

A Hexagon Tile Map Algorithm for Displaying Spatial Data

by Stephanie Kobakian, Dianne Cook, and Earl Duncan

Abstract The algorithm presented creates a display that uses tessellated hexagons to represent a set of spatial polygons. It allocates these hexagons in a manner that preserves the spatial relationship of the geographic units, in light of their positions to points of interest. The display showcases spatial distributions, by emphasising the small geographical regions that are often difficult to locate on geographic maps. Spatial distributions have been presented on alternative representations of geography for many years. In modern times, interactivity and animation have allowed alternative displays to play a larger role. Alternative representations have been popularised by online news sites, and atlas websites with a focus on public consumption. Applications are increasingly widespread, especially in the areas of disease mapping, and election results.

Introduction

Many cancer atlases present geospatial cancer data on a choropleth map display. The Australian Cancer Atlas ([Cancer Council Queensland](#), [Queensland University of Technology](#), and [Cooperative Research Centre for Spatial Information](#), 2018) is a recent addition to the many cancer atlas maps worldwide. The ground-breaking atlas for Australia presents a central map that shows the landmass overlaid with administrative boundaries. This choropleth display can highlight the geographic patterns in geospatially related cancer statistics ([Moore and Carpenter](#), 1999).

Over time the density of living in major cities has increased as residents have gathered to live near urban areas ([Dorling](#), 2011). The division of these residents into approximately equal population areas results in dramatically different square meterage of administrative geographic units, such as states or electorates. In a choropleth map display the geographic units are filled with colour to represent the value of the statistic for each unit ([Tufte](#), 1990). This can allow unequal emphasis on the statistical value of each of the geographic units. This allows choropleth map displays to misrepresent the spatial distributions of human related statistics due to area-size bias ([Skowronnek](#), 2016). Figure 1 is a choropleth map that uses colour to display the estimated Standardised Incidence Ratios (SIRs) of thyroid cancer for females, in each of the Statistical Areas at Level 2 (SA2) used by the [Australian Bureau of Statistics](#) (2011). The Australian choropleth map display draws attention to the expanse of dark and light blue areas across the rural communities in all states. The SA2s around Brisbane and in northern New South Wales stand out as more orange and red. However, this display neglects the vast amount of Australia residents living in the densely populated capital cities.

The solutions to this visualisation problem begin with the geography. Alternative maps shift the focus from land area and shape, to the value of the statistics in a collection of areas ([Dougenik et al.](#), 1985). Each style of cartogram applies a different transformation to the geographic areas, to highlight the values of the statistic of interest. These displays result in a distortion of the map space to represent features of the distribution across the areas ([Dougenik et al.](#), 1985) as the statistic of interest is used to determine the cartogram layout.

Alternative mapping methods allow increased understanding of the spatial distribution of a variable across the population, by fairly representing the population in each administrative area ([Levison and Haddon Jr](#), 1965). This acknowledges that the number of residents can be different but recognises that each area, or person within it is equally important.

This paper contains a discussion of Existing Mapping Practices in Section 2. This is followed by details of the Algorithm in Section 3. Section 4 describes the implementation of the algorithm in the *sugarbag* package. Section 5 outlines the use and benefits of animation.

Existing Mapping Practices

The familiar choropleth map display is often used to allow the user to orient themselves. However, the emphasis on the land mass diminishes the features of the distribution in densely populated communities due to the small size on the display ([Dorling](#), 2011). The unique shapes of boundaries can be helpful for orienting users but may not contribute to their understanding of the spatial disease distribution as many of the communities are not visible in a choropleth display ([Dorling](#), 2012). Each Statistical Area at Level 2 (SA2) ([Australian Bureau of Statistics](#), 2011) was designed to represent a community. These administrative areas are often used to aggregate census data and used to understand demographics within communities of the Australian population. This presents an opportunity for

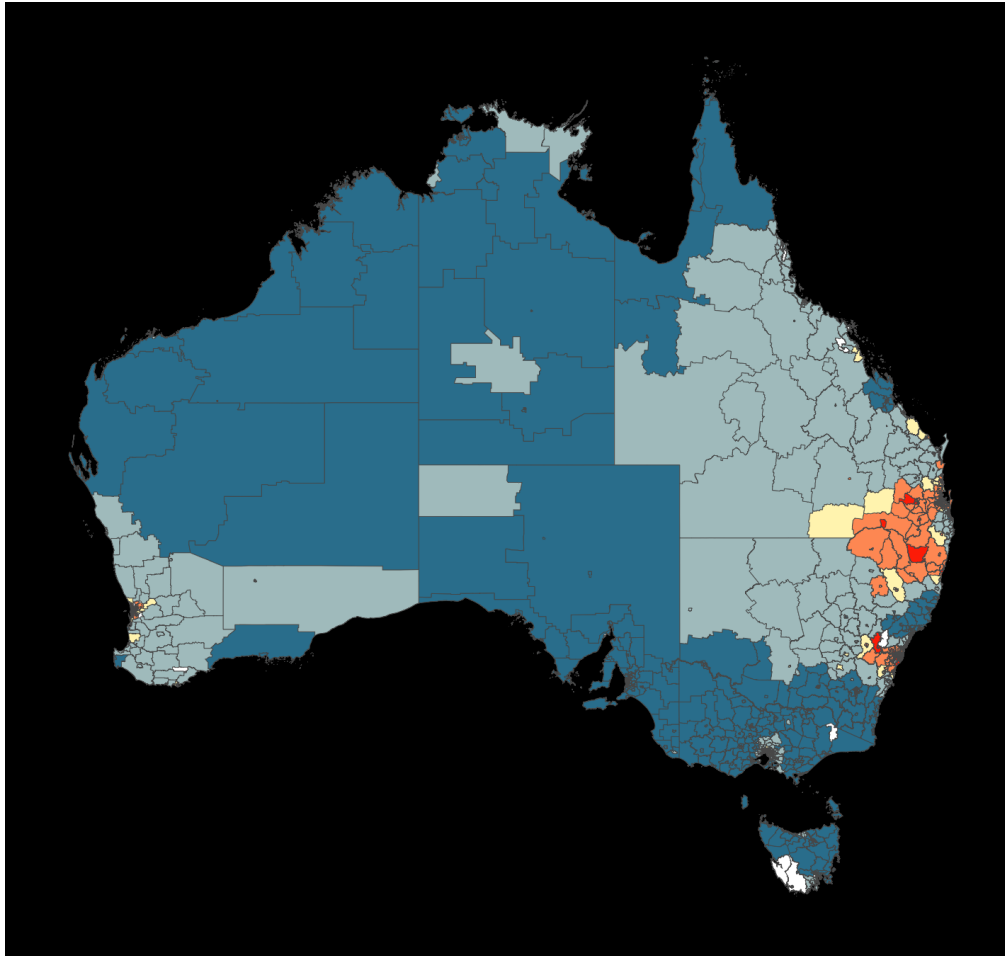


Figure 1: A choropleth map of the Statistical Areas of Australia at Level 2. The colours communicate the value of the estimated SIR of Thyroid cancer for females, ranging from much lower than average (blue) to much higher than average (red)

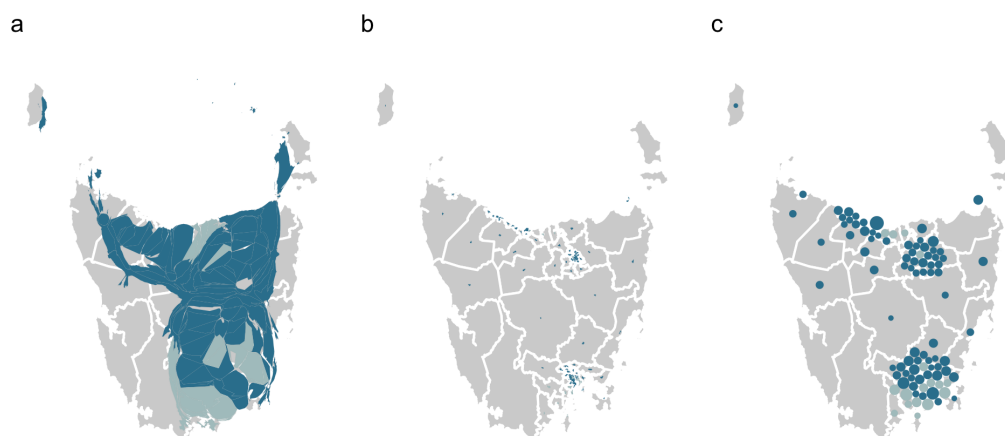


Figure 2: The three displays show alternative maps of the Statistical Areas at Level 2 in the Australian state of Tasmania. These alternatives show the (a) contiguous cartogram, (b) non-contiguous cartogram and (c) Dorling cartogram of Tasmania.

explicit map transformations to improve communication of distributions (Kocmoud and House, 1998). In Figure 2 the choropleth map can be seen underneath each alternative map display to allow for comparisons to be made.

There are several established alternative visualisation methods. Tile maps, Rectangular cartograms (van Kreveld and Speckmann, 2007) and Dorling cartograms (Dorling, 2011) all use one simple shape to represent each geographic unit. They all minimise the emphasis on the size or shape of the geographic areas. These alternative map displays focus on the relationship between neighbours, attempting to preserve connections, and disregard the unique shapes of the administrative boundaries. Figure 2 shows a collection of alternative map displays, this includes a) a contiguous cartogram, b) a non-contiguous cartogram, and c) a Dorling cartogram.

When communicating information that is relevant to the population, each member of the population can be given equal representation by transforming the map (Dorling, 2012). The impact of a disease over connected communities can be seen by allowing the boundaries to maintain connection in the transformed display. The contiguous cartogram displayed in Figure 2a) draws attention to smaller geographic units when they are rescaled according to the population (Walter, 2001). These new shapes can now be coloured to represent a second variable. The algorithm uses the geographic shape of the areas and iterates toward sizing the areas to represent the population. This display can create twisted and unfamiliar shapes from the geographic units as the algorithms must satisfy the topology conditions, especially when there are communities located geographically far from their neighbours (Dorling, 2012).

The non-contiguous cartogram in Figure 2b) also uses the population to rescale the geographic units. Unlike the contiguous cartogram, the SA2 areas maintain their geographic shape, but they may not retain the connection with their neighbours. The population of the SA2 areas is used to scale the geographic units in the non-contiguous cartogram (Olson, 1976). The amount of white space can be meaningful in non-contiguous cartograms (Keim et al., 2002), in this example the disparity between large rural areas and small urban areas means that the city of Hobart may look reasonable but these units are not visible in the context of the Tasmania state map. Depending on the difference between the population and geographic land size, the amount of white space can also prevent meaningful understanding of the distribution (Dorling, 2012).

The Dorling cartogram presents each geographic unit as a circle, the size of the circle is scaled according to the population value of each area (Dorling, 2011). Figure 2c) shows the Tasmania SA2 areas as an individual units located as close as possible to the geographic centroid location. This map draws attention to the collection of coastal cities in Tasmania that were not apparent in Figure 2 a), or b). This display also highlights that the approximation of equal population is not easy to fulfil, as there is some disparity in the sizes of the circles.

Algorithm

The purpose of this algorithm is to create an alternative map display that highlights the spatial distributions for populations. There has been an increasing need for displays that recognise the large number of people that live in dense urban environments. The algorithm intends to maintain the spatial relationships of a group of geographic units using the relationship between each unit and the closest

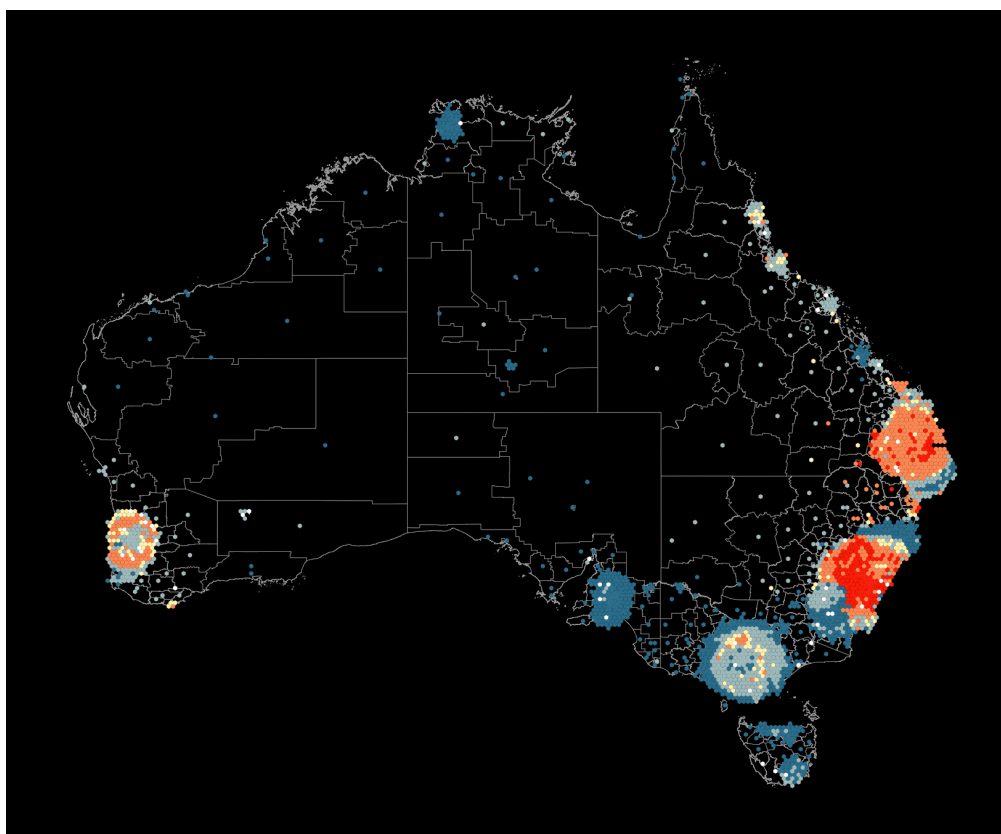


Figure 3: A hexagon tile map of the Statistical Areas of Australia at Level 2. The colours communicate the value of the estimated SIR of thyroid for females, they range from much lower than average (blue) to much higher than average (red).

focal point. The algorithm allocates geographic units to a representative hexagon, in order of their proximity to the closest focal point.

The algorithm is implemented in the *sugarbag* package for R, this is the common name for the *Trigona carbonaria* bee species. Native to northern and eastern Australia this stingless species builds flat layers of hexagonal brood cells, spiralling out from a central point (Vit et al., 2013). This hive design inspired the use of multiple focal points in the algorithm, individual spirals are constructed by building out from various points on the geographic map base.

The result of the algorithm on rates of thyroid cancer for females in each of the SA2 areas of Australia is shown in Figure 3. The difference to the choropleth map in Figure 1 are clear. Now the thyroid cancer incidence in the densely populated areas in Sydney, Brisbane and Perth are visible. Interestingly there is no clear divide between rural and urban SA2 areas as many rural areas, and the cities of Melbourne, Darwin, Adelaide and Hobart have low rates of thyroid cancer diagnosis for females. The hexagon tile map display provides a more accurate representation of the spatial distribution of thyroid cancer incidence across Australia.

When the algorithm is applied to the complete set of all Australian Statistical Areas at Level 2 it results in the hexagon tile map display shown in Figure 3. This display highlights the density of Australian capital cities, as it draws attention to the many communities in Sydney, Melbourne and Hobart. This display also highlights the disparity in the burden of thyroid cancer for females in the communities of these cities. There are several collections of red hexagons in Sydney that represent the communities with much higher rates of diagnosis than the Australian average. Brisbane also experiences higher than average rates of diagnosis, but has more orange than red. The females in the cities of Adelaide and Perth show much lower rates of diagnosis.

Compared to the choropleth map display, the low rates in the rural Australian communities are no longer dominating the display. While the representative hexagons are still visible against the black background, the much lower than average rates in the Northern Territory are less noticeable.

There are several key steps needed to create the hexagon tile map as described in the flow chart in Figure 4. First, derive the set of centroids from the polygons provided, then create the grid of hexagon locations. These two processes are defined in the blue left column of the flow chart in Figure 4. Each

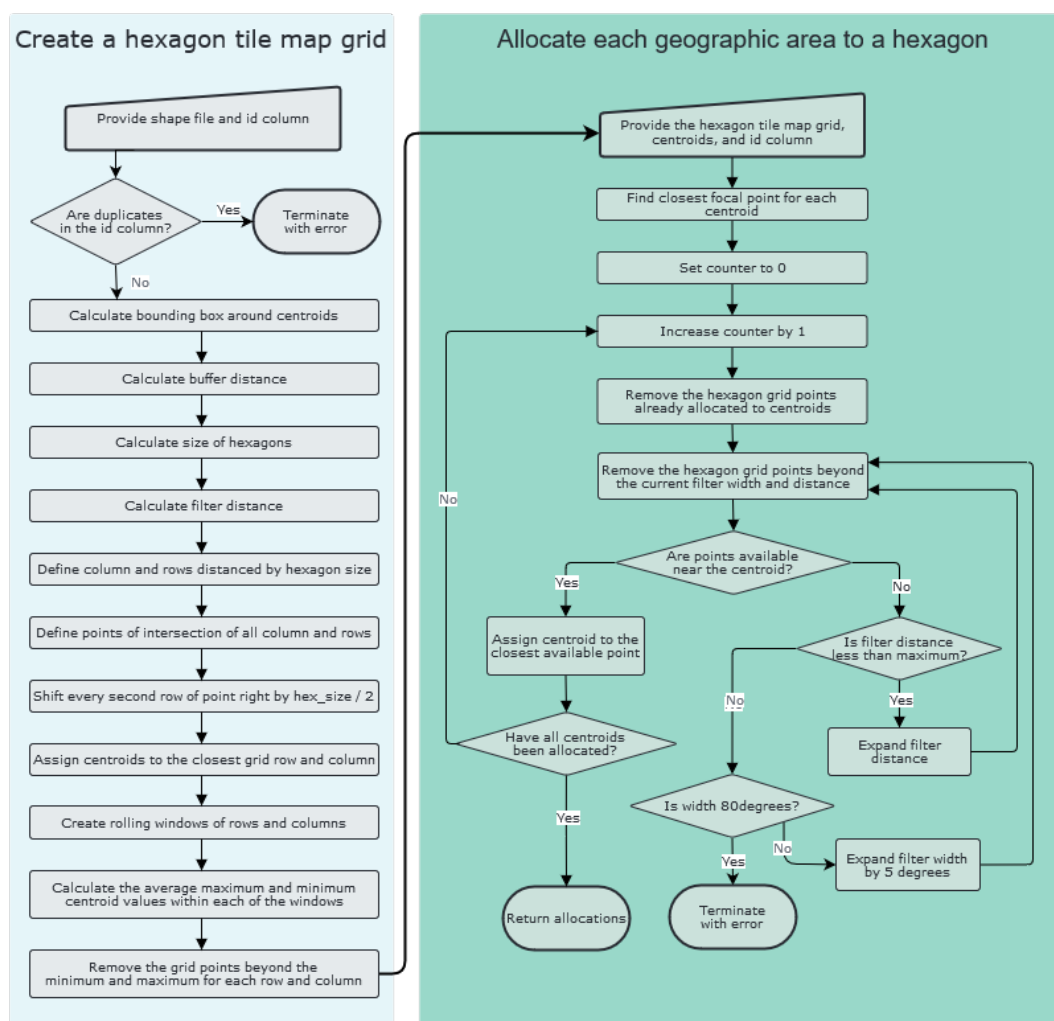


Figure 4: A flow diagram detailing the necessary steps to create a hexagon tile map.

centroid can then be allocated to an available hexagon location. The steps for the allocation process are detailed in the right column of Figure 4. There are several filter steps to speed the process of selecting an appropriate hexagon to represent each geographic unit. To make tessellated plots with the hexagon allocations, the point locations are converted into hexagon shapes.

User choices

Only two inputs are necessary to begin the algorithm; the shape file, and the id variable. The id variable should uniquely identify each geographic unit in the shape file.

The centroids are derived from the shape file. The amount of centroids within the geographic map is used to determine an appropriate hexagon size if one is not provided. The derived centroids are a necessary input for an appropriate grid to be constructed. Each grid point is located a hexagon size distance from the next closest point in all six directions. The grid will initially cover the entire map space, encompassing all the centroid locations and will extend in all directions to the extent of the buffer distance. This buffer distance can be helpful to account for densely populated coastal areas, allowing the use of the sea for hexagon locations.

The centroid set and hexagon tile map grid are necessary for the allocation process. Additionally, a set of reference locations can be provided as focal points, such as the capital cities of states or countries. The algorithm will use the focal points to create the order of allocation, prioritising the closest centroid locations to the focal points. Alternatively, a user can specify the variable that should be used to determine the order for the allocation.

When allocating representative hexagons, the width parameter can be used to determine the flexibility of positioning. Using the relationship with the nearest focal point, a larger width parameter will increase the amount of available hexagons nearer to the original centroid of the geographic unit.

A smaller width will maintain the orientation from the focal point to the centroid when selecting the hexagon location. However, this could mean it is placed further away.

Implementation

Hexagon tile maps can be useful to understand a distribution across a collection of geographic areas. However, these maps are not easy to create manually, especially as the number of areas increases. This algorithm was created to automate this process, and reduce the manual load involved in creating and implementing alternative displays. This allows map makers and data communicators to spend their time to choosing the most effective display.

The `sugarbag` package has been written for R software, it contains a set of functions that help R users to create a hexagon tile map. The algorithm presented in the `sugarbag` package operates on a set of simple feature geometry objects (Pebesma, 2018), this package allows R users to create `sf` objects by importing polygons stored in various formats. Users should provide a set of polygons that define geographic units by their administrative boundaries. The functions arrange the geographic units in order of proximity to a set of locations provided, such as the centre of capital cities. The centroid location of each geographic unit is used to measure the proximity. It emphasises the capital cities as population hubs, rather than emphasizing the size of large, rural geographic units.

A user can tweak the parameters of the hexagon map using the additional arguments to the `create_hexmap` function, these options may affect the speed of the creation. The hexagon size may need adjustments depending on the density of the population, users can provide an appropriate hexagon size and re-run the algorithm. The buffer distance may need to be increased if the coastal cities need to extend beyond the geographic land mass.

Algorithm steps

The package can be installed from CRAN:

```
install.packages("sugarbag")
```

and the development version can be install from the GitHub repository:

```
devtools::install_github("srkobakian", "sugarbag")
```

The following steps create the hexagon tile map for all the Statistical Areas at Level 2 in Tasmania. These steps can be executed by the main function, `create_hexmap`, or can be run separately for more flexibility.

If a user would like to perform each of the steps of the algorithm themselves, the necessary inputs will change for each function. For example, the `allocate` function can be used directly when a user wishes to use a set of centroids, rather than polygons.

The polygon set of Tasmania in 2011 is accessed from the `absmappedata` package. A single column of the data set is used to identify the unique areas. In this case, the unique SA2 names for each SA2 have been used.

The Australian capital cities are used as focal points to allocate each geographic area around the closest capital city. Hobart will be the focal point for this example because only the state of Tasmania is being processed.

The buffer distance, hexagon size, hexagon amount to filter and width of angle are parameters that will be determined within `create_hexmap()`, if they are not provided. They are created as they are needed throughout the following example.

The results from various steps in the process are illustrated in Figure 5.

Step 1: Derive the set of centroid points

The set of polygons should be provided as an `sf` object, this is a data frame containing a geometry column. The `read_shape()` function can assist in creating this object for use in R.

The centroids can be derived from an `sf` object using the `create_centroids()` function:

Step 2: Create the hexagon grid points

A grid is created to allow tessellation of the hexagons that represent the geographic units. For a hexagon tile map, the grid of possible hexagon locations is made using the `create_grid()` function. It

uses the centroids, the hexagon size and the buffer distance. Figure 5a) shows the initial locations of grid points created for Tasmania.

Creating a tessellated grid A set of longitude columns, and latitude rows are created to define the locations of the hexagons. The distance between each row and column is the size specified by `hex_size`. Equally spaced columns are created from the minimum longitude minus the buffer distance, up to the maximum longitude plus the buffer distance. Similarly, the rows are created from the latitude values and the buffer distance. A unique hexagon location is created from all intersections of the longitude columns and latitude rows. Figure 5a shows the hexagon grid after every second latitude row on the grid is shifted right, by half of the hexagon size.

Filtering the grid Not all of the grid points will be used, especially if islands result in a large grid space. To filter the grid for appropriate hexagon locations for allocation, the `create_grid` function will call the `create_buffer` function. It finds the grid points needed to best capture the set of centroids on a hexagon tile map. Reducing the buffer can help to decrease the amount of time taken, as there will be less points in the water to consider for each centroid allocation.

The closest latitude row and longitude column are found for each centroid location. Then rows and columns of centroids are divided into 20 groups. The amount of rows in each latitude group and the amount of columns in each longitude group are used as the width of rolling windows. The first rolling window step finds the minimum and maximum centroid values within each of the sliding window groups of longitude columns, and the groups of latitude rows. The second rolling window step finds the average of the rolling minimum and maximum centroid values, for the longitude columns and latitude rows.

The hexagon grid points are kept only if they fall within the rolling average of the minimum and maximum centroid values after accounting for the buffer distance, for each row and column of the grid. Figure 5b displays remaining hexagon grid points after applying the buffer filter. The sparsely populated South-West region of National Park has fewer points available in the water compared to the South-East region near the city of Hobart.

Centroid to focal point distance The distance is calculated between each centroid in the set, and each of the focal points provided. The order for allocation is determined by the distance between the polygon centroid and it's closest focal point. For this example, this distance is only calculated for the capital city of Hobart, represented in Figure 5c and d as a white cross.

Step 3: Allocate each centroid to a hexagon grid point

Allocation of all centroids requires the set of polygon centroids and the hexagon map grid. The polygon centroids are ordered from the centroid closest to the focal point(s), to the furthest. This will preserve spatial relationships with the focal point, as the inner city areas are allocated first and placed closest to the capital, the areas that are further will then be accommodated. The following example considers the first of the Statistical Areas at Level 2. Within the algorithm, these steps are repeated for each polygon.

Filter the grid for unassigned hexagon points The set of possible hexagon grid points reduces by one after each centroid is allocated, then only the grid points that have not yet been selected are considered. Keeping only the available hexagon points prevents multiple geographic units from being allocated to the same hexagon. This is demonstrated in Figure 5c and d by the black hexagons that represent the seven closest polygons to the capital city of Hobart. As the allocation process begins for the eighth closest centroid there are seven unavailable hexagon locations.

Filter the grid points for those closest to the centroid A selection of possible hexagon locations around the centroid allows only the closest points that are not yet assigned to be considered. The number of possible hexagon locations to consider for a centroid is determined by the `hex_filter`. This is the maximum amount of hexagons between the centroid and the furthest considered hexagon. It is used to subset possible grid points to only those surrounding the polygon centroid within an appropriate range. A smaller distance will increase speed, but can decrease accuracy when width of the angle increases.

The algorithm creates a circle of points, by only keeping points within a certain radial distance around the original centroid location.

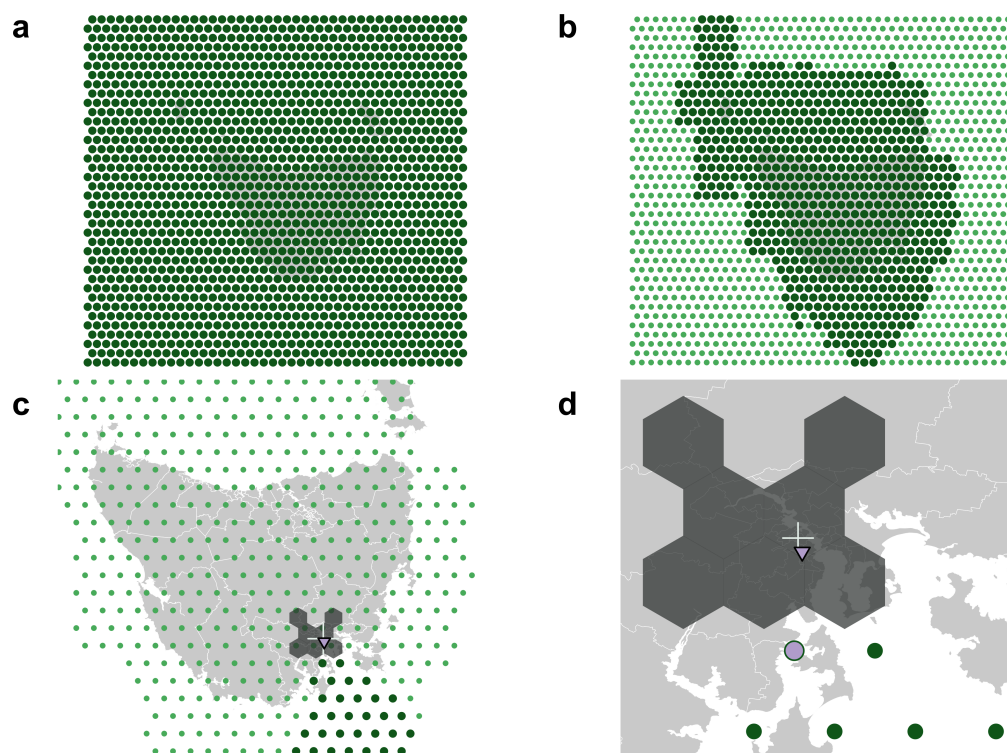


Figure 5: Filter for grid points within a square, then circular, distance for those closest to the centroid.

The width parameter is used to take a slice of the remaining points. The width is the amount of degrees used on either side of the angle from the focal point to centroid location. This uses the angle from the closest capital city, to the current centroid as seen in Figure 5c. This allows the spatial relationship to be preserved, even when it is allocated to a hexagon that is further from the focal point than the original centroid location.

If no available hexagon grid point is found within the original filter distance and angle, the distance is expanded, only when a maximum distance is reached will the angle expand to accommodate more possible grid points. By default the angle filter for the hexagon grid points create plus and minus 30 degrees bounds of the angle from the focal point to the geographic centroid. This will increase if no points can be found within the `hex_filter` distance. The default angle of 30 was chosen to allow the algorithm to find hexagons that best maintained the spatial relationship between the focal point and geographic centroid.

Animation

Creating an animation connecting these two displays can allow users to grasp the meaning of the connected hexagons. It will also highlight the density of the communities using the rapid expansion of the inner-city areas, these hexagons will move the furthest and will move rapidly in the animation from their geographic locations. The rapid decrease in size of the large rural areas can show how large the landmass of Statistical Areas can be, and how few people occupy rural areas. The `gganimate` (Pedersen and Robinson, 2019) package can be used to make an animation. It connects the polygons for each area in the two displays using the `sf_id` variable, the names of the statistical areas, and creates intermediate displays as they geographic areas shrink or expand to the chosen hexagon size.

Conclusion

A choropleth map display is the current common practice for spatial data visualisation. As the choropleth map display relies on the geographic shapes, it does not effectively present the density of the population in small geographic areas. The common solution to enhance the communication of spatial distributions is to incorporate the population values as a cartogram or another alternative map display. Many alternative map displays do not work well for Australia due to the large rural geographic space between the densely populated capital cities. The administrative boundaries may also distract from the statistics communicated using colour due to the small size of some areas. This

paper provides an algorithm for Australia called the hexagon tile map.

Alternative maps highlight the value of the statistics across the geographic units. These methods allow increased understanding of the spatial distribution of a variable across the population, by fairly representing each administrative area. This acknowledges that the amount of residents can be different but recognises that each population area is equally important. The solution to this visualisation problem can be using equally sized areas, with neighbourhood boundary connections. The *sugarbag* package for R creates tessellated hexagon tile maps by automating an algorithm to design these displays. The Australian application preserves the spatial relationships, emphasising capital cities.

These hexagons equally represent each area. However, similar to the choropleth map display the tessellation does not allow the size of the hexagons to represent another variable. The algorithm is heavily dependent on the focal points used, as this determines the order of allocation. It works on the assumption that viewers can identify their neighbourhoods using the relative directional relationships but this can be aided by the use of animation. With careful consideration of the choropleth map, the small geographic inner city areas may have been noticed by viewers, but the hexagon tile map display emphasises them. The communities in northern Tasmania and the Northern Territory do not draw attention because of their size as in the choropleth, but their colour is still noticeably below average when contrasted with the hexagons in New South Wales.

This algorithm should be general enough to apply to other countries and problems

This algorithm has only been tested using single countries, and does not consider definite borders of countries. While the buffer allows extension beyond the furthest centroids, there is no mechanism to protect the borders and ensure centroids are placed within the geographic borders of a country.

It would be also possible to take a logarithmic function in the filter steps for selection of a hexagon grid point. A direct angle is currently used, but this may narrowly miss appropriate hexagons, a logarithmic function may help to choose a closer hexagon to the original centroid location, before increasing the width of the angle used to filter the hexagons. This algorithm is an effective start to creating hexagon tile maps for many geographic units.

Bibliography

- Australian Bureau of Statistics. Statistical Area Level 2 (SA2) ASGS Ed 2011 Digital Boundaries in ESRI Shapefile Format, 2011. [p1]
- Cancer Council Queensland, Queensland University of Technology, and Cooperative Research Centre for Spatial Information. Australian Cancer Atlas, 2018. URL <https://atlas.cancer.org.au>. [p1]
- D. Dorling. *Area Cartograms: Their Use and Creation*, volume 59, pages 252 – 260. 04 2011. ISBN 9780470979587. doi: 10.1002/9780470979587.ch33. [p1, 3]
- D. Dorling. *The Visualisation of Spatial Social Structure*. John Wiley and Sons Ltd, 2012. [p1, 3]
- J. A. Dougenik, N. R. Chrisman, and D. R. Niemeyer. An algorithm to construct continuous area cartograms. *The Professional Geographer*, 37(1):75–81, 1985. doi: 10.1111/j.0033-0124.1985.00075.x. URL <https://doi.org/10.1111/j.0033-0124.1985.00075.x>. [p1]
- D. Keim, S. North, C. Panse, and J. Schneidewind. Efficient Cartogram Generation: A Comparison. In *IEEE Symposium on Information Visualization*, 2002. INFOVIS 2002, volume 2002, pages 33–36. IEEE, 2002. ISBN 076951751X. [p3]
- C. Kocmoud and D. House. A Constraint-based Approach to Constructing Continuous Cartograms. In *Proc. Symp. Spatial Data Handling*, pages 236–246, 1998. [p3]
- M. E. Levison and W. Haddon Jr. The Area Adjusted Map. An Epidemiologic Device. *Public Health Reports*, 80:55–59, 1965. ISSN 00946214. [p1]
- D. A. Moore and T. E. Carpenter. Spatial Analytical Methods and Geographic Information Systems: Use in Health Research and Epidemiology. *Epidemiologic Reviews*, 21(2):143–161, 07 1999. ISSN 1478-6729. doi: 10.1093/oxfordjournals.epirev.a017993. URL <https://doi.org/10.1093/oxfordjournals.epirev.a017993>. [p1]
- J. M. Olson. Noncontiguous Area Cartograms. *The Professional Geographer*, 28(4):371–380, 1976. doi: 10.1111/j.0033-0124.1976.00371.x. URL <https://doi.org/10.1111/j.0033-0124.1976.00371.x>. [p3]
- E. Pebesma. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1): 439–446, 2018. doi: 10.32614/RJ-2018-009. URL <https://doi.org/10.32614/RJ-2018-009>. [p6]

- T. L. Pedersen and D. Robinson. *gganimate: A Grammar of Animated Graphics*, 2019. URL <https://CRAN.R-project.org/package=gganimate>. R package version 1.0.3. [p8]
- A. Skowronnek. Beyond Choropleth Maps – A Review of Techniques to Visualize Quantitative Areal Geodata, 2016. URL https://alsino.io/static/papers/BeyondChoropleths_AlsinoSkowronnek.pdf. [p1]
- E. R. Tufte. *Envisioning Information*. Graphics Press, 1990. [p1]
- M. van Kreveld and B. Speckmann. On rectangular cartograms. *Computational Geometry*, 37(3):175 – 187, 2007. ISSN 0925-7721. doi: 10.1016/j.comgeo.2006.06.002. URL <http://www.sciencedirect.com/science/article/pii/S0925772106000770>. Special Issue on the 20th European Workshop on Computational Geometry. [p3]
- P. Vit, S. R. M. Pedro, and D. Roubik. *Pot-Honey: A legacy of stingless bees*. Springer-Verlag New York, 2013. ISBN 978-1-4614-4960-7. URL <http://dx.doi.org/10.1007/978-1-4614-4960-7>. [p4]
- S. D. Walter. *Disease Mapping: A Historical Perspective*. Oxford University Press, 2001. ISBN 9780191723667. doi: <https://dx.doi.org/10.1093/acprof:oso/9780198515326.003.0012>. [p3]

Stephanie Kobakian
Monash University
Department of Econometrics and Business Statistics

stephanie.kobakian@monash.edu

Dianne Cook
Monash University
Department of Econometrics and Business Statistics

dicook@monash.edu

Earl Duncan
Queensland University of Technology
School of Mathematical Sciences

earl.duncan@qut.edu.au