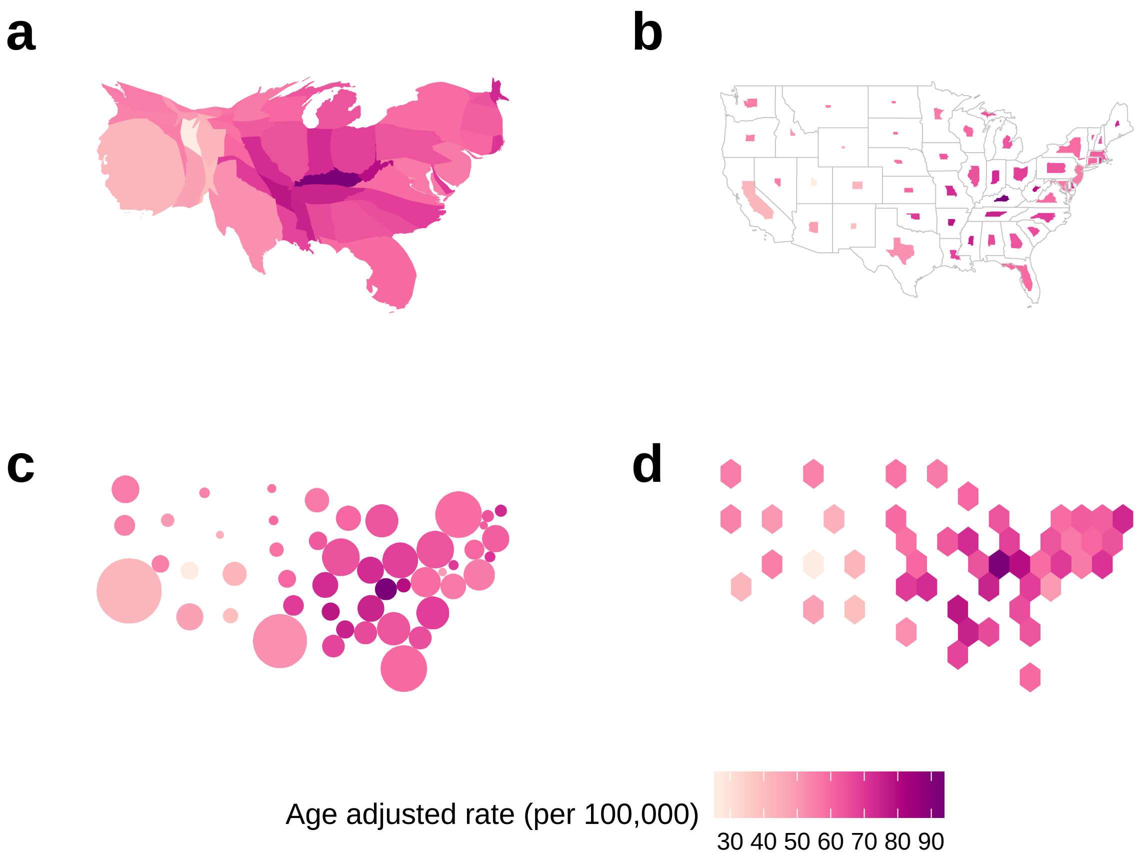

Figure 3: Common alternatives to maps, showing the same information as in Figure efchoroCRS for the United States of America: (a) contiguous cartogram, (b) non-contiguous, shape-preserved cartogram, (c) Dorling cartogram (non-contiguous), (d) hexagon tilemap (non-contiguous). Maps (a) - (c) are created by resizing and reshaping the states of the USA 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 ease losing information about physically small but population-dense states. Map creators give each state equal size and thus equal emphasis in (d) the hexagon tile map.

The distortion of region shapes on the contiguous cartogram presents an additional hurdle to visual recognition, and the non-contiguous cartogram eliminates this hurdle, and presents the meaningful empty-space property [34]. b shows small overlaps in the North-East region of the United States. These overlaps are also shown in the non-contiguous cartogram of Australia in Figure 5 b. –>

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. Daniel Dorling puts forward a simple question and his approach opposes preserving the intricate shape details [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. Contiguous cartograms use a gravity model to apply distortions to create the 'nonsense' shapes that have 'no meaning' [30]. This method applies a gravity model to produce a compelling map, that avoids overlaps and keep spatial relationships with neighboring areas over many iterations. This implementation can work for up to 'one hundred thousand' areas.

Raisz laid the groundwork for this approach in the mid-1930s, and his 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 may sacrifice contiguity but allows for tiling where geographic neighbors placed in suitable relative positions also share borders [36]. Rectangular cartograms communicate bi-variate displays of the population by the size of each rectangular, and they use color to communicate a second variable [37].

A similar method to a rectangular cartogram, represents each geographic area using a square. The squares are tessellated to create a 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. d shows a hexagon tilemap representing each state using a single hexagon.

Recommended criteria to contrast mapping methods include average cartographic error, and maximum cartographic error, correct adjacency, maximum aspect ratio, and suitable relative positions [37]. However, this does not consider the issues with actually producing rectangular cartograms. Algorithms for the creation of rectangular cartograms –>

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 that use a consistent measure of population for each tile. Regions with over four neighbors require some necessary displacement. The tile map uses color to represent a value of a statistic for each area. A similar method to a rectangular cartogram represents each geographic area using a square of the

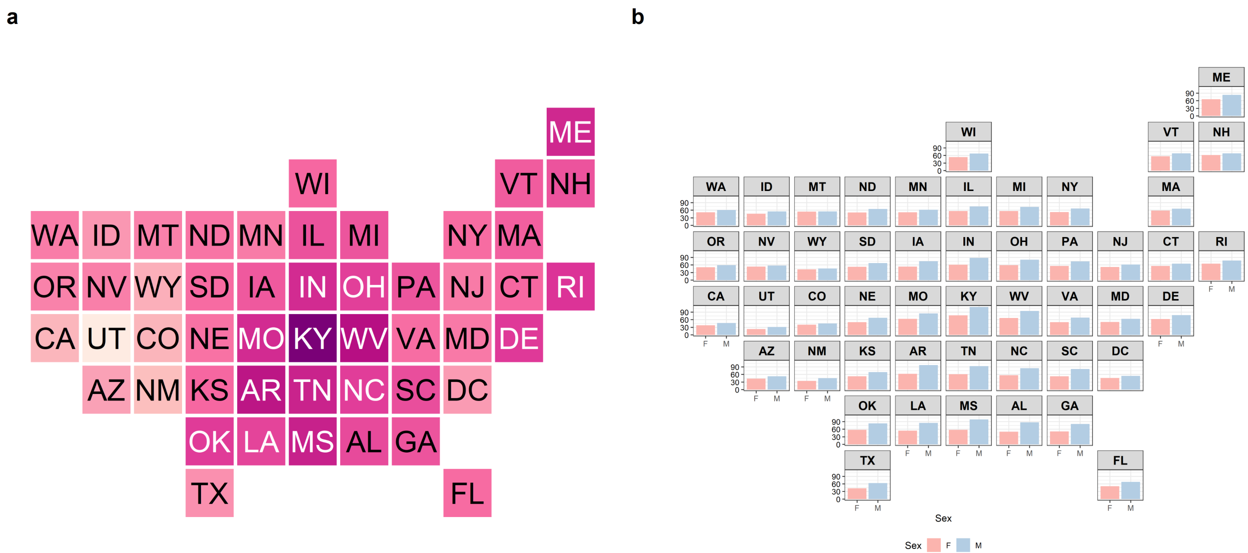

Figure 4: Two alternative displays, tile map (left) and geofaceted map (right), showing state age-adjusted rate of incidence for lung and bronchus in the USA. In the tile map, the layout approximates spatial location, with each state being an equal box filled with color representing cancer incidence.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

same dimensions. The squares tessellate to create a grid. There are online media sources using this method, such as FiveThirtyEight, Bloomberg Business, The Guardian, The Washington Post, The New York Times and NPR. Tile maps may be difficult to create as they are best created manually, they require additional time and care 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 or the color of them can communicate the statistic of regions. When using several tile per regions, map makers can adjust the complexity of the boundaries in the resulting display. They can also make a trade-off between boundary complexity and simplicity by the size of the tiles used.

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 tile maps, they can reuse the arrangement of the areas for other spatial distribution visualizations. Geofaceting has the functionality of facets, often used to replicate visualizations for each subset of the data. Here, the data subsets are geographic. A tile map communicates one value per region in a visualization, geofaceting is a more flexible visualization for communication as it increases the amount of information displayed. Virtually any plot display can be shown in the tile representing the areas, allowing displays of multiple variables or values per geographic entity. Creating a tile map or a geofacet both require manual creation of the grid, except for the grids that other have already contributed.

4.4 Multivariate displays

Pickle and others [40] present linked micro map plots to match geographic and statistical data visually, this serves as a solution to multi-dimensionality issues. An online atlas of US states uses this display. These maps group areas 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 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 micro maps predominantly use the choropleth map for displays of spatial relationships. These maps show spatial relationships by allotting spatial neighbors 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 suggest map pixilation, it breaks each region into small pixels, and allocates values to the individual pixels to create texture. This reflects the uncertainty around the area's estimate by randomly sampling from the confidence interval of the estimate of the area. Animating these displays involves resampling the pixels for each frame. Areas with uncertain values will flicker more dramatically than areas with more certain values.

5 Comparison and critique of alternative displays

5.1 Neither choropleth maps 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 it has reduced all areas to tiny spots on the map. Zooming in on a high-resolution output shows it does preserve the areas. The Dorling cartogram has lost all geographic context. The hexagon tile map 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 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 micro maps, which allow some spatial continuity while also zooming into small areas, are good solutions for difficult geographies. Table 3 summarizes the key criteria for testing maps and alternative displays. Moore and Carpenter [1] and Bell et al. [5] provide suggestions and comments to help map creators best communicate their health data and spatial analysis.

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. –>

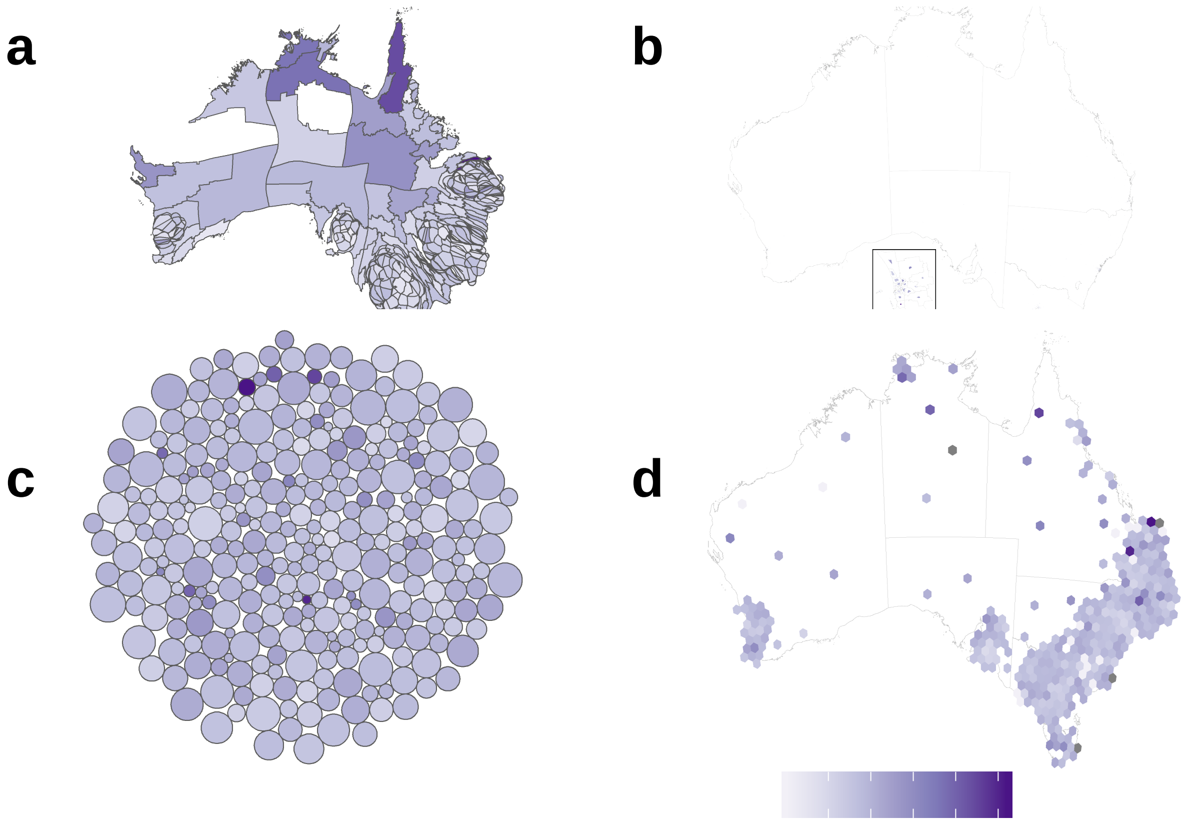

Figure 5: Cartograms showing melanoma incidence in Australia: (a) contiguous, partially population transformed, (b) non-contiguous shape preserved, (c) Dorling, (d) hexagon tile map. The contiguous cartogram has expanded the highly populated areas while preserving the full shapes of rural areas. If it accurately sized areas by population, the country would be unrecognizable. The shape-preserved is unreadable due to the small area sizes. The Dorling cartogram presents all areas but many are difficult to compare. The hexagon tile map provides a reasonable spatial distribution despite having isolated hexagons in the outback areas.

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	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

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-centred 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 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 can be found in various online cancer maps [11].

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 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.

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.

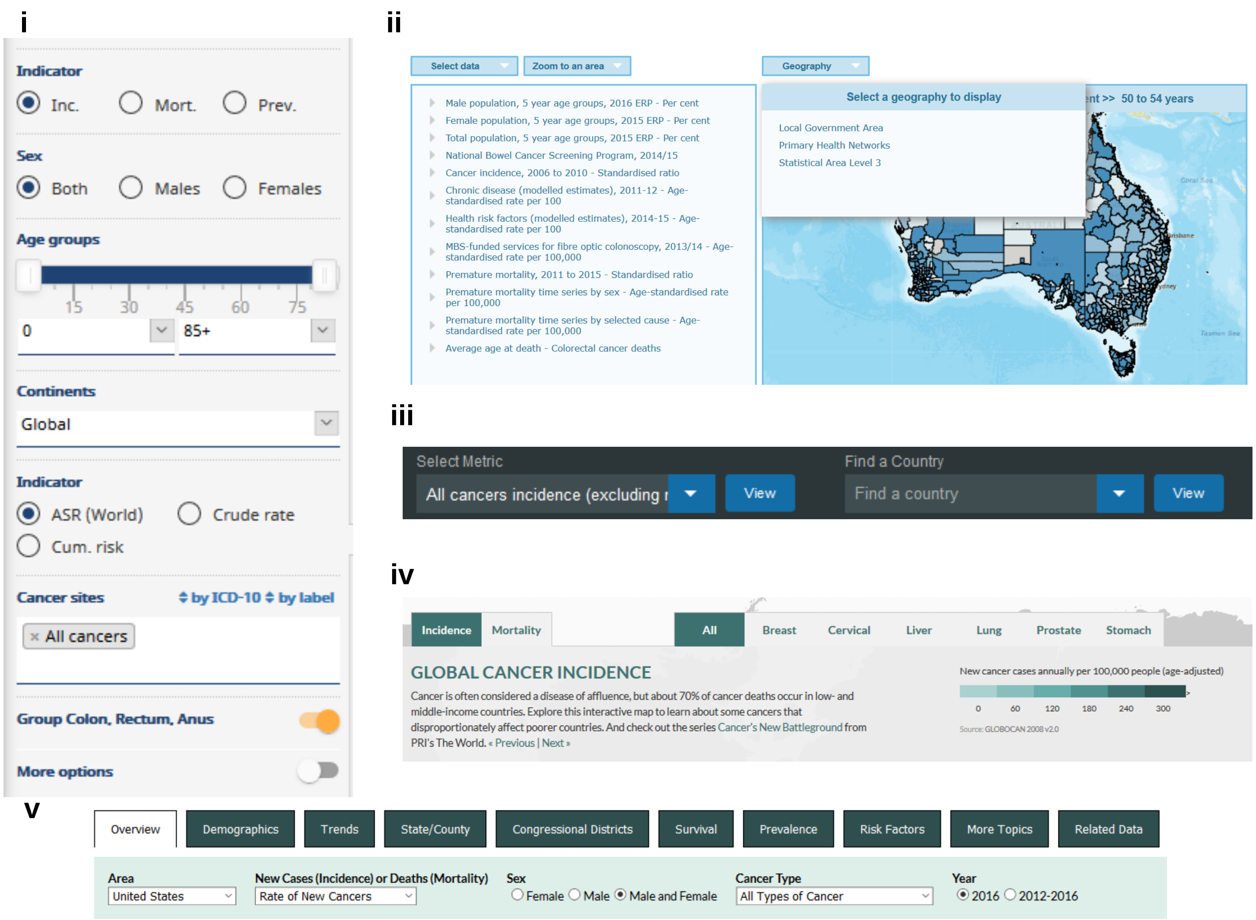

Figure 6: Interactive controls of displays in publicly available choropleth cancer maps: (i) GUI controls for statistic, sex, age groups, continents, and cancer types for Globocan 2018: Cancer Today, (ii) Menus for variable selection and zooming on Bowel Cancer Australia Atlas, (iii) Menus for choosing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

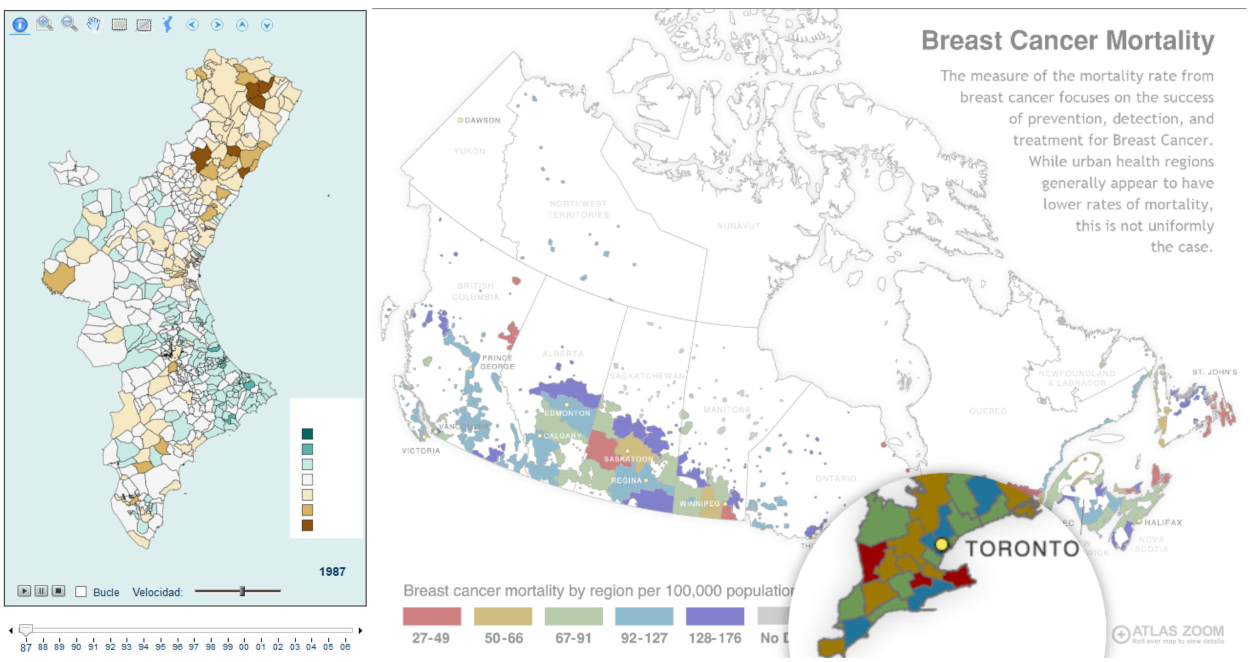

Figure 7: Two examples of advanced interactivity (and animation) in publicly available choropleth cancer maps: a. Linked maps and time-series line plots, with temporal animation in Map of Cancer Mortality Rates in Spain, b. A highly responsive magnifying glass on a map of Breast Cancer Mortality in Canada.

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 (1999) Spatial Analytical Methods and Geographic Information Systems: Use in Health Research and Epidemiology. Epidemiologic Reviews 21:143–161

2. Exeter DJ (2016) Spatial Epidemiology. International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology 1–4

3. Tufte ER (1990) Envisioning Information. Graphics Press

4. Walter SD (2001) Disease Mapping: A Historical Perspective. https://doi.org/https://dx.doi.org/10.1093/acprof:oso/9780198515326.003.0012

5. Bell BS, Hoskins RE, Pickle LW, Wartenberg D (2006) Current Practices in Spatial Analysis of Cancer Data: Mapping Health Statistics to Inform Policymakers and the Public. International Journal of Health Geographics 5:49

6. Brewster MB, Subramanian SV (2010) Cartographic Insights into the Burden of Mortality in the United Kingdom: A Review of "The Grim Reaper's Road Map". International Journal of Epidemiology 39:1120–1122

7. Skowronnek A (2016) Beyond Choropleth Maps – A Review of Techniques to Visualize Quantitative Areal Geodata. In: Infovis Reading Group WS 2015/16. https://alsino.io/static/papers/BeyondChoropleths_AlsinoSkowronnek.pdf.

8. d'Onofrio A, Mazzetta C, Robertson C, Smans M, Boyle P, Boniol M (2016) Maps and Atlases of Cancer Mortality: A Review of a Useful Tool to Trigger New Questions. Ecancermedicalscience 10:670–670

9. Burbank F (1971) Patterns in Cancer Mortality in the United States 1950-67. National Cancer Institute Monograph Vol. 33, NCI, Washington DC

10. Howe G (1989) Historical Evolution of Disease Mapping in General and Specifically of Cancer Mapping. In: Cancer mapping. Springer, pp 1–21

11. Roberts J (2019) Communication of Statistical Uncertainty to Non-expert Audiences. https://doi.org/10.5204/thesis.eprints.130786

12. World Health Organization's International Agency for Research on Cancer (2017) Globocan 2012: Estimated cancer incidence, mortality and prevalence. http://globocan.iarc.fr/Pages/Map.aspx. Accessed 26 Sep 2019

13. Bowel Cancer Australia (2016) Bowel Cancer Australia Atlas. http://www.bowelcanceratlas.org/. Accessed 26 Sep 2019

14. El Pais (2014) Map of Cancer Mortality Rates in Spain. http://elpais.com/elpais/2014/10/06/media/1412612722_141933.html. Accessed 26 Sep 2019

15. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute - Cancer Statistics Working Group (2019) U.S. Cancer Statistics Data Visualizations Tool (data 1999-2016). http://www.cdc.gov/cancer/dataviz. Accessed 26 Sep 2019

16. Emperial College London - Small Area Health Statistics Unit (2010) The environmental and health atlas of england and wales: National male lung cancer rate. http://www.envhealthatlas.co.uk/eha/Breast/. Accessed 26 Sep 2019

17. Queensland Cancer Registry (2011) The Atlas of Cancer in Queensland (1998 - 2007). https://cancerqld.org.au/research/queensland-cancer-statistics/queensland-cancer-atlas/. Accessed 26 Sep 2019

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

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

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

21. Kronenfeld BJ, Wong DWS (2017) Visualizing Statistical Significance of Disease Clusters Using Cartograms. International Journal of Health Geographics 16:19

22. Dorling D (2011) Area Cartograms: Their Use and Creation. In: Concepts and techniques in modern geography (catmog). pp 252–260

23. Dougenik JA, Chrisman NR, Niemeyer DR (1985) An Algorithm to Construct Continuous Area Cartograms. The Professional Geographer 37:75–81

24. Griffin T (1980) Cartographic Transformation of the Thematic Map Base. Cartography 11:163–174

25. Berry BJL, Morrill RL, Tobler WR (1964) Geographic Ordering of Information: New Opportunities. The Professional Geographer 16:39–44

26. Olson JM (1976) Noncontiguous Area Cartograms. The Professional Geographer 28:371–380

27. Levison ME, Haddon Jr W (1965) The Area Adjusted Map. An Epidemiologic Device. Public Health Reports 80:55–59

28. Monmonier M (2018) How to Lie with Maps (Third Edition). https://doi.org/10.1191/0309132505ph540pr

29. Kocmoud C, House D (1998) A Constraint-based Approach to Constructing Continuous Cartograms. In: Proc. Symp. Spatial data handling. pp 236–246

30. Dent BD (1972) A Note on the Importance of Shape in Cartogram Communication. Journal of Geography 71:393–401

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

32. Nusrat S, Kobourov SG (2016) The State of the Art in Cartograms. Computer Graphics Forum 35:619–642

33. Min Ouyang, Revesz P (2000) Algorithms for Cartogram Animation. In: Proceedings 2000 International Database Engineering and Applications Symposium (Cat. No.PR00789). pp 231–235

34. Keim D, North S, Panse C, Schneidewind J (2002) Efficient Cartogram Generation: A Comparison. In: IEEE Symposium on Information Visualization, 2002. INFOVIS 2002. IEEE, pp 33–36

35. Tobler W (2004) Thirty Five Years of Computer Cartograms. Annals of the Association of American Geographers 94:58–73

36. Monmonier M (2005) Cartography: Distortions, World-views and Creative Solutions. Progress in Human Geography 29:217–224

37. Kreveld M van, Speckmann B (2007) On rectangular cartograms. Computational Geometry 37:175–187

38. Cano RG, Buchin K, Castermans T, Pieterse A, Sonke W, Speckmann B (2015) Mosaic Drawings and Cartograms. In: Computer graphics forum. Wiley Online Library, pp 361–370

39. Hafen R (2018) Introducing geofacet. ryanhafen

40. W. PL, Carr DB, Pearson JB (2015) micromapST: Exploring and Communicating Geospatial Patterns in US State Data. Journal of Statistical Software 63:1–25

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

42. McGranaghan M (1993) A Cartographic View of Spatial Data Quality. Cartographica: The International Journal for Geographic Information and Geovisualization 30:8–19

43. Cliburn DC, Feddema JJ, Miller JR, Slocum TA (2002) Design and Evaluation of a Decision Support System in a Water Balance Application. Computers & Graphics 26:931–949

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

45. Goodchild M, Buttenfield B, Wood J (1994) On Introduction to Visualizing Data Validity. Visualization in geographical information systems 141–149

46. MacEachren AM (1992) Visualizing Uncertain Information. Cartographic Perspectives 10–19

47. Van der Wel FJ, Hootsmans RM, Ormeling F (1994) Visualization of Data Quality. In: Modern cartography series. Elsevier, pp 313–331

48. Dang G, North C, Shneiderman B (2001) Dynamic Queries and Brushing on Choropleth Maps. In: Proceedings Fifth International Conference on Information Visualisation. pp 757–764

49. Carr DB, Wallin JF, Carr DA (2000) Two New Templates for Epidemiology Applications: Linked Micromap Plots and Conditioned Choropleth Maps. Statistics in Medicine 19:2521–2538

50. Pedersen TL, Robinson D (2019) gganimate: A Grammar of Animated Graphics.

51. Pedersen TL (2018) The Grammar of Animation. https://youtu.be/21ZWDrTukEs. Accessed 16 Nov 2018

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

53. R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

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

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

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

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

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

59. Pebesma E (2018) Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10:439–446

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

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

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

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

64. Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R (2019) rmarkdown: Dynamic Documents for R.