



Mapping cancer: the potential of cartograms and alternative map displays

Stephanie Kobakian¹, Dianne Cook¹, Jessie Roberts²

¹Department of Econometrics and Business Statistics, Monash University, Victoria, Australia; ²Science and Engineering Faculty, Queensland University of Technology, Queensland, Australia

Contributions: (I) Conception and design: All authors; (II) Administrative support: None; (III) Provision of study materials or patients: J Roberts, S Kobakian; (IV) Collection and assembly of data: None; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Stephanie Kobakian, MPhil. Department of Econometrics and Business Statistics, Monash University, Melbourne VIC 3800, Australia. Email: stephanie.kobakian@monash.edu.

Abstract: Cancer atlases communicate cancer statistics over geographic domains, typically with a choropleth map. A choropleth map subdivides these domains into administrative regions such as countries, states, or suburbs. When communicating human-related statistics, the choropleth has a disadvantage in that it draws attention to sparsely populated rural areas to the neglect of small inner-city areas. The smaller geographic areas are important to consider if they are densely populated. Alternative map displays, such as a cartogram or a hexagon tile map, can shift the attention of map users from the large rural areas by decreasing their size on the map display. This means alternative displays can be more effective at accurately communicating spatial patterns across spatial areas. This study summarizes current practices for cancer atlases and investigates the alternative map displays that could be used to accurately represent the distribution of cancer statistics across a population, as many cancer atlases lack appropriate displays for population statistics. It is recommended that alternative displays are included in cancer atlases for a perceptually accurate display of the distribution of the burden of cancer over the population, in addition to the familiar choropleth map, if possible given time and budget constraints. With the ease of today's technology, user interaction with the displays is also encouraged. Users should also be able to interactively display different statistics, such as incidence rate or relative incidence, or filter by demographic variables.

Keywords: Cancer; cancer atlas; choropleth; cartogram; spatial

Received: 20 November 2019; Accepted: 09 June 2020; Published: 20 November 2020.

doi: 10.21037/ace-19-31

View this article at: <http://dx.doi.org/10.21037/ace-19-31>

Introduction

Researchers, health authorities, governments, not-for-profits and the media are common communicators of cancer statistics. They often present statistics to the public as aggregated values for geopolitical areas. Counts and incidence rates are often obtained by state health registries. Data privacy and ethics are key concerns even when combined to small area statistics. Presenting these statistics requires aggregating individual observations for the geographical units, also for political and policy purposes.

Examples of typical geographical units include states, provinces, local government areas, and post/zip codes. This type of data is routinely collected for public health reasons and may be made available to the general public as a service to the community.

A cancer atlas is a map, or collection of maps, commonly representing cancer incidence or mortality patterns across a country, or group of countries. Atlases are key to developing hypotheses regarding areas with unusually high rates, and geographic correlations (1). The data collection methods

Table 1 Common measures for reporting cancer information (*Figure 1*)

Measure	Details	Figure
1. Cancer incidence counts	The number of new cases occurring in a specified period and geographic area (4)	–
2. Incidence rate (per 100,000)	The frequency of cases relative to a fixed population size, usually 100,000 people, within a specific period of time (4)	<i>Figure 1E,G</i>
3. Incidence ratio	The incidence rate in a region divided by the average incidence rate for all of the regions	<i>Figure 1C</i>
4. Age-standardized incidence rate (per 100,000)	The weighted mean of the age-specific incidence rates; the weights are taken from population distribution of the standard population. The most frequently used standard population is the World Standard Population (4)	<i>Figure 1B</i>
4. Age-SIR	The ratio of the observed Age-Standardized Incidence Rate, to the expected number of cases. Having adjusted for differences in population and differences in age structure of the population within each area (5)	<i>Figure 1F</i>
5. Age-standardized relative risk	The relative risks represent the risk of an area (ward) relative to average risk, adjusted for age, deprivation and chance fluctuations due to small numbers (6)	<i>Figure 1A,D</i>

Each measure is given along with a description and the Atlases in *Figure 1* that display each measure. SIR, standardized incidence ratio.

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, with more modern atlases directly evolving from early work in US cancer atlases, appearing in 1971 (2). The presentation of cancer statistics has increased with greater access to computational power and the availability of geographic information systems software (3).

This paper considers the current visualization techniques to communicate statistics to the public and their applications to cancer statistics. Alternative approaches are posed because they may be more effective than contemporary techniques. The differences and historic use of these displays is discussed, highlighting the potential and the limitations of the visualization methods.

The paper is structured as follows. The next section “measures mapped” describes the common statistics displayed in disease mapping. Section “visualization approaches” focuses on disease map visualizations. It describes cancer atlases, presents examples of atlases in use today. It discusses the limitations of the most commonly used technique for disease mapping, the choropleth map. This section also describes alternative displays, including the cartogram, which are useful when the map has heterogeneously sized geographic units. Section “comparison and critique of alternative displays” presents the limitations in the production and use of alternative displays. Disease maps are more useful when made interactive, and common options are described in “user interaction”, along with a discussion of benefits and

disadvantages. The “conclusions” section summarizes this survey of the literature, and provides recommendations.

Measures mapped

Epidemiologists and statisticians have developed a range of statistics to communicate the burden of cancer, and the choice of statistics used in maps has changed in recent decades. *Table 1* summarizes the measures commonly presented in published cancer atlases. Mortality rates are commonly presented as relative rates of risk across the population, and age-adjusted to correct for the higher prevalence of cancers in older populations. As described by Howe (7), the Englishman P. Stocks advanced the field of mortality statistics by introducing the standardized mortality ratios in the 1930s, which was an improvement on crude death rates.

The measures displayed are typically aggregations of data values over small areas or model estimates. This is to protect patient privacy, and for numerical stability. The counts of cases in small areas can be difficult to obtain as they are often protected for privacy reasons. The information released and the statistics presented in worldwide cancer atlases are often model estimates. The measures described in *Table 1* use incidence as the statistic of interest, which reports the number of new cases. These same statistics can be used to describe the mortality, that is, standardized incidence rate is analogous to the standardized mortality rate.

Obtaining reliable estimates becomes more difficult as

data becomes more sparse, either by increasing geographic resolution or stratifying by age/sex. Many atlases analyzing smaller areas such as the Australian Cancer Atlas (8), use statistical modelling to produce estimates, although some continue to use simple calculations and suppress regions where estimates are unstable.

Visualization approaches

Traditional approaches for cancer map displays

A choropleth map is the common display to visualize and communicate geospatial cancer statistics over geographic domain. Choropleth maps are a type of thematic map that show polygons for each of the groups of data points representing the geographic units, where each polygon is shaded with a color according to the area-specific values of the statistic being conveyed. Visualizing this data is helpful as geographic patterns of disease may be obscured when reported in a table (9). Providing a visual representation of cancer outcomes allows identification of geographic patterns of the disease that can then be addressed with public health policy and actions. The spatial distribution of the disease incidence can be examined using a choropleth map display 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. One of the key challenges with mapping spatial patterns of disease is the design of visualizations (3). It is important to consider the strengths and weaknesses of designs, as visualizing diseases on maps is often the first step in exploratory spatial data analysis and helps in the formulation of hypotheses.

A choropleth map displays the geographical distribution of data over a set of spatial units by shading areas of a map (10,11). Faithful rendering of the geography, when combined with an appropriate color scheme, can 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 (9). Early versions of choropleth maps used symbols or patterns instead of color. Choropleth maps can be used for displaying disease data (12), including cancer data (13). In epidemiology, choropleth maps are often used as a tool to study the spatial distribution of cancer incidence and mortality.

Displaying familiar state boundaries can make a map easier to read (14) and allow viewers to infer the spatial relationships visually in the data using their mental model

of the geography. The users of disease maps may include researchers, the public, policy makers, and the media (13). For these users, the familiarity of the geography is a worthy consideration when presenting results of spatial analysis.

Cancer maps are effective tools for communicating incidence, survival, and mortality to a wide range of audiences, including the 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 (12), discusses using visuals as a '*formal statistical assessment of the spatial pattern*'. Overwhelmingly, choropleth maps are visualizations chosen to communicate cancer statistics to members of the public and other non-expert audiences.

A review of modern cancer atlases (15) identified 33 cancer atlases published between 2010 and 2016. Atlases published between 2016 and 2018 have also been considered. Each of these online atlases uses choropleth maps. All except 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, the selection varies in the geographies explored. *Figure 1B* shows Globocan 2018 (4) which explores estimated cancer incidence, mortality and prevalence worldwide using estimates based on available country cancer registries. Most atlases allow users to view sex-specific distributions.

There is large variation in the resolution of the maps. *Figure 1B* shows global information at a national level. The United States Cancer Statistics (18) shows data aggregated at the state level (n=51). The Environment and Health Atlas of England and Wales (6) (*Figure 1A*) shows the relative risk for women developing lung cancer at a neighborhood (small-area) scale (n=8,850). The Australian Cancer Atlas (*Figure 1F*) shows the relative incidence ratio of lung cancer in males for each Statistical Area at Level 2 (19) (n=2,292). The Atlas of Cancer in Queensland (*Figure 1C*) shows a subset of the Australian Statistical Local Areas (SLAs) located in the state of Queensland within Australia (16) (n=478).

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 (15).

Australia presents an extreme case of an urban rural

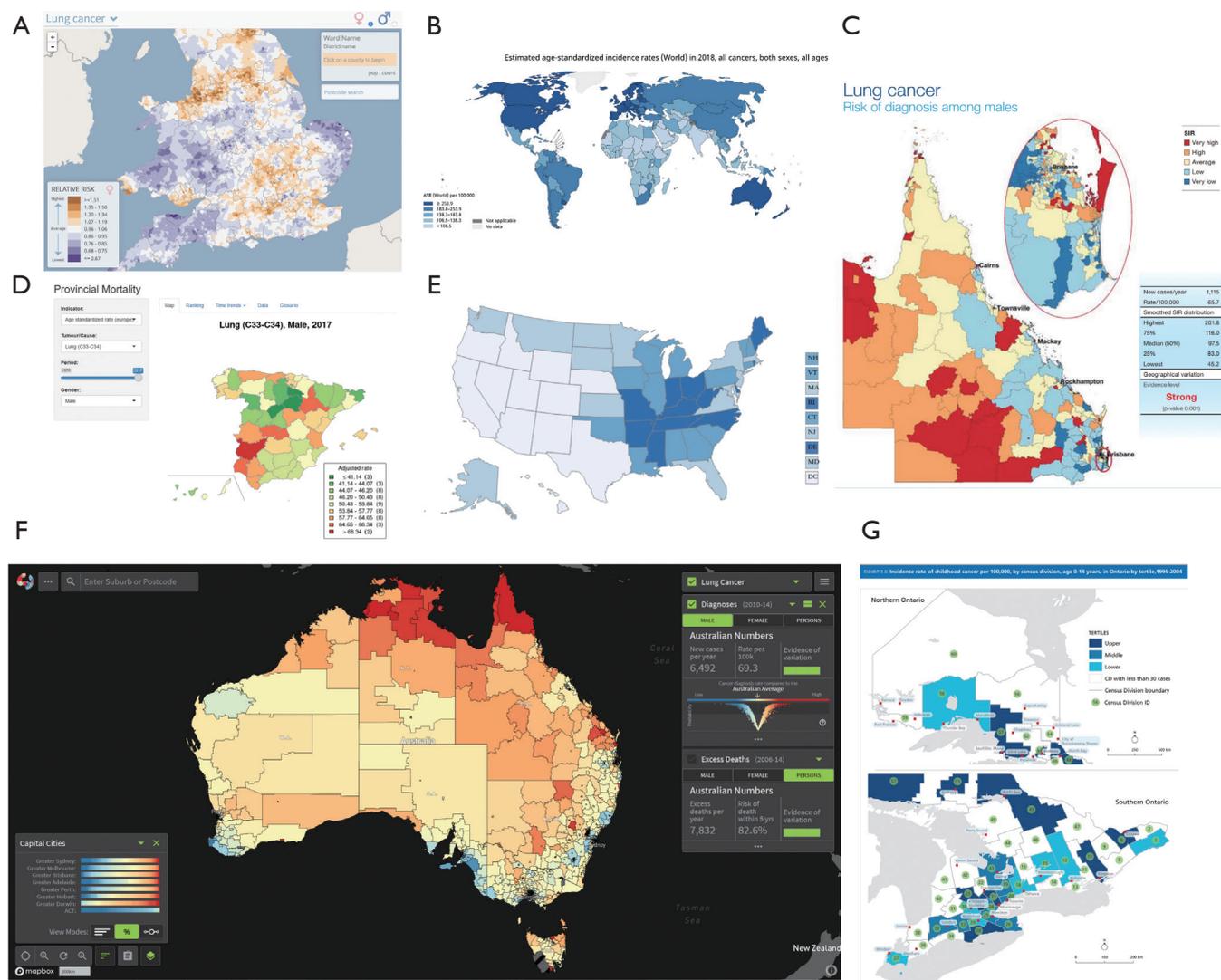


Figure 1 A selection of choropleth cancer maps from online atlases that are publicly available. These include: the Environment and Health Atlas of England and Wales (A), Globocan 2018: Estimated Cancer Incidence, Mortality and Prevalence Worldwide (B), Atlas of Cancer in Queensland (C), Map of Cancer Mortality Rates in Spain (D), United States Cancer Statistics: An Interactive Cancer Statistics Website (E), The Australian Cancer Atlas (F), and the Atlas of Childhood Cancer in Ontario (G). These atlases are described in *Table 2*.

divide. The land mass occupied by urban electoral districts is only 10% of Australia, yet 90% of the population live in these urban areas (20). Choropleth maps provide a familiar display, that shows data in a geographically recognizable way. A disadvantage is that the different population and geographical sizes of administrative areas can attract attention to the shades of the underpopulated but large areas (10). Skowronnek (11) also discusses how choropleth maps suffer from area-size bias, as they give a ‘stronger visual weight’ to large administrative units. The administrative boundaries

used to define regions may limit a choropleth display, as this display unfaithfully represents the disease distribution across the region by obscuring small geographic areas. Sparsely populated rural areas are emphasized, whereas the areas representing inner city communities are very small. This is especially true for Australia.

Choropleth maps color each geographic unit to allow map users to measure the value of the statistic (10). Map users contrast the colors in neighboring areas to understand the spatial distribution. Pickle’s (21) suggestions for

Table 2 A selection of choropleth cancer maps from online atlases displayed in *Figure 1*

Figure	Atlas	Data	Statistic	Source
<i>Figure 1A</i>	The Environment and Health Atlas of England and Wales (6)	Sex: women; year: 2010; cancer: lung	Relative risk	Office for National Statistics (ONS) (England) and from the Welsh Cancer Intelligence and Surveillance Unit (WCISU)
<i>Figure 1B</i>	Globocan 2018: Estimated Cancer Incidence, Mortality and Prevalence Worldwide (4)	Sex: both; year: 2018; cancer: all invasive	Age standardized incidence rates (per 100,000)	World Health Organization's International Agency for Research on Cancer
<i>Figure 1C</i>	Atlas of Cancer in Queensland (16)	Sex: males; year: 1998–2007; cancer: lung	Relative incidence ratio	Queensland Cancer Registry
<i>Figure 1D</i>	Map of Cancer Mortality Rates in Spain (17)	Sex: males; year: 1975–2017; cancer: lung	Age Standardized Rate	Map of cancer mortality rates in Spain
<i>Figure 1E</i>	United States Cancer Statistics: An Interactive Cancer Statistics Website (18)	Sex: both; year: 2016; cancer: all types	Incidence rate per 100,000	With data from state cancer registries: Centers for Disease Control and Prevention
<i>Figure 1F</i>	Australian Cancer Atlas (8)	Sex: males; year: 2010–2014; cancer: lung	Average standardized incidence rate	Australian Cancer Atlas
<i>Figure 1G</i>	Atlas of Childhood Cancer in Ontario (15)	Sex: both; year: 1995–2004; cancer: all types	Incidence rate per 100,00	The Pediatric Oncology Group of Ontario Networked Information System

choropleth map displays include directions to categorize the statistic presented on the map according to percentiles. A color scheme that communicates high and low rates can be useful in displays of cancer statistics, using a double ended color scheme such as those provided by the ColorBrewer system (22) and viridis (23) palettes provide effective color schemes for qualitative, sequential and diverging data. When communicating information using color, a map creator should use a scheme with perceptually uniform color spaces that match equal steps in data space with equal steps in the color space (24). It is possible to allow for data to progress uniformly in both positive and negative directions from a mid-point, such as the mean of the data. These diverging color schemes pair two sequential schemes that use a common light color at the mid-point, each sequential scheme progresses to a dark hue at the extreme value in each direction (25), it preferable that the scheme use darker or warmer colors for higher cancer rates (21). A linear color gradient is appropriate for incidence counts and rates. The linear gradient can also be transformed using the log scale, to show appropriate colors for ratio measures. The Australian Cancer Atlas example in *Figure 1F* implements dark red for areas with a standardized incidence ratio (SIR) value with a risk level more than 50% (e.g., SIR = 1.5) above

the Australian average. Areas were colored dark blue if they had an SIR value below the inverse of the risk value specified (e.g., ~0.67). The use of borders and backgrounds, and their colors, can also change the appearance of the colors representing the value of the statistics (22). These supports can be used to implement a reference point in the color scheme as well as orient users to the geographic regions.

Contemporary alternatives to choropleth maps

Cartograms

Choropleth maps imply uniformity of data across the geographic space but population densities are unlikely to be uniform (11). Cartographers developed the cartogram to draw the attention to the population by transforming the map (26). 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.

Cartograms provide an alternative visualization method for statistical and geographical information. Monmonier (27) suggests that map creators can use white lies to create useful spatial displays. An area cartogram (28), or population-

by-area cartogram (29) is produced from the distortion of the geographical shape according to population. Event cartograms (30) change the area of regions on a map depending on the amount of disease-related events, rather than population. It is easy for the reader to disregard the impact of transformations used to create cartograms, for the benefit of reading the statistical distribution more accurately with approximate geographic information. The spatial transformation of map regions relative to the data emphasizes the data distribution instead of land size (31). When visualizing population statistics, Dorling considers this design ‘more socially just’ (20), or honest (32), giving equitable representation and attention to all members of the population and reducing the visual impact of large areas with small populations (12). Howe (7) 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 (33).

The creation of cartograms was historically in the hands of professional cartographers (34). Early approaches include Hunter and Young’s (35) wooden tile methods, Skoda and Robertson’s (36) steel ball-bearing approach and Tobler’s (37) computer programs. Howe (7) 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’ (34).

The intended audience and the communication purpose are important to consider when creating alternative map displays. Nusrat and Kobourov (38) provided a framework to investigate implementations of the many algorithms presented, and the “*statistical accuracy, geographical accuracy, and topological accuracy*”. The alternative map displays in *Figure 2A,B,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 prevent overlooking the impact of the disease within physically small but population-dense states. Map creators give each state equal size and thus equal emphasis in *Figure 2D* the hexagon tile map.

Figure 2 shows four different cartograms for the average age-adjusted rate of incidence for lung and bronchus for females and males in the United States 2012–2016. Each cartogram highlights different aspects of the population and

relation to the average age-adjusted rate of incidence.

In the contiguous cartogram map (*Figure 2A*) 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.

In the non-contiguous (*Figure 2B*) 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.

In the Dorling cartogram (*Figure 2C*) the North-East states remain closer to their neighbors and are slightly 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.

In the Hexagon Tile Map (*Figure 2D*) it is simple 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.

Contiguous

A contiguous cartogram alters the choropleth according to a statistic and maintains connectivity of the map regions. Ouyang and Revesz (8) 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 (20), Radial Expansion Method, Rubber Sheet Method, Line Integral Method, Constraint-Based Method (31).

Figure 2A shows a population contiguous cartogram of the United States. All states are visible and the shape of the United States overall is still recognizable.

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 (28). The shapes of small areas on a choropleth map and a cartogram are preserved using Tobler’s Conformal mapping method. Kocmoud and House (31) present this issue as conflicting tasks or aims, to adjust region sizes and retain region shapes.

Non-contiguous

Non-contiguous cartograms prioritize the shapes of the

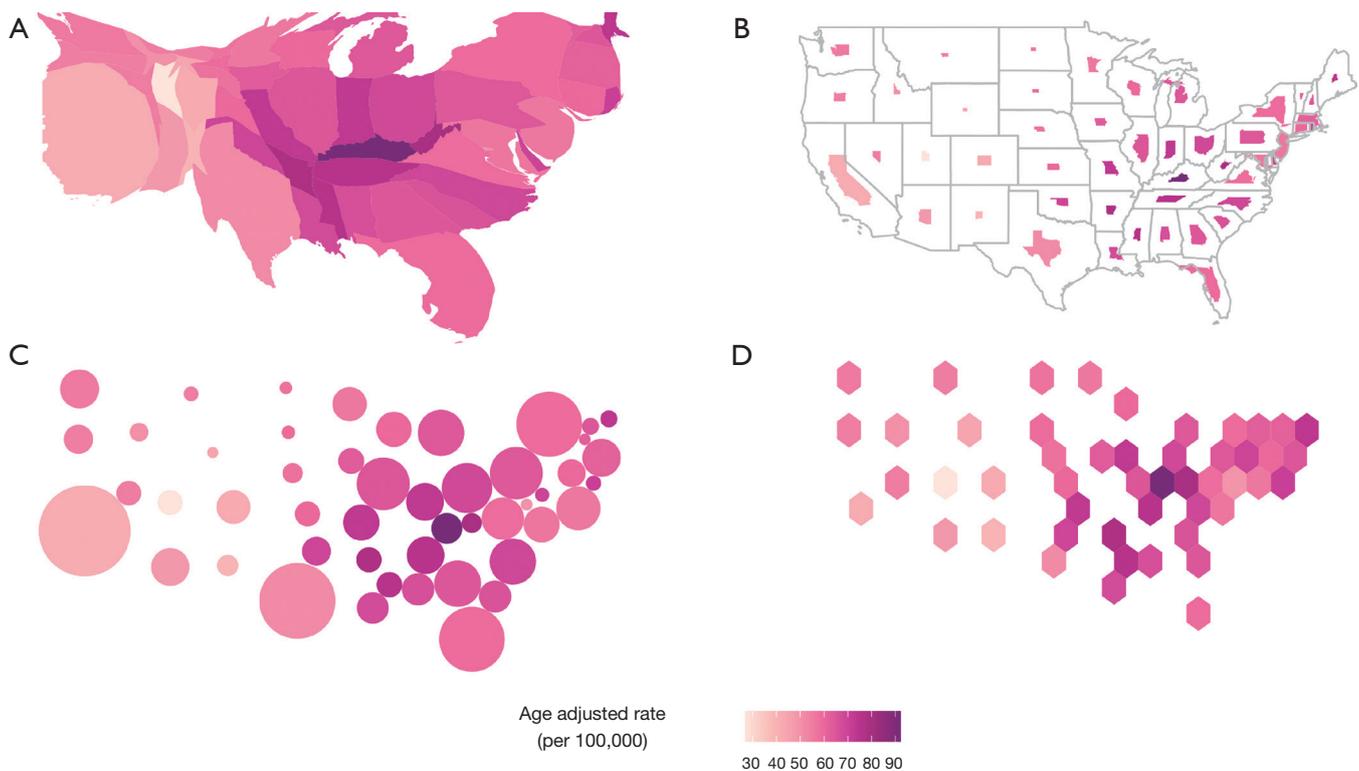


Figure 2 Common alternatives to maps, showing the same information for the United States of America. The color of each state communicates the average age-adjusted rate of incidence for lung and bronchus for females and males in the United States 2012–2016: (A) contiguous cartogram distorted each state’s shape by the population of the state in 2015, (B) non-contiguous cartogram preserves the shape of the cartogram, but the size now reflects the population of the state in 2015, (C) Dorling cartogram are non-contiguous, circles are used to represent each state and the population of the state determines the size in 2015, (D) hexagon tile map (non-contiguous) uses a hexagon of equal size for each state, and colors the state by the average age-adjusted rate of incidence.

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 (28) discusses the creation of these maps and the significance of the empty areas left between the geographic boundaries and the new shape. The white space presents the meaningful empty-space property (39,40).

Dorling

Daniel Dorling presents an alternative display engineered to highlight the spatial distribution and neighborhood relationships without complex distortions of borders and boundaries (20): “*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 answered this with his implementation of maps created using ‘the simplest of all shapes’. Circular cartograms use the same circle shape for every region represented, resized according to the statistic represented or the population. This simple shape may be more effective for understanding the spatial distribution than contiguous cartograms. Contiguous cartograms create ‘nonsense’ shapes that have ‘no meaning’ (32). Both methods apply a gravity model to produce a layout, that avoids overlaps and keep spatial relationships with neighboring areas over many iterations. The circular cartogram is relatively fast to compute.

Raisz (41) laid the groundwork for this approach in the

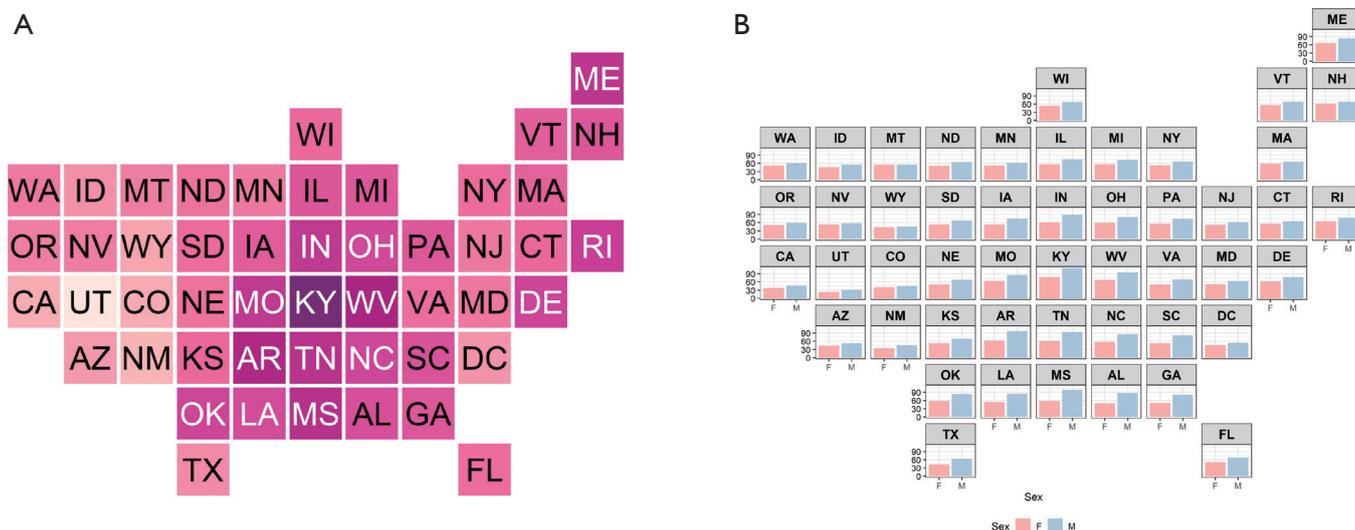


Figure 3 Two alternative displays, tile map (A) and geofaceted map (B), showing state age-adjusted rate of incidence for lung and bronchus in the USA. In the tile map, the layout approximates spatial location, with each state being an equal box filled with color representing cancer incidence. The geo-faceted 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.

mid-1930s, drawing rectangular cartograms that provide simple comparisons, effective for correcting misconceptions communicated by geographic maps. Tobler (42) names and defines these as Value-Area Cartograms. This rectangular display may sacrifice contiguity but allows for tiling where geographic neighbors placed in suitable relative positions also share borders (43). Rectangular cartograms communicate bivariate displays of the population by the size of each rectangle, and they use color to communicate a second variable (44).

Tile map

A tile map provides a tessellated display of consistent shapes. A similar method to a rectangular cartogram, representing each geographic area using a square. The squares are tessellated to create a grid as shown in *Figure 3A*. Each area is represented by a square of the same dimensions, 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 same dimensions. There are online media sources that use this

method (45-48). 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 (49) 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 tiles per region, 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.

Geofacet

Hafen (50) introduces the term geofacet to describe a grid display of small plots. The arrangement of tiles in *Figure 3B* mimics the geographic topology. Geofaceting has the functionality that a statistical plot can be constructed in each facet for each geographic area. A tile map can communicate only one value per region in a visualization, while geofaceting is a more flexible visualization for communication as it increases the amount of information displayed. Virtually any type of plot can be shown in the tile, allowing displays of multiple variables or values per geographic entity. Creating the layout of a geofacet is manual, but once created can be used for any data on that geographic base.

Multivariate displays

Pickle and others (51) present linked micromap plots to match geographic and statistical data visually, this serves as a solution to multi-dimensionality issues. 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 micromaps 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.

Lucchesi and Wikle (52) present bivariate choropleth maps blend color schemes to convey the intersection of categorized levels of an estimate and the associated uncertainty for each spatial area. They also suggest map pixilation, which 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.

Comparison and critique of alternative displays

Performance of cartograms for Australia

Figure 4 shows four main types of cartograms using melanoma incidence on Australian Statistical Areas at Level 3 (53). The version of a contiguous cartogram (Figure 4A) has expanded the highly populated areas while preserving the full shapes of rural areas. The South-East is enlarged, but high population areas are still small, and low population areas are still large on the map. It has not fully resolved the population transformation of areas, because the algorithm can't reach an optimal configuration where area matches population—Australia is too heterogeneous. 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 shapes. The Dorling cartogram and the hexagon tile map provide

reasonable displays of the spatial distribution, despite having a very large amount of white space in the outback areas.

Limitations of alternative displays

Cartograms provide the spatial distortion to more accurately convey the statistical distribution, focusing on the human impact of the disease. However, the transformation of contiguous cartograms often occurs at the expense of the shape of areas (31). When the population density of the geographic units is highly dissonant with geographic density, the cartogram will lose all spatial context. Dorling (20) contains a cartogram showing the 1966 general election results for Australia, produced by Hughes and Savage (54), which looked very little like the geographical shape of Australia. The reader is encouraged to access the freely available pdf of Dorling's book, and this image can be found on page 41.

The most common aesthetics employed in alternative map displays are shape, color and size. Each alternative display allows for some combination of these. Color is used most often to represent the variable of interest. The size and shape are often used to scale the areas to draw attention to regions of interest.

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. Bell *et al.* (13) provide suggestions and comments to help map creators best communicate their health data and spatial analysis. The authors highlighted that the map design chosen should be tested on a representative sample of potential consumers, to ensure that the target audience is not misinformed by the display. The authors encourage the consideration of map types beyond the traditional classed choropleth map, but warn that sound cartographic principles must be employed to ensure effective communication to the public. A clear definition of the purpose of the display can help map creators to select the design that best communicates the statistic of interest (13). Table 3 lists several features, or limitations, of each alternative display in contrast to the commonly used choropleth map. The desirable features of each display can be contrasted within the table, this can be used to help inform the choice of map creators as they consider each alternative display. Map creators should choose a display that best communicates the statistic according to the purpose of their display.

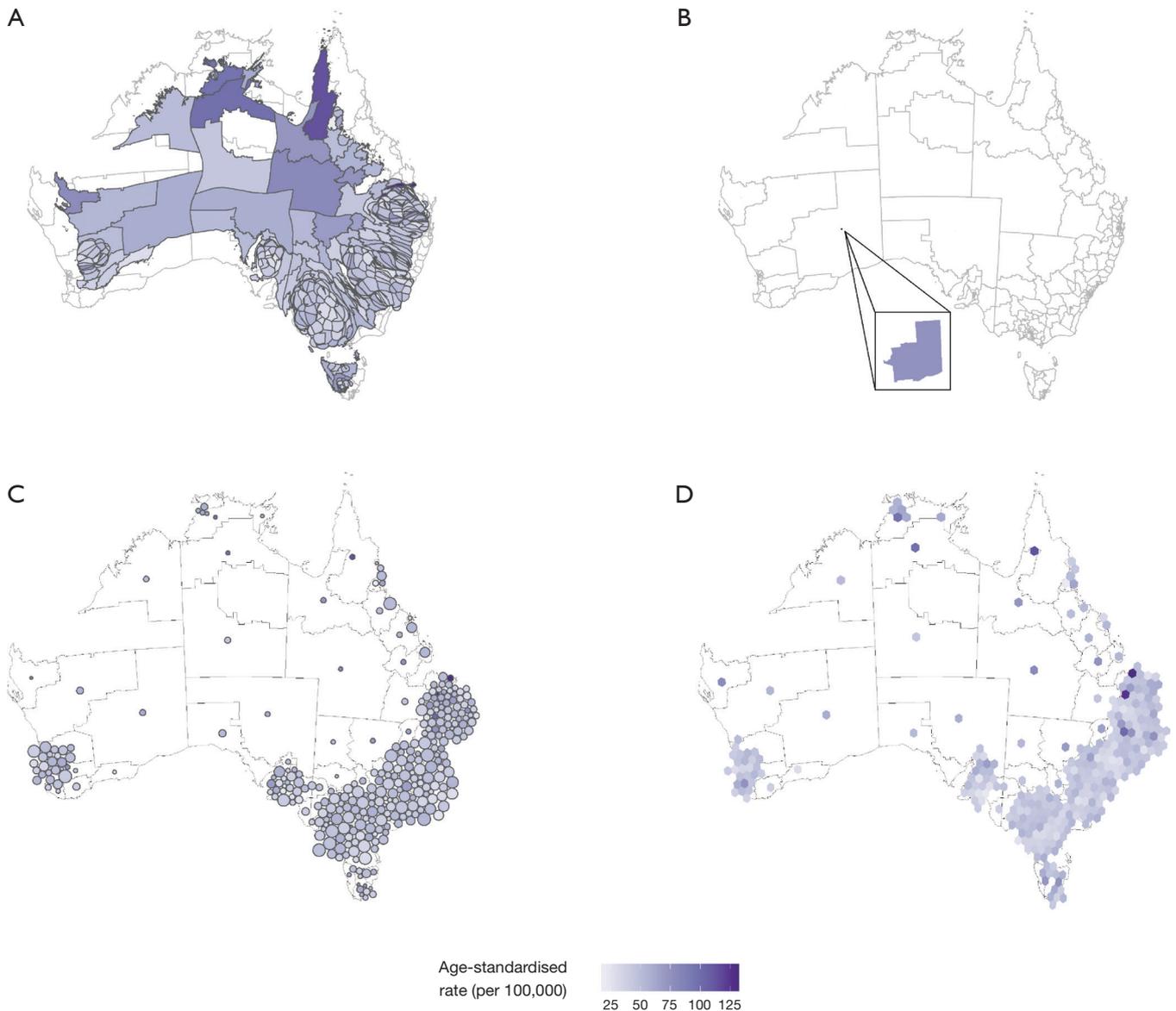


Figure 4 Cartograms showing melanoma incidence in Australia: (A) contiguous, partially population transformed, (B) non-contiguous shape preserved, (C) Dorling, (D) hexagon tile map. The contiguous cartogram has expanded the highly populated areas while preserving the full shapes of rural areas. If it 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.

Additional considerations

Cancer atlases often display supplementary graphs and plots to add more 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 (13). The many displays of statistical summaries, including dot plots, bar plots, box plots, cumulative distribution plots, scatter plots, and 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

Table 3 Summary of features and constraints of common mapping methods used to display cancer statistics

Features	Choropleth	Contiguous	Non-contig	Dorling	Tile maps	Geofacets
Spatial distortion	No	Yes	Yes	Yes	Yes	Yes
Preserves neighbors	Yes	Yes	Yes	Possible	Possible	Possible
Conceals small areas	Yes	Possible	No	No	No	No
Uniform shapes	No	No	No	Yes	Yes	Yes
Univariate data	Yes	Yes	Yes	Possible	Possible	No
Manual construction	No	No	No	No	Yes	Yes

“Yes” indicates the characteristic is always used, “No” indicates never used, and “Possible” it may be employed, but it is not necessary.

sizes, and standard deviation.

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 (55). World atlases can allow for displays of data aggregated into continents, countries, states, provinces and congressional districts (18). Each population area will probably have a different number of people, which is typically used to calibrate the statistic. Cancer atlases may also communicate the distribution of the population living in all areas in a table or histogram display (56). Atlases can connect the population to the land available to them by communicating population density.

Maps can also be used to focus on demographic strata, such as age and sex. Some of the digital atlases surveyed allow subsets to be selected for display, for example the Australian Cancer Atlas (8) allows filtering for the selection of statistics regarding males and females specifically.

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, a cancer atlas often communicates uncertainty about the value of a statistic. There are several potential sources of uncertainty: sampling error, errors arising from the disease reporting or data collection processes, and uncertainty arising from the statistical modelling or simulation process. The most common measures used to present uncertainty are credible intervals or confidence intervals. Displaying the uncertainty associated with reported statistics is a vital feature of a cancer map, but it is difficult to display effectively, the Australian Cancer Atlas (8) uses transparency to communicate uncertainty. Providing an adjacent map

or overlaying maps with symbols (30) are two common solutions.

User interaction

One of the concerns of adding too much information to a map is the fear of cognitive overload (57) in which the user reaches an information threshold, beyond which they become confused. It can be a juggling act for a diverse audience, with experts probably preferring more detail (58) while a simpler display is more broadly readable. 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 (59).

Monmonier (27) recommends using interactivity to allow users to explore the map for more information and provide flexibility for the display. The user can toggle between different variables, map views or even multiple realizations of future scenarios (60). This provides additional mechanisms for the users to digest the uncertainty of the available information (61,62). 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 (63). 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 (64).

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,

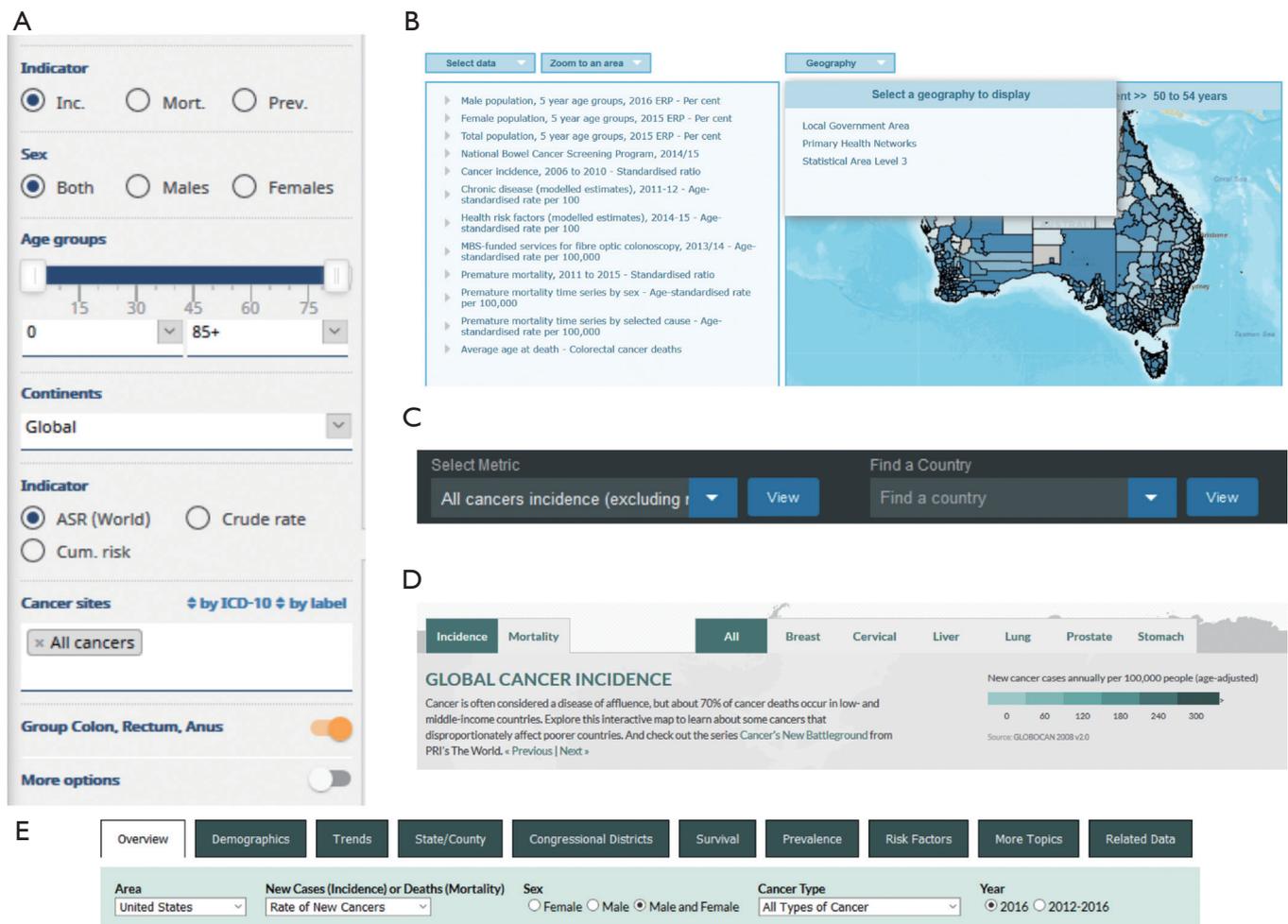


Figure 5 Interactive controls of displays in publicly available choropleth cancer maps: (A) GUI controls for statistic, sex, age groups, continents, and cancer types for Globocan 2018 (4), (B) menus for variable selection and zooming on Bowel Cancer Australia Atlas, (C) menus for choosing variables and countries in The Cancer Atlas, (D) tabs for different indicators and cancer types in Global Cancer Map, (E) menus and toggles for variable and subset selection in United States Cancer Statistics: Data Visualizations.

zooming, and panning to allow users to explore the map as they want more information and allow flexibility in the display (27). The use of these supports can be found in various online cancer maps and are shown in *Figure 5* (39).

Animation, in contrast to interactivity, usually involves pre-computing views and showing these in a sequence. Lin Pedersen (65) provides an overview of animation for maps using the R package *gganimate* (66). 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 (65), and the map image is updated/replaced as the animation progresses.

Weather maps are thoroughly developed examples of animation of spatial displays to communicate information to the general public (13). 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 (67) provides tours that change the display to draw users' attention to areas on the map that are relevant to the interpretation of the statistic displayed. This implementation of animation gives users tools to plan their exploration.

Figure 6 shows two examples of more sophisticated interactive maps. The Spanish Cancer map (*Figure 6A*)

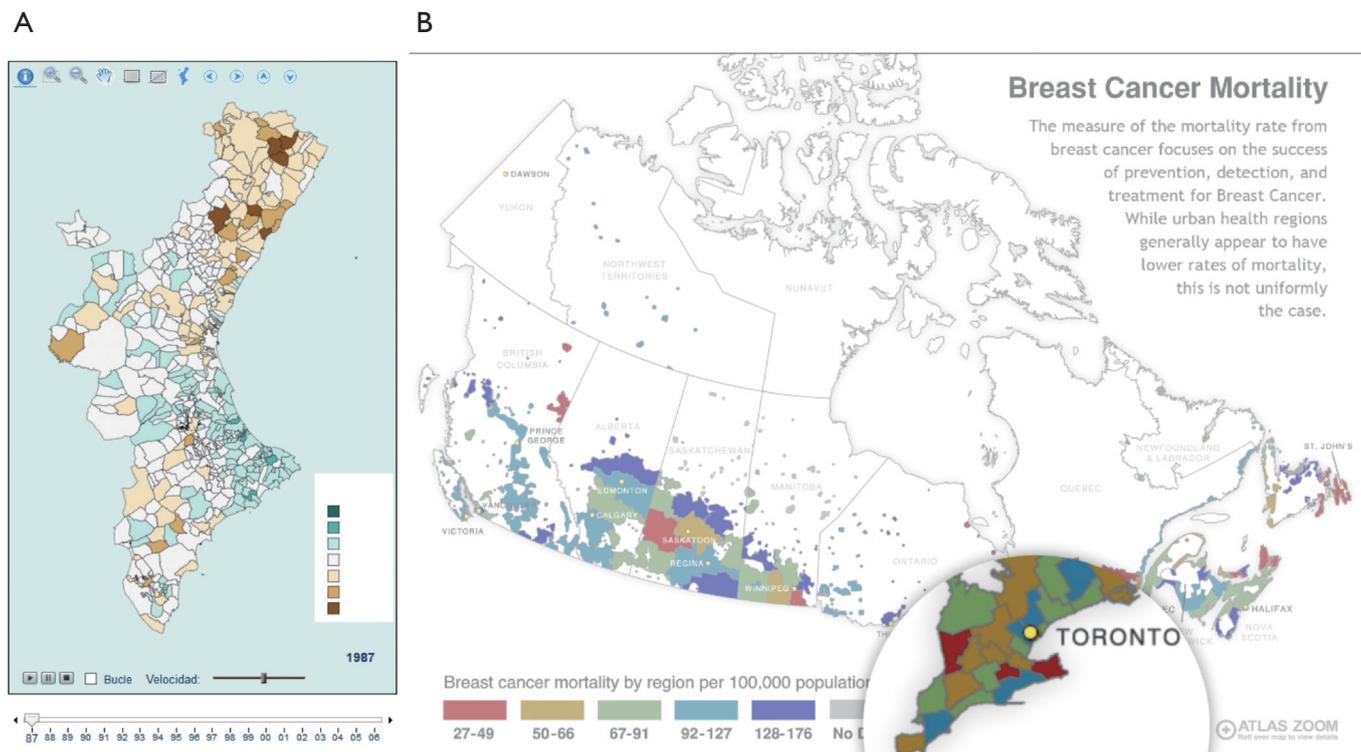


Figure 6 Two examples of advanced interactivity (and animation) in publicly available choropleth cancer maps: (A) linked maps and time-series line plots, with temporal animation in Map of Cancer Mortality Rates in Spain; (B) a highly responsive magnifying glass on a map of Breast Cancer Mortality in Canada.

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 (Figure 6B) 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.

Conclusions

This paper provides an overview of mapping practices as commonly used for cancer atlases. The conventional approach is the choropleth map, and it is widely used. The choropleth map suffers when there are small geographic units, as occurs in Australia. The population of Australia is concentrated in small areas on the coast, and a choropleth map can hide information about the burden of cancer on

those communities. Making an inset can clarify congested regions but this breaks the viewers' attention as they shift focus from the map to the inset, and if there are many congested areas, many insets would be needed. The map alternatives implement trade-offs between the familiar shapes, and the importance of the geographic areas. Given the population or a cancer statistic for each area, the geographic size or shape will change. Alternative displays allow the spatial distribution of cancer data to be digested by map users.

Additional mapping methods should be considered by map creators during the development of a cancer atlas, as alternative displays may align better with the communication purpose. Other considerations need to be taken into account, including audience, budget, time, maintenance. Of primary importance is that information about cancer statistics be effectively communicated to the public. The choropleth has an advantage in that it is more familiar in form for more people, but we have seen that it can give an incorrect perception of the information. Public

atlases can be useful educational tools. A combination of the choropleth alongside an alternative display, could provide a balance of the familiar along with a perceptually accurate display, and provide an opportunity to educate the reader. This is especially recommended for Australia, because of the vast difference between spatial area and population density.

Many statistics are commonly used in cancer displays. It is common to see incidence rates, or ratios which displays how far a region is above or below the average. The purpose of using an age standardized ratio is, perhaps the desire to pinpoint the areas that need attention because they have higher than expected rates. Ratios may negatively impact the interpretability of the actual rates of incidence or mortality. However, this impact is offset by the value of seeing the relativity of the values in reference to the mean, especially when the mean value is also given on the display. This helps to put individual areas in perspective, as a region might have a value higher than the average, but it may not be a health concern if all regions have a low incidence rate. Supplementary materials such as displaying the mean, or a distribution plot, can allow map users to recognise when this occurs.

Interaction with maps is an important component of public atlases, and is becoming increasingly straightforward to add with today's technology. The purpose of interaction in public atlases is to provide access to more information than is possible to display in a single map, without overwhelming the viewer. Too many choices can similarly overwhelm a viewer, and thus decisions do need to be made about content to provide 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.

Software used

The following R (68) packages were used to produce this paper: tidyverse (69), RColorBrewer (70), ggthemes (71), png (72), cowplot (73), sf (74), spData (75), cartogram (76), sugarbag (77), knitr (78), rmarkdown (79) and absmappedata (80). Files to reproduce the paper, and code to reproduce the plots, are available at <https://github.com/srkobakian/review>.

Acknowledgments

The authors would like to thank Dr. Earl Duncan for his contributions in editing and refining the drafts of this

article. They would also like to thank Professor Kerrie Mengersen, Dr. Susanna Cramb and Dr. Peter Baade for conversations on the content of this article.

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the Guest Editors (Peter Baade and Susanna Cramb) for the series "Spatial Patterns in Cancer Epidemiology" published in *Annals of Cancer Epidemiology*. The article has undergone external peer review.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (<http://dx.doi.org/10.21037/ace-19-31>). The series "Spatial Patterns in Cancer Epidemiology" was commissioned by the editorial office without any funding or sponsorship. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. d'Onofrio A, Mazzetta C, Robertson C, et al. Maps and atlases of cancer mortality: a review of a useful tool to trigger new questions. *Ecancelmedicallscience* 2016;10:670.
2. Burbank F. Patterns in cancer mortality in the United States: 1950-1967. *Natl Cancer Inst Monogr* 1971;71:1-594.
3. Exeter DJ. Spatial Epidemiology. In: *International Encyclopedia of Geography: People, the Earth, Environment and Technology*. Hoboken: John Wiley & Sons, 2017:1-4.
4. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer

- statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
5. Duncan EW, Cramb SM, Aitken JF, et al. Development of the Australian Cancer Atlas: spatial modelling, visualisation, and reporting of estimates. *Int J Health Geogr* 2019;18:21.
 6. Imperial College London, Small Area Health Statistics Unit. The environmental and health atlas of England and Wales: National male lung cancer rate. 2010. Available online: <http://www.envhealthatlas.co.uk/eha/Breast/> (Accessed 26 Sep 2019).
 7. Howe GM. Historical evolution of disease mapping in general and specifically of cancer mapping. In: Boyle P, Muir CS, Grundmann E. editors. *Cancer mapping*. Heidelberg: Springer, 1989:1-21.
 8. Ouyang M, Revesz P. Algorithms for cartogram animation. In: *Proceedings 2000 International Database Engineering and Applications Symposium (Cat. No. PR00789)*. Piscataway: IEEE, 2000:231-5.
 9. Moore DA, Carpenter TE. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol Rev* 1999;21:143-61.
 10. Tufte ER, Goeler NH, Benson R. *Envisioning information*. Cheshire: Graphics Press, 1990.
 11. Skowronnek A. Beyond choropleth maps: a review of techniques to visualize quantitative areal geodata. 2016. Available online: https://alsino.io/static/papers/BeyondChoropleths_AlsinoSkowronnek.pdf
 12. Walter SD. *Disease Mapping: A Historical Perspective*. 2001. Available online: <https://dx.doi.org/10.1093/acprof:oso/9780198515326.003.0012>
 13. Bell BS, Hoskins RE, Pickle LW, et al. Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. *Int J Health Geogr* 2006;5:49.
 14. Brewster MB, Subramanian SV. Cartographic Insights into the Burden of Mortality in the United Kingdom: A Review of 'The Grim Reaper's Road Map'. *Int J Epidemiol* 2010;39:1120-2.
 15. Pole JD, Greenberg ML, Sung L, et al. *Survival*. In: Greenberg ML, Barnett H, Williams J. editors. *Atlas of Childhood Cancer in Ontario*. Toronto: Pediatric Oncology Group of Ontario, 2015.
 16. Cramb S, Mengersen K, Baade PD. *Atlas of cancer in Queensland: geographical variations in incidence and survival 1998-2007*. Brisbane: Viertel Centre for Research in Cancer Control, Cancer Council Queensland, 2011.
 17. Carlos III Institute of Health, National Center for Epidemiology. Interactive epidemiological information system (ARIADNA). 2014. Available online: <http://ariadna.cne.isciii.es/evindex.html> (Accessed 27 Sep 2019).
 18. U.S. Cancer Statistics Working Group. *United States Cancer Statistics: Data Visualizations*. 2019. Available online: <http://www.cdc.gov/cancer/dataviz> (Accessed 26 Sep 2019).
 19. ABS. Statistical area level 2 (SA2) ASGS Ed 2011 digital boundaries in ESRI shapefile format. Available online: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.001July%202011>
 20. Dorling D. Area Cartograms: Their Use and Creation. In: Dodge M, Kitchin R, Perkins C. editors. *The map reader*. 2011:252-60.
 21. Pickle LW. A history and critique of U.S. mortality atlases. *Spat Spatiotemporal Epidemiol* 2009;1:3-17.
 22. Harrower M, Brewer CA. ColorBrewer. org: an online tool for selecting colour schemes for maps. *Cartogr J* 2003;40:27-37.
 23. van der Walt S, Smith N. *mpl colormaps*. 2015. Available online: <https://bids.github.io/colormap/>
 24. Madsen R. *Programming Design Systems*. 2019. Available online: <https://programmingdesignsystems.com/>
 25. Brewer C. *Diverging Color Schemes*. 2020. Available online: <http://www.personal.psu.edu/cab38/ColorSch/SchHTMLs/CBCColorDiv.html>
 26. Dougenik JA, Chrisman NR, Niemeyer DR. An algorithm to construct continuous area cartograms. *Prof Geogr* 1985;37:75-81.
 27. Monmonier M. *How to Lie with Maps*. 3rd ed. Chicago: University of Chicago Press, 2018.
 28. Olson JM. Noncontiguous area cartograms. *Prof Geogr* 1976;28:371-80.
 29. Levison ME, Haddon W Jr. The area adjusted map: an epidemiologic device. *Public Health Rep* 1965;80:55-9.
 30. Kronenfeld BJ, Wong DWS. Visualizing statistical significance of disease clusters using cartograms. *Int J Health Geogr* 2017;16:19.
 31. Kocmoud C, House D. A constraint-based approach to constructing continuous cartograms. *Proc Symp Spatial Data Handling* 1998:236-46.
 32. Dent BD. A note on the importance of shape in cartogram communication. *J Geog* 1972;71:393-401.
 33. Griffin TLC. Cartographic transformation of the thematic map base. *Cartography* 1980;11:163-74.
 34. Kraak MJ. Cartographic design. In: *International Encyclopedia of Geography: People, the Earth,*

- Environment and Technology. Hoboken: John Wiley & Sons, 2017:1-16.
35. Hunter JM, Young JC. A technique for the construction of quantitative cartograms by physical accretion models. *Prof Geogr* 1968;20:402-7.
 36. Skoda L, Robertson JC. Isodemographic map of Canada. Information Canada, 1972.
 37. Tobler WR. A continuous transformation useful for districting. *Ann N Y Acad Sci* 1973;219:215-20.
 38. Nusrat S, Kobourov S. The state of the art in cartograms. *Computer Graphics Forum* 2016;35:619-42.
 39. Roberts J. Communication of statistical uncertainty to non-expert audiences. 2019. Available online: <https://doi.org/10.5204/thesis.eprints.130786>
 40. Keim DA, North SC, Panse C, et al. Efficient cartogram generation: a comparison. In: *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002. IEEE, 2002:33-6.*
 41. Raisz E. Rectangular statistical cartograms of the world. *J Geogr* 1936;35:8-10.
 42. Tobler W. Thirty five years of computer cartograms. *Ann Am Assoc Geogr* 2004;94:58-73.
 43. Monmonier M. Cartography: distortions, world-views and creative solutions. *Prog Hum Geogr* 2005;29:217-24.
 44. Van Kreveld M, Speckmann B. On rectangular cartograms. *Comput Geom* 2007;37:175-87.
 45. Montanaro D. NPR Battleground Map: Hillary Clinton Is Winning—And It's Not Close. 2016.
 46. Kanjana J, Mehta D. Who will win the presidency? 2016.
 47. Zitner A, Yeip R, Wolfe J. Draw the 2016 Electoral College Map. 2016.
 48. Gamio L, Cameron D. Poll: Redrawing the electoral map. *The Washington Post*, 2016.
 49. Cano RG, Buchin K, Castermans T, et al. Mosaic drawings and cartograms. *Comput Graph Forum* 2015;34:361-70.
 50. Hafen R. Geofacet: 'Ggplot2' faceting utilities for geographical data. 2019.
 51. Pickle LW, Pearson JB Jr, Carr DB. micromapST: Exploring and communicating geospatial patterns in US State data. *J Stat Softw* 2015;63:1-25.
 52. Lucchesi LR, Wikle CK. Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation. *Stat* 2017;6:292-302.
 53. ABS. Statistical area level 3 (SA3) ASGS Ed 2011 digital boundaries in ESRI shapefile format. Available online: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.001July%202011>
 54. Hughes CA, Savage EE. The 1955 federal redistribution. *Australian Journal of Politics & History* 1967;13:8-20.
 55. Ferlay J, Ervik M, Lam F, et al. Global cancer observatory: cancer today. Lyon: International Agency for Research on Cancer, 2018.
 56. Northern Ireland Cancer Registry. All-Ireland Cancer Atlas 1995-2007. 2011. Available online: <http://www.ncri.ie/publications/cancer-atlases>
 57. McGranaghan M. A cartographic view of spatial data quality. *Cartographica: The International Journal for Geographic Information and Geovisualization* 1993;30:8-19.
 58. Cliburn DC, Feddema JJ, Miller JR, et al. Design and evaluation of a decision support system in a water balance application. *Comput Graph* 2002;26:931-49.
 59. Perin C. Direct manipulation for information visualization. Paris: Université Paris Sud, 2014.
 60. Goodchild M, Buttenfield B, Wood J. On introduction to visualizing data validity. *Visualization in Geographical Information Systems* 1994:141-9.
 61. MacEachren AM. Visualizing uncertain information. *Cartographic Perspectives* 1992;(13):10-9.
 62. Van der Wel FJM, Hootsmans RM, Ormeling F. Visualization of data quality. In: *Modern Cartography Series*. Cambridge: Academic Press, 1994;2:313-31.
 63. Dang G, North C, Shneiderman B. Dynamic Queries and Brushing on Choropleth Maps. In: *Proceedings Fifth International Conference on Information Visualisation*. 2001:757-64.
 64. Carr DB, Wallin JF, Carr DA. Two new templates for epidemiology applications: linked micromap plots and conditioned choropleth maps. *Stat Med* 2000;19:2521-38.
 65. Pedersen TL. The Grammar of Animation. 2018. Available online: <https://youtu.be/21ZWDrTukEs> (Accessed 16 Nov 2018).
 66. Pedersen TL, Robinson D. ganimate: A Grammar of Animated Graphics. 2019.
 67. Cancer Council Queensland, Queensland University of Technology, and Cooperative Research Centre for Spatial Information. Australian Cancer Atlas. 2018. Available online: <https://atlas.cancer.org.au>
 68. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2019.
 69. Wickham H. tidyverse: R packages for data science. 2017. Available online: <https://CRAN.R-project.org/package=tidyverse>
 70. Neuwirth E. RColorBrewer: ColorBrewer palettes. 2014. Available online: <https://CRAN.R-project.org/>

- package=RColorBrewer
71. Arnold JB. ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. 2019. Available online: <https://CRAN.R-project.org/package=ggthemes>
 72. Urbanek S. png: Read and write PNG images. 2013. Available online: <https://CRAN.R-project.org/package=png>
 73. Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. 2019. Available online: <https://CRAN.R-project.org/package=cowplot>
 74. Pebesma E. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 2018;10:439-46.
 75. Bivand R, Nowosad J, Lovelace R. spData: Datasets for Spatial Analysis. 2019. Available online: <https://CRAN.R-project.org/package=spData>
 76. Jeworutzki S. cartogram: Create Cartograms with R. 2018. Available online: <https://CRAN.R-project.org/package=cartogram>
 77. Kobakian S, Cook D. sugarbag: Create Tessellated Hexagon Maps. 2019. Available online: <https://CRAN.R-project.org/package=sugarbag>
 78. Xie Y. knitr: A General-Purpose Package for Dynamic Report Generation in R. 2019.
 79. Allaire J, Xie Y, McPherson J, et al. rmarkdown: Dynamic Documents for R. 2019.
 80. Mackey WF. wfmackey/absmapi: A catalogue of ready-to-use ASGS mapping data. 2019.

doi: 10.21037/ace-19-31

Cite this article as: Kobakian S, Cook D, Roberts J. Mapping cancer: the potential of cartograms and alternative map displays. *Ann Cancer Epidemiol* 2020;4:9.