

Cartogram Mapping and its Application to Cancer Data Visualisation

Stephanie Kobakian, Jessie Roberts and Dianne Cook

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 2 |
| 2 | Map displays for disease data | 3 |
| 3 | Cancer atlases | 4 |
| 3.1 | Overview of publicly available atlases | 5 |
| 3.1.1 | The Environment and Health Atlas of England and Wales | 6 |
| 3.1.2 | Globocan 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide | 6 |
| 3.1.3 | Atlas of Cancer in Queensland | 6 |
| 3.1.4 | Bowel Cancer Australia Atlas | 6 |
| 3.1.5 | United States Cancer Statistics: An Interactive Cancer Statistics Website | 6 |
| 3.1.6 | Map of Cancer Mortality Rates in Spain | 7 |
| 3.1.7 | Atlas of Childhood Cancer in Ontario | 7 |
| 3.2 | Common statistics displayed | 7 |
| 3.2.1 | Geographic hierarchies | 8 |
| 3.2.2 | Population distribution | 8 |
| 3.2.3 | Statistical uncertainty | 8 |
| 3.2.4 | Demographics | 10 |
| 3.2.5 | Socio-economic indicators | 10 |
| 3.3 | Interaction with atlases | 10 |
| 4 | Cartograms and other spatially warped displays | 10 |
| 4.1 | Cartograms | 10 |
| 4.1.1 | Contiguous | 13 |
| 4.1.2 | Non-Contiguous | 15 |
| 4.1.3 | Dorling | 15 |
| 4.2 | Tilegram | 16 |

| | |
|--|-----------|
| 5 Comparison of maps and cartograms | 16 |
| 6 Animation and Interactivity | 17 |
| 7 Acknowledgements | 19 |
| 8 References | 19 |

Abstract

1 Introduction

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

A common approach to communicate cancer statistics, is to display statistics on a map. This is a choropleth map: the statistic is mapped to colour and the geographic region is filled with this colour. The viewer would then be able to examine the spatial distribution of the disease incidence, where there is a trend in longitude or latitude, or rural vs urban, or coastal vs inland, or even specific hotspots of the disease. Visualising diseases on maps is often the first step in exploratory spatial data analysis and effectively helps in the formulation of hypotheses (Jahan et al. 2018). Disease maps help to present geographic patterns that may be overlooked in a table, obscuring the geospatially related statistics (Moore and Carpenter 1999). By providing a visual representation of cancer outcomes, geographic patterns of disease are able to be identified and effectively addressed with public health policy and actions. Exeter (2017) recognises one of the key challenges with mapping spatial patterns of disease is the design of visualisations. This paper addresses the visualisation techniques and their applications to cancer statistics. Highlighting the differences and historic use of these displays.

The paper is organised as follows. The next section describes the choropleth map which is the common approach to disease maps. Section 3 surveys atlases in use today. Section 4 describes an alternative display, the cartogram what may be useful for countries that have heterogeneously sized geographic units. The pros and cons of these approaches is discussed in Section 5. Disease maps are more useful when made interactive, and common options are described in Section 6, along with a discussion of benefits and disadvantages. The last section summarises the paper and discusses future directions.

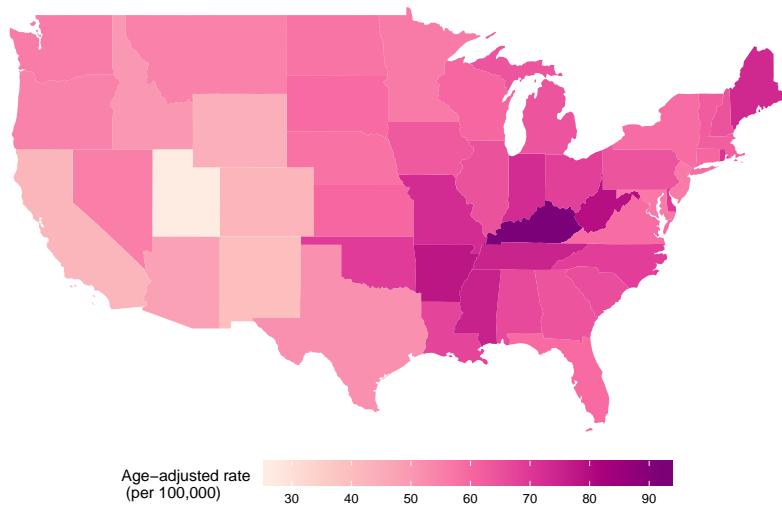


Figure 1: Choropleth map of the average age-adjusted rate per 100,000 people, of new cases of lung and bronchus cancer 2012-2016 in the USA. Utah has an unusually low rate and Kentucky is high. There is generally a west to east trend, with higher rates towards the eastern states.

2 Map displays for disease data

A choropleth map is used to display the spatial relationship of measurements by shading areas of a map. The geography is faithfully rendered, and the colour rendering is designed to reveal spatial patterns among data values. A choropleth is constructed by drawing the geographic or political boundaries, and filling the shapes with colours to represent values of a measured variable (Tufte 1990). Figure 1 shows a choropleth of age-adjusted rate (per 100,000 people) of new cases of lung and bronchus in the USA, averaged over 2012 through 2016. The data was extracted from the official federal statistics by U.S. Cancer Statistics Working Group. (n.d.) on cancer incidence and deaths, produced by the Centers for Disease Control and Prevention (CDC) and the National Cancer Institute (NCI).

Early versions of choropleth maps used symbols or patterns instead of colour. Bell et al. (2006) discuss the use of choropleths to visualise cancer data, and Walter (2001) gives an overview of the development of these maps for displaying disease data.

Utilising the state boundaries can make a map familiar to read (Brewster and Subramanian 2010), and allows viewers to visually infer the spatial relationships in the data. The familiarity of the geography is a worthy consideration when presenting results of spatial analysis. Just as geographers are no longer the only creators of maps, Bell et al. (2006) suggests the audiences of spatial health data analysis have extended beyond researchers to the public, policymakers and the media.

Identifying and explaining spatial structures, patterns, and processes involves considering the individuals in communities and organising communities into representable units (Moore and Carpenter 1999). In Figure 1, a west to east spatial trend of increasing rates, can be seen. There is also a spatial outlier – Utah has a noticeably lower rate than its neighbours. Also Kentucky has a noticeably high rate, and Maine also has a higher rate than its neighbours. There is something of a cluster of higher rates around the tobacco states.

While the areas are recognisable shapes, they are often politically driven boundaries with individual

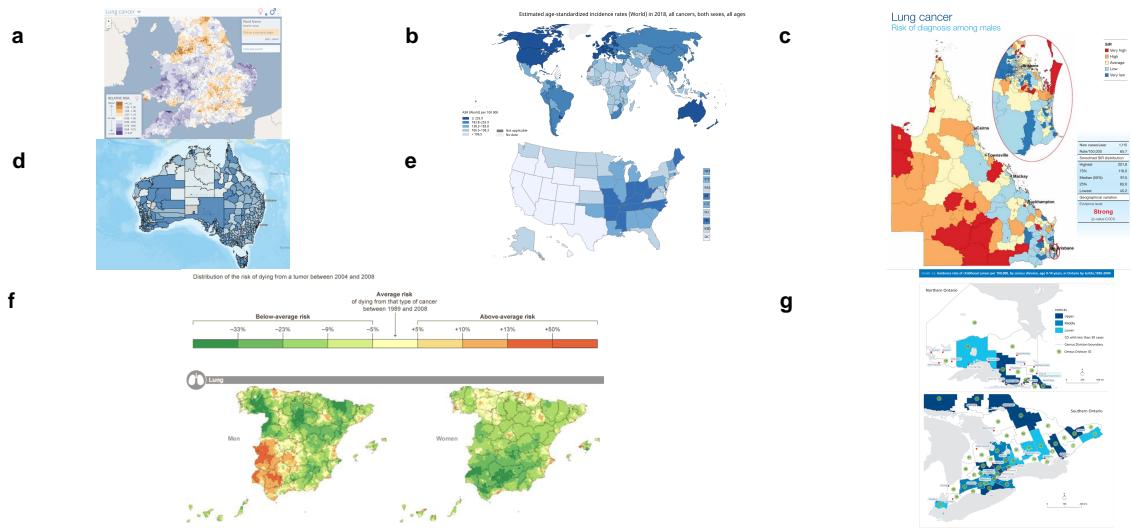


Figure 2: Publicly available choropleth cancer maps, published between 2010 and 2015. Sourced from: @roberts2019communication

areas being of non-uniform size, containing different population densities and subject to change over time. The different population and geographical sizes of administrative areas can attract attention to the shades of the unpopulated but large areas (Tufte 1990). Choropleths can inhibit visual inference when presenting human related statistics as the display may draw attention away from the ‘potentially more important results in the more populous communities’ that are geographically smaller (Exeter 2017).

3 Cancer atlases

Choropleth maps can be useful devices for communicating information to public on a familiar map base. A cancer atlas is a choropleth map, or collection of maps, representing cancer incidence and mortality for a country, or group of countries. In epidemiology, choropleths are often used as a tool to study the spatial distribution of cancer incidence and mortality. The data collection methods of cancer mortality rates across regions, and the administrative control within regions lends itself to choropleth visualisation. d’Onofrio et al. (2016) provides the definition of a cancer atlas, beginning with Haviland’s maps in 1875, they attribute UK cancer atlases to Howe (1963), and early work in US cancer atlases can be attributed to Burbank (1971). The increasing development and use of disease maps can be attributed to the availability of geographic information system software (Exeter 2017). The choropleth maps presented levels via hatchings or dots on a black and white scale. These atlases were key to developing hypotheses regarding areas with unusually high rates, geographic correlations, work related exposures, and high risk diets (d’Onofrio et al. 2016).

The presentation of cancer statistics has changed over time with greater access to computational power. Mortality rates are now often presented as relative rates of risk across the population, and age adjusted to correct for the higher prevalence of cancers in older populations. Howe (1989) describes Stock’s development of the standardised mortality ratios through the 1930s. Table 1 summarises the measures presented in published cancer atlases, and provides a definition of each measure.

Table 1: Measures used to report cancer statistics

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

The atlas of Cancer in Queensland (Cramb, L Mengersen, and Baade 2011) focussed on highlighting the difference in experience for those living in rural and disadvantaged areas. They used Standardised Incidence Ratios modelled via Bayesian methods, and explored the information presented in a range of selected atlases. These atlases provided Incidence, Survival and Mortality rates to the public on Areal or Isopleth map bases.

The presentation of these rates considers not only the shapes of the areas, but also appropriate choices of colour blind friendly colour schemes, and categories of the values to highlight significantly different areas.

Cancer maps are effective tools for communicating to wide range of audiences, including the general public and others not trained in statistical analyses. These visualisations enable non-expert audiences to interpret the outputs of sophisticated statistical analyses. Cruickshank (1947) as cited by Walter (2001), discusses using visuals as a ‘formal statistical assessment of the spatial pattern’. Overwhelmingly, cancer maps utilised to communicate to the public and other non-expert audiences are choropleths.

3.1 Overview of publicly available atlases

Roberts (2019) identified 33 publicly available cancer atlases, published between January 2010 and November 2015. All of these use choropleth maps. All but one of these were published by non-commercial organisations, including not-for-profits, government, research organisations, advocacy groups or government funded partnerships. The use of choropleths within the public domain mirrors the heavy use of choropleth maps within the research literature, discussed above. The cancer atlases

identified by Roberts (2019) covered geographies from all around the world, most focused on single nations. Figure 2 displays a global (b), national (a,d,e,f) and state (c,g) choropleths. The sections below provide details on each of the maps within Figure 2.

3.1.1 The Environment and Health Atlas of England and Wales

Figure 2.a contains The Environmental and Health Atlas of England and Wales. This map shows the relative risk for women developing lung cancer across England and Wales in 2010. The cancer data used to generate this map came from Office for National Statistics (ONS) (England) and from the Welsh Cancer Intelligence and Surveillance Unit (WCISU).

3.1.2 Globocan 2012: Estimated Cancer Incidence, Mortality and Prevalence World-wide

The map seen in Figure 2.b is from the Globocan 2012: Estimated Cancer Incidence, Mortality and Prevalence. This global map shows age standardized incidence rates (per 100,000) of all cancers for both men and women at national level, for 2018. This map is published by the World Health Organisation's International Agency for Research on Cancer. Data was sourced from cancer registries in each country, contributing registries can be seen in the supplementary material on the cancer atlas website.

3.1.3 Atlas of Cancer in Queensland

Figure 2.c shows the The Atlas of Cancer in QLD. This map was published by the Queensland Cancer Council and shows the relative incidence ratio of lung cancer in males in the state of QLD within Australia (1998 - 2007). Data to generate this map was sourced from the Queensland Cancer Registry.

3.1.4 Bowel Cancer Australia Atlas

Figure 2.d shows the Bowel Cancer Australia Atlas. Published by Bowel Cancer Australia (Australia). This map shows the percentage of males between 50 - 54 years of age that diagnosed with bowel cancer in 2016 in Australia. The source of the data is not provided.

3.1.5 United States Cancer Statistics: An Interactive Cancer Statistics Website

The United States Cancer Statistics: An Interactive Cancer Statistics Website can be seen in Figure 2.e. This map contains the Incidence rate per 100,000, of all cancer types for man and women in the United States for 2016, aggregated at the state level. The map was published by the Centers for Disease Control and Prevention. Incidence data seen in this map were compiled from cancer registries meeting U.S. Cancer Statistics data quality criteria covering 100% of the U.S. population.

3.1.6 Map of Cancer Mortality Rates in Spain

The Map of Cancer Mortality Rates in Spain can be seen in Figure 2.f. The side by side maps show relative risk of lung cancer for men vs women based on data from 2004 - 2008. The source of the data and statistical methods are unknown.

3.1.7 Atlas of Childhood Cancer in Ontario

Figure 2.g shows the incidence rate of childhood cancers per 100,000, by census division, age 0-14, in Ontario, from 1995 to 2004.

Modern mapping methods have the opportunity to incorporate interactive features. This enables supplementary information to be incorporated into online atlases without cluttering the screen. Tool tip features and drop down menus allow for user exploration with zooming and panning features to allow flexibility in the display (Monmonier 2018). The use of these supports were found in many online disease map sources. This directly enables epidemiologists to explore spatial distributions of cancer statistics. Basic interactivity is often implemented by controls outside of the plot space (Pedersen 2018), this exchanges an image or set of data for another. Some more advanced interactions include direct interactions with the plot via the use of overlaid tool tip features, very few cancer atlases involve complex selection interactivity.

These interactive features provide an opportunity for users to explore the additional information available. This helps users to understand and interpret the spatial distribution presented, as well as validate, explain or explore the presented statistics and their relationships to each other and/or their underlying spatial distribution. This allows relationships between spatial areas and diseases to be explored with sophistication in nontraditional but still ‘cognitively accessible’ ways (Carr, Wallin, and Carr 2000). The interactive features of the publicly available maps identified by Roberts (2019) allow the exploration of geographic hierarchies, population distribution, statistical uncertainty, demographics and socio-economic indicators. Carr, Wallin, and Carr (2000) suggested LM plots as a solution to linking cartography and statistical graphics.

XXX more details on the examples in the next figure

3.2 Common statistics displayed

Cancer maps are powerful visualisations that summarise complex statistical analyses, however the statistics represented in these maps cannot tell the entire story. Supplementary graphs and plots are often included to add more depth and information to the map. Bell et al. (2006) suggests 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. There are many visualisations used for displays of statistical summaries, these may be dot plots, bar plots, box plots, timeseries plots, cumulative distribution plots, scatter plots, Q-Q plots. These additional displays of the cancer distribution can provide alternative views of the cancer statistics, as well as the supporting statistics including error, confidence intervals, distributions, sample or population sizes, standard deviation and other measures. When presenting cancer maps, d’Onofrio et al. (2016) believes the intuition derived from maps must be ‘validated by rigorous statistical analyses’, the supplementary statistics help for this validation.

The interactivity of modern mapping methods enable supplementary information to be incorporated without cluttering the screen. Tool tip features and drop down menus allow for user exploration, small screens are not a problem with the flexibility provided by zooming and panning (Monmonier 2018). This allows relationships between spatial areas and diseases to be explored with sophistication in nontraditional but still ‘cognitively accessible’ ways (Carr, Wallin, and Carr 2000). The use of these supports were found in many online disease map sources. These interactive features provide an opportunity for users to explore additional information to help users understand and interpret the spatial distribution presented, as well as validate, explain or explore the presented statistics and their relationships to each other and/or their underlying spatial distribution. This directly enables epidemiologists to explore spatial distributions of cancer statistics. The interactive features of the publicly available maps identified by Roberts (2019) include exploration of geographic hierarchies, population distribution, statistical uncertainty, demographics and socio-economic indicators. Carr, Wallin, and Carr (2000) suggested LM plots as a solution to linking cartography and statistical graphics.

3.2.1 Geographic hierarchies

While atlases are often used to describe differences between areas, statistics may be displayed at different levels of aggregation. Global health statistics can be aggregated to administrative and arbitrarily defined regions, such as those used by the World Health Organisation and the United Nations (Ferlay J 2018). World atlases can allow for displays of data aggregated into continents, countries, states, provinces and congressional districts (Group 2019).

3.2.2 Population distribution

It is extremely likely that each population area will have a different number of people. The distribution of the population residing in all areas may also be communicated in a table or histogram display (“All-Ireland Cancer Atlas 1995-2007” 2011). Atlases have the opportunity to connect the population to the land available to them by communicating population density. Atlases can also connect the population to the land available to them by communicating population density.

3.2.3 Statistical uncertainty

Additional statistics that accompany an atlas often include a measure of the statistical uncertainty surrounding the statistics presented in a choropleth. In the review of atlases in the public domain, atlases were considered to report uncertainty to the non-expert user if they included a measure of statistical uncertainty either within or alongside the map (Roberts 2019). The maps considered used standard and well known measures including credible intervals and standard deviation, statistical significance, box plots and distributions. Other methods involve providing adjacent maps or overlapping maps with symbols (Kronenfeld and Wong 2017). The maps employing uncertainty ranged from static documents or infographics, to interactive online visualisations. Communicating the statistical uncertainty associated with the estimates occurs using confidence intervals (CI), credible intervals (CrI), statistical significance levels, boxplots, distribution plots, and reporting sample size and standard deviations. Close to half of the atlases identified (42%, n=14) included some measure of uncertainty. The most common measure used to represent uncertainty were credible or confidence intervals (CIs).

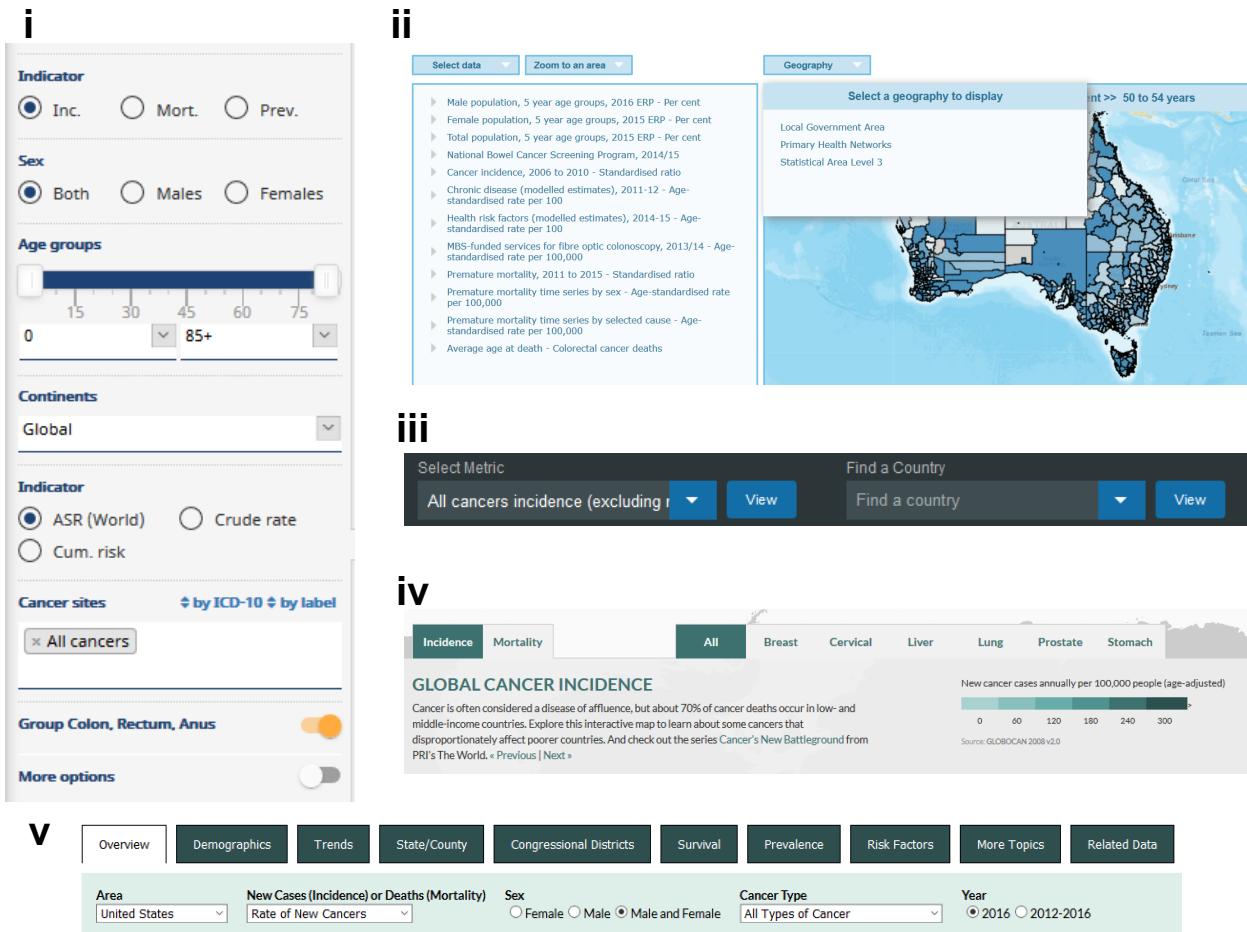


Figure 3: Interactive controls of displays in publicly available choropleth cancer maps. a. Globocan 2018: Cancer Today(Controls for indicator, sex, age groups, continents, and cancer types), b. Bowel Cancer Australia Atlas (Controls for indicator, age, year, and geographical areas), c. The Cancer Atlas (Controls for predetermined combinations of cancer type or risk factors, mortality or incidence, and gender) d. Global Cancer Map (Controls for indicator, and cancer type), e. United States Cancer Statistics: Data Visualizations (Controls for demographics, trends, geography, indicators, risk factors, sex, cancer type, and year).

3.2.4 Demographics

Demographics include information regarding the age and sex distribution of the areas presented. Sex is an important cofactor for cancer atlases. As some cancers are sex specific, and others may be found in both males and females, atlases often specify the relevant sex as part of the visual output in the displays. Digital atlases allow for users to interact with the controls of the displays, they can select males, females or both depending on the type of cancers explored.

3.2.5 Socio-economic indicators

Socio-economic indicators can explain how the experience of cancer prevalence varies for various members of a society. These indicators include unemployment rates, poverty rates, remoteness, and education levels achieved though, only a few atlases also explored the impact of rurality on cancer rates. These rates may also be explored as percentages above or below the mean or median value for the set of spatial areas. The Human Development Index can be used to understand the socio-economic experience of a community, as well as Income levels which can be sourced from the World Bank list of economies (Ferlay J 2018). The areas are often ranked and allocated to quintiles, each quintile can be presented as categories describing the ranking.

3.3 Interaction with atlases

4 Cartograms and other spatially warped displays

4.1 Cartograms

A cartogram alters the map base with the intention of improving the presentation of the statistic of interest. For a single variable of interest, each map area is changed to emphasise the distribution by representing the corresponding value, in comparison to the value of the other areas (Dougenik, Chrisman, and Niemeyer 1985). Changes in the map base are implemented by altering the boundaries, and therefore shapes, of individual areas.

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

Choropleths may be considered true topological maps, however, if the land mass displayed covers enough of the globe, there must be a transformation or distortion to display the land in 2D (Monmonier 2018). The amount of distortion is related to the distance covered by the landmass displayed Tobler (1963). World map projections reflect the frequent perspectives used to view the earth. Choropleth maps will always be distorted if they cover enough of the globe, just like photographs of the globe from space. Choropleth creation requires choosing a map projection that shows a favourable distortion of the geography for presenting the set of spatial information. Selecting a display can prevent misinterpretation of global statistics, as global maps face the challenge of

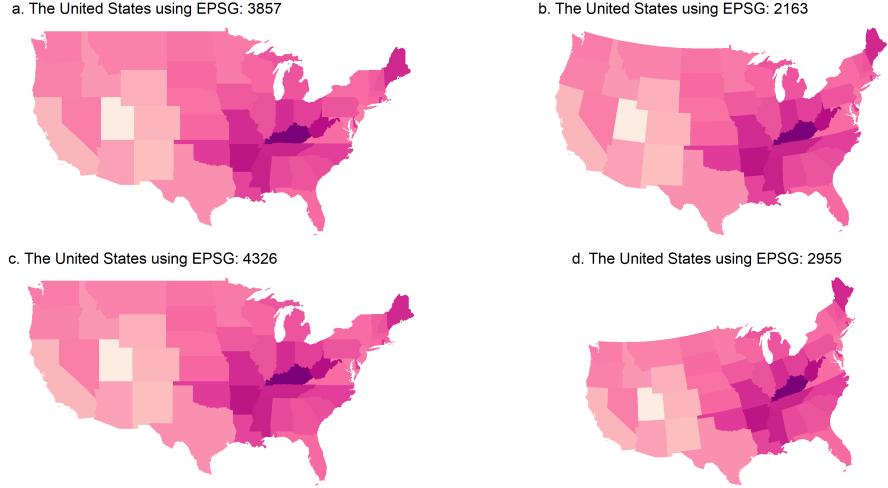


Figure 4: Four choropleth maps of the United States of America using various coordinate reference systems. Each state has been coloured according to the average age-adjusted rate of incidence for lung and bronchus for females and males in the United States 2012-2016. The map projections alter the shapes and angles of the boundaries of each state. Maps a and b are similar in their straight edges, unlike maps c and d which curve on the northern United States border.

equitable displays of land mass on maps (Raisz 1963). If the statistic presented on the map base relies on physical distance and is influenced by the topology there is no transformation needed, beyond choosing a reasonable projection.

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 (Kronenfeld and Wong 2017). The purposeful distortion of the map space, transformed according to population density, is beneficial when a uniform density of the map base is desired. Population then becomes a uniformly distributed background for the statistic presented (Berry, Morrill, and Tobler, n.d.). Dorling (2011) suggests ‘population distribution is often extremely uneven in former British colonies’, this makes the distortion necessary (Griffin 1980). When implementing a distortion of the geographical shape according to population, the resulting display is an area cartogram (Olson 1976), or population-by-area cartogram (Levison and Haddon Jr 1965).

Cartograms provide an alternative visualisation 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 (Dorling 2011). Monmonier (2018) suggests that white lies may be employed to create useful displays and map creators have the ability to draw lines that may distort the geometry and suppress features and it is easy for the average person to disregard the impact of transformations used to create cartograms. Cartograms may be seen as an extension of map transformations and projections. The favourable distortion is proportional to a value other than actual earth size area (Olson 1976). A disadvantage of the conventional map is that sparsely populated rural areas may be emphasized, whereas the areas representing cities are very small, making interpretation of spatial patterns very difficult. The distortion of a cartogram accounts for the population density, preventing it from obscuring the spatial patterns (Levison and Haddon Jr 1965). The spatial transformation of map regions relative to the data emphasises the data distribution instead of land size (Kocmoud and House 1998). When visualising population statistics, Dorling (2011) considers this equitable representation design ‘more socially just’, or

honest (Dent 1972), giving due attention to all members of the population and reducing the visual impact of large areas with small populations (Walter 2001). Howe (1989) suggests that ‘cancer occurs in people, not in geographical areas’ and Griffin (1980) believe that spatial socio-economic data, like cancer rates, are best presented on a cartogram for urban areas as the population map base avoids allocating ‘undue prominence’ to rural areas. Jahan et al. (2018) encourage the use of cartograms to highlight small areas and uncover local-level inequalities.

The creation of cartograms was historically in the hands of professional cartographers (Kraak 2017). Early approaches including John Hunter and Jonathan Young (1968) and Durham’s wooden tile method, Skoda and Robertson’s (1972) steel ball bearing approach and Tobler’s (1973) computer programs (Dorling 2011). Geographical information systems allowed map users, and researchers to implement their own cartograms, but these systems are utilised depending on ‘the effectiveness, efficiency, and satisfaction of the map products (Nielsen 1994), (Kraak 2017). Howe (1989) discusses the impact of electronic computer-assisted techniques.

There are many alternatives to consider, the intended audience of the map, and its purpose are key points in cartogram use and creation. Dorling (2011) reiterates: “There is no “best” cartogram or method of creating cartograms just as there is no “best” map” (Monmonier and Schnell, 1988). Nusrat and Kobourov (2016) provided a framework to investigate implementations of the many algorithms presented, and the “statistical accuracy, geographical accuracy, and topological accuracy”.

Table 2: Maps used to present statistics for the United States of America

| Map display | Details |
|-------------------------|--|
| a. contiguous cartogram | A contiguous cartogram map of the United States of America. Each state has been coloured according to the average age-adjusted rate of incidence for lung and bronchus for females and males in the United States 2012-2016. Each stated shape has been distorted according to the population of the state in 2015. The state of California has become much larger due to it’s large population density. This draws attention to the densely populated North East region, and detracts from the less populated Mid West. |
| b. Non - Contiguous | A Non - contiguous cartogram map of the United States of America. Each state has been coloured according to the average age-adjusted rate of incidence for lung and bronchus for females and males in the United States 2012-2016. Each state shape has been maintained, but the size has altered according to the population of the state in 2015. The state of California has remained closer to it’s original size than it’s surrounding states. The North East states have remained closer to their geographical size, in the case of Massachusetts and Connecticut. This draws attention to the densely populated North East region, and the sparse Mid West. |

| Map display | Details |
|---------------------|--|
| Dorling | A dorling cartogram of the United States of America. Each state has been coloured according to the average age-adjusted rate of incidence for lung and bronchus for females and males in the United States 2012-2016. Each state has been represented by a circle, but the size was determined by the population of the state in 2015. The North East states remain closer to their neighbours, and may be displaced from their geographic location. The sparsity of the population in the Mid West is highlighted by the distance between the circles, located at the geographic centroids. |
| Hexagon Tesselation | A hexagon tessellation map of the United States of America. Each state has been coloured according to the average age-adjusted rate of incidence for lung and bronchus for females and males in the United States 2012-2016. Each state has been represented by a square, each square is the same size. The neighbouring states are easily contrasted, however the north east regions have been displaced from their geographic location. The sparsity of the population in the Mid West is highlighted by the light yellow colour, the age adjusted rate in Kentucky is the darkest and its neighbours are similar. |

4.1.1 Contiguous

A contiguous cartogram maintains connectivity of the map regions while areas are altered according to a statistic. This transformation often occurs at the expense of the shape of areas (Kocmoud and House 1998, @NAC, @TAAM). From a computer graphics perspective, Min Ouyang and Revesz (2000) explain the application of ‘map deformation’ to account for the value assigned to each area, they provide three methods for creating value-by-area cartograms. Examples include Tobler’s Pseudo-Cartogram Method, Dorling’s Cellular Automaton Method (2011), 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) (1998).

An intentional goal when creating the 1966 Census population cartogram for Canada was to maintain contiguity, while attempting to keep the actual shape of places. The end result was a ‘very accurate isodemo-graphic map of Canada’. This intentional design goal coincided with the rising interest in urban geography and presentation of social statistics.

To be able to recognise the significant changes, a reader will usually have to know the initial geography to find the differences in the new cartogram layout (Olson 1976). 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. Kocmoud and House (1998) presents this issue as conflicting tasks or aims, to adjust region sizes and retain region shapes. Cano et al. (2015) define the term ‘mosaic cartograms’ for hexagonal tile displays, where the number of tiles for each area can be used to communicate the statistic of regions. The complexity of the boundaries can be adjusted in the resulting display, as the size of the tiles used allows a trade-off to be made between boundary complexity and simplicity.

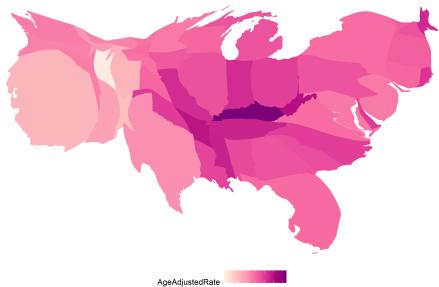
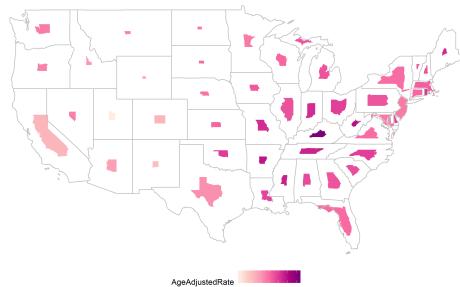
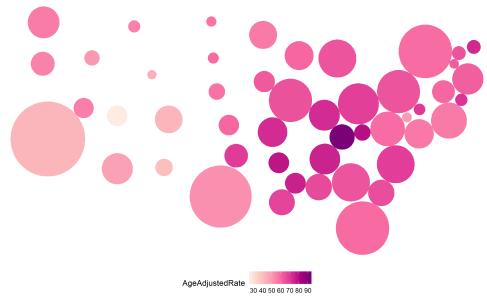
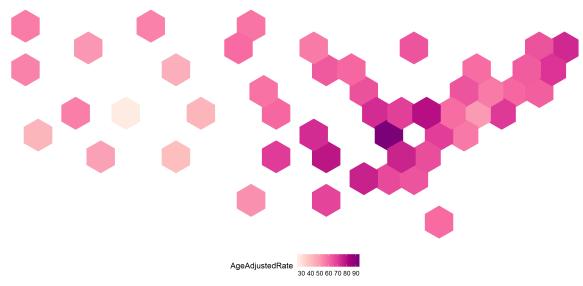
a**b****c****d**

Figure 5: Alternative maps of the United States of America. Each state has been coloured according to the average age-adjusted rate of incidence for lung and bronchus for females and males in the United States 2012-2016. Each stated shape has been distorted according to the population of the state in 2015. The state of California has become much larger due to it's large population density. This draws attention to the densely populated North East region, and detracts from the less populated Mid West.

4.1.2 Non-Contiguous

Non-Contiguous cartograms succeed in maintaining the shape of the areas presented. Each area stays in a similar position to their location on a choropleth map. The choropleth map base is often also plotted as a comparison point to highlight the change in area. The addition is the gap between areas, created as each individual area shrinks or grows according the associated value of the statistic. Olson (1976) discusses creation of these maps, the significance of the empty areas left between the geographic boundaries and the new shape, and the ‘degree of difference from the original map that is the real message’ of these displays.

As the trade-off regarding boundaries approaches simplicity, the distortion of region shapes on the contiguous cartogram presents an additional hurdle to visual recognition and this hurdle is not only eliminated on the non-contiguous cartogram but is replaced by the meaningful empty-space property (Olson 1976, @ECGC). The shapes are valuable for recognition and allows users to orient themselves on the display. Map creators can efficiently communicate with this kind of map by keeping the outlines or particular elements of the original in the new shape (Dent 1972). The scale of the areas does not impact on the shape recognition. However it may impact on the visibility of all areas if small areas expand beyond their boundaries.

4.1.3 Dorling

Daniel Dorling presented an alternative display engineers to highlight the spatial distribution and neighbourhood relationships without complex distortions of borders and boundaries. This approach opposes preserving the intricate shape details and is founded in the simple question put forward by Daniel Dorling (2011):

“If, for instance, it is desirable that areas on a map have boundaries which are as simple as possible, why not draw the areas as simple shapes in the first place?”

He acknowledged the sophistication of contiguous cartograms but critiqued their ‘very complex shapes,’ he answers this with his implementation of maps created using ‘the simplest of all shapes’. Circular cartograms use the same simple shapes for every region represented, and resizes the shapes according to the statistic represented or the population for a base map. This familiar shape may be more effective for understanding the spatial distribution than contiguous cartograms, as the ‘nonsense’ shapes used have ‘no meaning’ after distortions are applied Dent (1972). To produce a compelling map, a gravity model is applied to avoid overlaps, and keep spatial relationships with neighbouring areas over many iterations. This implementation can work for up to ‘one hundred thousand’ areas.

The groundwork for this approach had been laid in the mid 1930’s by Raisz, and rectangular cartograms provide dramatic comparisons and are especially useful for correcting misconceptions communicated by geographic maps. Tobler (2004) quotes the official definition of Value-Area Cartograms, the simplistic displays which represent each area as a single rectangle, sized according to the value of the statistic. This rectangular display also allows for tiling, where geographic neighbours placed in suitable relative positions also share borders, however contiguity may be sacrificed (Monmonier 2005). Rectangular cartograms allow for bivariate displays, population can be effectively communicated by the size of each rectangular, and a second variable can be communicated using colour (Kreveld and Speckmann 2007).

Figure 6: A tilegram of the United States of America. Each state has been coloured according to the average age-adjusted rate of incidence for lung and bronchus for females and males in the United States 2012-2016. Each state has been represented by a square, each square is the same size. The neighbouring states are easily contrasted, however the north east regions have been displaced from their geographic location. The sparsity of the population in the Mid West is highlighted by the light yellow colour, the age adjusted rate in Kentucky is the darkest and its neighbours are similar.

A similar method, where each geographic area is represented using a square, tessellated to create a square grid. This method has been used by 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. Figure 6 shows a square tilegram representing each state using a single square.

Recommended criteria to contrast mapping methods include average cartographic error, and maximum cartographic error, correct adjacencies, maximum aspect ratio, and suitable relative positions (Kreveld and Speckmann 2007). However, this does not consider the issues with actually producing rectangular cartograms. Algorithms for the creation of rectangular cartograms

4.2 Tilegram

Bob Rudis, geogrid

5 Comparison of maps and cartograms

Visualisation properties have been explored across several disease mapping displays. To choose an appropriate map display, the map creator must consider the intended user, and message the map will communicate. Choropleth displays utilise more traditional cartographic methods, they are usually true to the topography, displaying familiar boundaries of countries, states or administrative areas. Choropleth methods require a decision to be made about the projection of the display, in this case aim to find a map projection that gives the least distorted representation of the geography (Tobler 1999). When there is a relationship between the cancer type being mapped and geographical features, a choropleth map would pair this information visually. These maps are best used when

the geography is of importance to the data in the display. Users can identify the areas relevant to them by the familiar boundaries, and can easily contrast the experience of geographic neighbours.

Pairing geography and cancer data was historically the work of epidemiologists, however it did not require them to be the map designer. This was expected as cartographers were the creators of geographic displays, and epidemiologists utilised the maps that had been prepared in advance. Cartograms with the population variable used for base map distortion can be seen as density equalising maps, that reduce the visual prominence of low-population areas (Kronenfeld and Wong 2017). This approach began a shift in the focus and responsibility of map creators. Cartogram displays began incorporating the statistics and population into the design of the display. Shifting and sometimes sacrificing familiar boundaries to draw attention to outliers in the data space or population density. The difference between the familiar map display helps highlight the impact of the disease on communities.

By removing the unexpected boundaries and shapes, the spatial distribution is the primary concept communicated by the display. Areas that are geographically close will maintain connectedness in some way, but the population and the statistic will take precedence. These alternatives are especially helpful for data aggregations where administrative boundaries break populations into groups. The experience of each collective may be worth considering as their experiences may be similar in terms of the services and facilities they share. Population affects are highlighted by these displays, as this display allows a more equitable view of each community, and does not minimise those that operate on a smaller geographic scale.

Table 3: Mapping methods used to display cancer statistics

| | Choropleth | Contiguous | Non-contiguous | Dorling | Tilegrams |
|--------------------------------|------------|------------|----------------|---------|-----------|
| Preserves shapes | Y | N | Y | N | N |
| Preserves neighbours | Y | Y | Y | S | S |
| Uniform use of shape | N | N | N | Y | Y |
| Bivariate display ¹ | N | Y | Y | Y | Y |

Creating maps of diseases now involves more decisions to be made by map makers, rather than cartographers. Technology has played an enormous role in increasing the opportunities for map makers. The computation and graphics power have made creation, alteration and interactivity possible. as these options have expanded and it is the objectives of the investigator that will drive the choices. Bell et al. (2006) and Moore and Carpenter (1999) have provided suggestions and comments to help map creators best communicate their health data and spatial analyses.

6 Animation and Interactivity

Epidemiologists consider the interactions between the spatial distribution of a disease, the environment and the people affected. Recent developments in technology have created an opportunity to migrate the presentation of epidemiological results from atlases in paper map collections to web atlases. When the map is supplementary to the discussion, static visualisations can help to aid

¹A bivariate display refers to a map that allows two variables to be plotted. One variable via the size of the area, and another by filling the area with a colour.

the written results provided by experts, and pairing an explanation with a map or diagram can help to draw attention to a point, group or area of interest. Static maps often need to include enough information for the display to be understandable out of the original context. There are now many suitable methods and mediums to communicate health data, the choice between them should always be driven by the intended message to communicate and the needs of the audience.

Bell et al. (2006) provide three questions that are generally when using a disease map:

- What is the mortality rate in a certain area?
- Are there geographic trends in the data, or regions of unusually high or low rates?
- Is the lung cancer mortality pattern similar to the pattern of smoking prevalence shown in a companion map?

A single map display historically created by epidemiologists, may not effectively answer all three at the same time. This is especially true when areas are small and difficult to identify in a display. However, with the ability to animate, map creators can specify and answer these questions by directing the attention of a map user to the areas of interest.

Animations provide a linear, fixed narrative to passive users, communicating a message by capturing and directing attention (Pedersen 2018). Pedersen (2018) suggests the use of moving objects demands attention by taking advantage of the behaviour of the visual cognition system. They can be produced to not tolerate or require user input, creating useful, directed descriptions of processes, such as highlighting a particular region or time period (Monmonier 2018). Animations ensure all users follow the same interpretive path but creators cannot control the level of engagement.

Animations can be implemented to allow users varying degrees of control. A basic form of interactivity with animations is the play, pause and rewind of films, where the film is intended to be watched in one direction but the position in the animation can be controlled. Bell et al. (2006) provide weather maps as well developed and appropriate examples of spatial displays to communicate to the general public, and encourage those creating disease maps to leverage the success of weather displays. The movement of a weather system will follow a forecasted path, the animation will communicate this so that all map users can follow the path of the weather system across the geography over a specified period of time. More complex animations may include steps, when user interaction via clicks animates the display between stages. The Australian Cancer Atlas (n.d.) provides tours which change the display to draw user's attention to areas on the map that are relevant to the story. This directs user's attention and gives them tools to plan their own exploration.

The move from static, print media displays toward interactivity allow users to develop maps to ask and answer their own questions. Where the needs of the audience is changeable and is the priority, the map creator can allow interactivity for map users to explore.

Developments and adoptions of this technology has allowed map users more control of the message displayed. Users are able to see the familiar shapes, zoom in to explore densely populated communities, or change the display completely to highlight the population space.

This opportunity provides room to move away from the traditional approach of a single map for a single use, into a more explorative territory. Interactivity allows users to drive their own exploration, controlling a display through variable controls, links, tool tips, and data selection tools (Pedersen 2018).

Where the message is most important, static or animated graphics allow control over the display and interpretation. However before publishing, it is important for map creators to develop and then test the message being communicated, this task requires expert knowledge.

7 Acknowledgements

We would like to acknowledge Dr Earl Duncan (Research Associate at ARC Centre of Excellence for Mathematical & Statistical Frontiers, QUT). His time and effort given to edit and comment on this paper was invaluable.

We thank Dr Susanna Cramb (Spatial Modeller, Cancer Council Queensland) and Dr Peter Baade (Senior Research Fellow, Cancer Council Queensland) for providing the opportunity to cover this exploration of disease mapping methods in writing. This literature review would not be possible without the opportunity provided by Queensland University of Technology, and Cancer Council Queensland, and the roles of Professor Kerrie Mengersen (Professor of Statistics, Science and Engineering Faculty, QUT) and Dr Earl Duncan in contributing to the Australian Cancer Atlas and supervision of Stephanie's Research Masters.

It was in the development of this online cancer atlas that methods for disease map displays, and visual communication strategies were explored. We are thankful for the opportunity to write about these visualisations and the situations in which they are appropriate.

8 References

Wickham (2017) for data analysis, Bivand, Nowosad, and Lovelace (2019) and Pebesma (2018) to implement plotting of spatial data. Kobakian and Cook (2019) to create hexagon tessellation. Arnold (2019) to enhance plot displays. Jeworutzki (2018) for contiguous and non-contiguous cartogram displays.

"All-Ireland Cancer Atlas 1995-2007." 2011.

Arnold, Jeffrey B. 2019. *Ggthemes: Extra Themes, Scales and Geoms for 'Ggplot2'*. <https://CRAN.R-project.org/package=ggthemes>.

Atlas, Australian Cancer. n.d. "Australian Cancer Atlas." <https://atlas.cancer.org.au>.

Bell, B Sue, Richard E Hoskins, Linda Williams Pickle, and Daniel Wartenberg. 2006. "Current Practices in Spatial Analysis of Cancer Data: Mapping Health Statistics to Inform Policymakers and the Public." *International Journal of Health Geographics* 5: 49. <https://doi.org/10.1186/1476-072X-5-49>.

Berry, Brian J. L., Richard L. Morrill, and Waldo R. Tobler. n.d. "GEOGRAPHIC Ordering of Information: NEW Opportunities" 16 (4): 39–44. https://doi.org/10.1111/j.0033-0124.1964.039_q.x.

Bivand, Roger, Jakub Nowosad, and Robin Lovelace. 2019. *SpData: Datasets for Spatial Analysis*. <https://CRAN.R-project.org/package=spData>.

Brewster, Mark B., and S. V. Subramanian. 2010. "Cartographic Insights into the Burden of Mortality in the United Kingdom: A Review of 'The Grim Reaper's Road Map.'" Journal Article. *International Journal of Epidemiology* 39 (4): 1120–2. <https://doi.org/10.1093/ije/dyp395>.

- Cano, R. G., K. Buchin, T. Castermans, A. Pieterse, W. Sonke, and B. Speckmann. 2015. "Mosaic Drawings and Cartograms." *Computer Graphics Forum* 34 (3): 361–70.
- Carr, D. B., J. F. Wallin, and D. A. Carr. 2000. "Two New Templates for Epidemiology Applications: Linkedmicromap Plots and Conditioned Choropleth Maps." *Statistics in Medicine* 19: 2521–38.
- Cramb, Susanna, Kerrie L Mengersen, and Peter Baade. 2011. "Developing the Atlas of Cancer in Queensland: Methodological Issues." *International Journal of Health Geographics* 10 (January): 9. <https://doi.org/10.1186/1476-072X-10-9>.
- Dent, Borden D. 1972. "A Note on the Importance of Shape in Cartogram Communication." *Journal of Geography* 71 (7): 393–401. <https://doi.org/10.1080/00221347208981697>.
- d'Onofrio, Alberto, Chiara Mazzetta, Chris Robertson, Michel Smans, Peter Boyle, and Mathieu Boniol. 2016. "Maps and Atlases of Cancer Mortality: A Review of a Useful Tool to Trigger New Questions." Journal Article. *Ecancermedicalscience* 10: 670–70. <https://doi.org/10.3332/ecancer.2016.670>.
- Dorling, Daniel. 2011. "Area Cartograms: Their Use and Creation." In *Concepts and Techniques in Modern Geography (CATMOG)*, 59:252–60. <https://doi.org/10.1002/9780470979587.ch33>.
- Dougenik, James A., Nicholas R. Chrisman, and Duane R. Niemeyer. 1985. "AN Algorithm to Construct Continuous Area Cartograms." *The Professional Geographer* 37 (1): 75–81. <https://doi.org/10.1111/j.0033-0124.1985.00075.x>.
- Exeter, Daniel J. 2017. "Spatial Epidemiology." Journal Article, 1–4. <https://doi.org/10.1002/9781118786352.wbieg0283>.
- Ferlay J, Lam F, Ervik M. 2018. "Global Cancer Observatory: Cancer Today." <https://gco.iarc.fr/today>.
- Griffin, T.L.C. 1980. "Cartographic Transformation of the Thematic Map Base." *Cartography* 11 (3): 163–74. <https://doi.org/10.1080/00690805.1980.10438102>.
- Group, U.S. Cancer Statistics Working. 2019. "U.S. Cancer Statistics Data Visualizations Tool, Based on November 2018 Submission Data (1999–2016)." www.cdc.gov/cancer/dataviz.
- Howe, G. M. 1989. "Historical Evolution of Disease Mapping in General and Specifically of Cancer Mapping." In *Cancer Mapping*, edited by Peter Boyle, Calum S. Muir, and Ekkehard Grundmann, 1–21. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Jahan, Farzana, Earl Duncan, Susanna Cramb, Peter Baade, and Kerrie Mengersen. 2018. "Making More of Spatial Maps: A Bayesian Meta-Analysis Approach." In.
- Jeworutzki, Sebastian. 2018. *Cartogram: Create Cartograms with R*. <https://CRAN.R-project.org/package=cartogram>.
- Keim, D.A, S.C North, C Panse, and J Schneidewind. 2002. "Efficient Cartogram Generation: A Comparison." In *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002*, 2002:33–36. IEEE.
- Kobakian, Stephanie, and Dianne Cook. 2019. *Sugarbag: Create Tessellated Hexagon Maps*. <https://CRAN.R-project.org/package=sugarbag>.
- Kocmoud, Christopher J., and Donald H. House. 1998. "A Constraint-Based Approach to Constructing Continuous Cartograms." In.

- Kraak, M. J. 2017. “Cartographic Design.” In *The International Encyclopedia of Geography : People, the Earth, Environment, and Technology*, edited by D. Richardson, 1–16. United States: Wiley.
- Kreveld, Marc van, and Bettina Speckmann. 2007. “On Rectangular Cartograms.” *Computational Geometry* 37 (3): 175–87. [https://doi.org/https://doi.org/10.1016/j.comgeo.2006.06.002](https://doi.org/10.1016/j.comgeo.2006.06.002).
- Kronenfeld, Barry J., and David W. S. Wong. 2017. “Visualizing Statistical Significance of Disease Clusters Using Cartograms.” Journal Article. *International Journal of Health Geographics* 16 (1): 19. <https://doi.org/10.1186/s12942-017-0093-9>.
- Levison, M. E., and W. Haddon Jr. 1965. “THE Area Adjusted Map. AN Epidemiologic Device.” Journal Article. *Public Health Reports* 80: 55–59.
- Min Ouyang, and P. Revesz. 2000. “Algorithms for Cartogram Animation.” In *Proceedings 2000 International Database Engineering and Applications Symposium (Cat. No.PR00789)*, 231–35. <https://doi.org/10.1109/IDEAS.2000.880581>.
- Monmonier, M. 2018. *How to Lie with Maps, Third Edition*. University of Chicago Press. <https://books.google.com.au/books?id=MwdRDwAAQBAJ>.
- Monmonier, Mark. 2005. “Cartography: Distortions, World-Views and Creative Solutions.” *Progress in Human Geography* 29 (2): 217–24. <https://doi.org/10.1191/0309132505ph540pr>.
- Moore, Dale A., and Tim E. Carpenter. 1999. “Spatial Analytical Methods and Geographic Information Systems: Use in Health Research and Epidemiology.” *Epidemiologic Reviews* 21 (2): 143–61. <https://doi.org/10.1093/oxfordjournals.epirev.a017993>.
- Nusrat, Sabrina, and Stephen G. Kobourov. 2016. “The State of the Art in Cartograms.” *CoRR* abs/1605.08485. <http://arxiv.org/abs/1605.08485>.
- Olson, Judy M. 1976. “NONCONTIGUOUS Area Cartograms.” *The Professional Geographer* 28 (4): 371–80. <https://doi.org/10.1111/j.0033-0124.1976.00371.x>.
- Pebesma, Edzer. 2018. “Simple Features for R: Standardized Support for Spatial Vector Data.” *The R Journal* 10 (1): 439–46. <https://doi.org/10.32614/RJ-2018-009>.
- Pedersen, Thomas Lin. 2018. “The Grammar of Animation.” 2018. <https://youtu.be/21ZWDrTukEs>.
- Raisz, Erwin. 1963. “Rectangular Statistical Cartograms of the World.” Journal Article. *Journal of Geography* 35 (1): 8–10. <https://doi.org/10.1080/00221343608987880>.
- Roberts, Jessie. 2019. “Communication of Statistical Uncertainty to Non-Expert Audiences.” PhD thesis, Queensland University of Technology.
- Tobler, W. 1999. *Unusual Map Projections*. Association of American Geographers.
- Tobler, Waldo. 1963. “Geographic Area and Map Projections,” January, 59–78. <https://www.jstor.org/stable/212809>.
- . 2004. “Thirty Five Years of Computer Cartograms.” *Annals of the Association of American Geographers* 94 (1): 58–73.
- Tufte, Edward R. 1990. *Envisioning Information*. Graphics Press.
- U.S. Cancer Statistics Working Group. n.d. “U.S. Cancer Statistics Data Visualizations Tool, based on November 2018 submission data (1999-2016).” www.cdc.gov/cancer/dataviz; U.S. Department

of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute.

Walter, S. D. 2001. "Disease Mapping: A Historical Perspective." Book.

Wickham, Hadley. 2017. *Tidyverse: Easily Install and Load the 'Tidyverse'*. <https://CRAN.R-project.org/package=tidyverse>.