Aust. N. Z. J. Stat. 2026 doi: 10.1111/j.1467-842X.XXX

Comparing the Effectiveness of the Choropleth Map with a Hexagon Tile Map for Communicating Patterns in Australian Spatial Statistics

Stephanie Kobakian[1] and Dianne Cook[2]

Queensland University of Technology and Monash University

Summary

The choropleth map is a common tool for communicating spatial distributions across geographic areas. However, the size of geographic units can distort interpretation, influencing how users perceive the distribution. A common alternative is the cartogram, which resizes areas based on population. Yet, in Australia, the stark disparities in geography and population make cartograms less suitable. This study explores the hexagon tile map as an alternative. We report results from a task-based experiment involving human participants, using the lineup protocol to assess how well hexagon tile maps and choropleths convey spatial patterns. Three spatial patterns were tested: one reflecting geography, with values increasing monotonically from the northwest to southeast of Australia, and two with clustered high concentrations. Results show that the hexagon tile map outperforms the choropleth map. These findings support the use of alternative map displays and suggest that hexagon tile maps are effective for visualising spatial distributions in heterogeneous regions.

Key words: data visualisation; visual inference; geospatial; statistical graphics; designed experiment

1. Introduction

This study compares the effectiveness of the spatial display, a hexagon tile map, against the standard, a choropleth map, for communicating information about disease statistics. The choropleth map is the traditional method for visualizing aggregated statistics across administrative boundaries. It works better for countries that have administrative areas that are relatively equally sized spatially, which is far from the

[1] Science and Engineering Faculty, Queensland University of Technology, 2 George St, Brisbane, Australia
[2] Econometrics and Business Statistics Faculty, Monash University, 29 Ancora Imparo Way, Clayton, VIC 3800, Australia
Email:

situation in Australia. The hexagon tile map builds on existing displays, such as the cartogram, and tessellated hexagon displays. A hexagon tile map forgoes the familiar boundaries, in favour of representing each geographic unit as an equally sized hexagon, placed approximately in the correct spatial location. It differs in the relaxed requirement to have connected hexagons, and allows sparsely located hexagons. This type of display may be generally useful for displaying government statistics spatially, and other spatial display purposes. The algorithm to construct a hexagon tile map is available in the R package sugarbag (Kobakian, Cook & Duncan 2023).

The hexagon tile map was designed for Australia, motivated by a need to display spatial statistics for the Australian Cancer Atlas. None of the existing approaches for creating cartograms or hexagon tiling perform well for the Australian landscape, which has vast open spaces and concentrations of population in small regions clustered on the coastlines.

The Australian Cancer Atlas (Cancer Council Queensland and Queensland University of Technology 2024) is an online interactive web tool created to explore the burden of cancer on Australian communities. There are many cancer types to be explored individually or aggregated. The Australian Cancer Atlas allows users to explore the patterns in the distributions of cancer statistics over the geographic space of Australia. It uses a choropleth map display and diverging colour scheme to draw attention to relationships between neighbouring areas. The hexagon tile map may be a useful alternative display to enhance the atlas.

The experiment was conducted using the lineup protocol, a visual inference procedure (Buja et al. 2009; Wickham et al. 2010), that can be used to objectively test the effectiveness of the two displays (Hofmann et al. 2012). A lineup embeds the data plot among a field of null data plots, and an independent observer is asked to select the most different plot. It was shown by Majumder, Hofmann & Cook (2013) to be an effective way to conduct a hypothesis test where a plot is treated as a test statistic, and utilised for this purpose in numerous studies (Fieberg, Freeman & Signer 2024; Green 2021; Li et al. 2024). If the data plot is selected it is analogous to a rejection of the null hypothesis (specifying no structure), and the observed data plot is unlikely to arise from the null scenario. The work of VanderPlas & Hofmann (2016) compares the lineup protocol to performance on standard tests of visual ability, concluding that participants' performance is related to general visual aptitude, to classification rather than spatial reasoning.

The paper is organised as follows. The next section discusses the background of geographic data display and visual inference procedures. Section 3 describes the methods for conducting the experiment and analysing the results. The results are summarised in the Section 4, followed by a discussion about the broader implications for the use of this map style.

2. Background

2.1. Spatial data displays

Spatial visualisations communicate the distribution of statistics over geographic landscapes. The choropleth map (Tufte 1990; Skowronnek 2016) is a traditional display. It is used to present statistics that have been aggregated on geographic units. Creating a choropleth map involves drawing polygons representing the administrative boundaries, and filling with colour mapped to the value of the statistic. The choropleth map places the statistic in the context of the spatial domain, so that the reader can see whether there are spatial trends, clusters or anomalies. This is important for digesting disease patterns. If there is a linear trend it may imply a relationship between disease and geographic location. If there is a cluster, or an anomaly, there may be a localized outbreak of the disease.

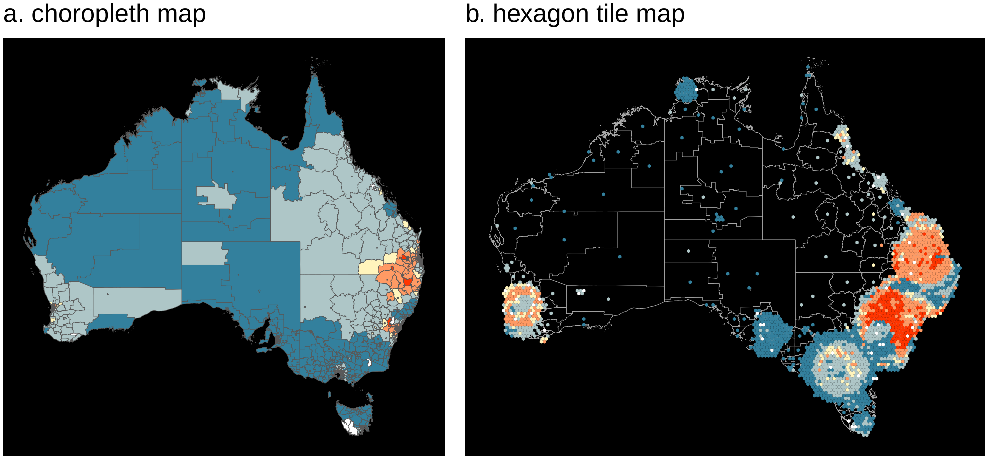

Figure 1. Thyroid cancer incidence among females across the Statistical Areas of Australia at Level 2, displayed using a choropleth map (a) and a hexagon tile map (b). Blue indicates lower than average, and red indicates higher than average incidence. The choropleth map suggests high incidence is clustered on the east coast but misses the high incidence in Perth and a few locations in inner Melbourne visible in the hexagon tile map.

The choropleth map is an effective spatial display if the size of the geographic units is relatively uniform. This is not the case for most countries. Size heterogeneity in administrative units is particularly extreme in Australia: most of the landscape of Australia is sparsely settled, with the population densely clustered into the narrow coastal strips. Figure 1 shows the choropleth map of thyroid cancer incidence rates in Australia. The choropleth map focuses attention on the geography, and for heterogeneously sized areas it presents a biased view of the population related distribution of the statistic (House & Kocmoud 1998). *Land does not get cancer, people do* – a more effective way to communicate the spatial distributions of cancer statistics is needed (sentiment motivated by Monmonier (2018)).

A cartogram is a general solution for adequately displaying a population-based statistic. It transforms the geographic map base to reflect the population in the geographic region, while preserving some aspects of the geographic location. There are several cartogram algorithms (Dorling 2011; House & Kocmoud 1998); each involves shifting the boundaries of geographic units, using the value of the statistic to increase or decrease the area taken by the geographic unit on the map. The changes to the boundaries result in cartograms that accurately communicate population by map area for each of the geographic units but can result in losing the familiar geographic information. For Australia, the transformations warp the country so that it is no longer recognisable (see Kobakian, Cook & Roberts (2020) for details).

Alternative algorithms make various trade-offs between familiar shapes and representation of geographic units. The non-contiguous cartogram method (Olson 1976) keeps the shapes of geographic units intact, and changes the size of the shape. This method disconnects areas creating empty space on the display losing the continuity of the spatial display of the statistic. The Dorling cartogram (Dorling 2011) represents each unit as a circle, sized according to the value of the statistic. The neighbour relationships are mostly maintained by how the circles touch. A similar approach was pioneered by Raisz (1963), using rectangles that tile to align borders of neighbours (Monmonier 2005). There have been thorough reviews of the array of methods, as suitable for cancer atlas displays (e.g. Kobakian, Cook & Roberts 2020; Skowronnek 2016), and experiments demonstrating cartograms to be more effective than choropleth maps (Kaspar, Fabrikant & Freckmann 2011).

The hexagon tile map algorithm, automatically matches spatial regions to their nearest hexagon tile, from a grid of tiles. It has the effect of spreading out the inner city areas while maintaining the spatial locations or regions in remote areas. The algorithm is available in the R package, sugarbag (Kobakian, Cook & Duncan 2023). Figure 1 shows the hexagon tile map, where the map is coloured from low incidence (blue) to high (red). The inner city areas have expanded, making it possible to see the cancer incidence in the small, densely populated areas. Remote regions are represented by isolated hexagons, which is not ideal, but maintains the spatial location of these data values. It is of interest to know how well the spatial distribution patterns are seen from this display, in comparison to how they are seen from the choropleth map.

Hexagon displays are growing in popularity. Two media outlets used variations of the hexagon displays to communicate the 2025 Australian federal election results (Green 2025; Evershed & Ball 2025). Both are effective, but have inadequacies. The Guardian preserves geography and allows inner city results to be seen but the overall sense of the result is skewed because the large rural areas dominate the display. The ABC's contiguous hexagon tile map gives the correct sense of the final results but loses the shape of Australia, and some hexagons are far from their true location.

2.2. Visual Inference

In order to assess the effectiveness of the hexagon tile map, the lineup protocol (Buja et al. 2009; Wickham et al. 2010) from visual inference procedures is employed. The approach mirrors classical statistical inference. The procedures for doing a power comparison of competing plot design, outlined in Hofmann et al. (2012), are followed. It is the only current human subjects testing protocol which quantitatively compares plot designs on the basis of detection of structure relative to null distributions (VanderPlas, Cook & Hofmann 2020; VanderPlas 2021). The premise for comparing two designs is that the only difference between the two lineups is the plot design, and hence difference in detection rate and time to detect is due to the effectiveness of the plot design for differentiating between data plot and null plots. The protocol has been used to quantitatively test plot designs numerous studies (e.g. Loy, Hofmann & Cook 2017; VanderPlas & Hofmann 2016; Kossmeier, Tran & Voracek 2019; Reda & Szafir 2021).

In classical statistical inference hypothesis testing is conducted by comparing the value of a test statistic on a standard reference distribution, computed assuming the null hypothesis is true. If the value is extreme, the null hypothesis is rejected, because the test statistic value is unlikely to have been so extreme if it was true. In the lineup protocol, the plot plays the role of the test statistic, and the data plot is embedded in a field of null plots. Defining the plot using a grammar of graphics (Wickham 2009) makes it a functional mapping of the variables and thus, it can be considered to be a statistic. With the same data, two different plots can be considered to be competing statistics, one possibly a more powerful statistic than the other.

Hypothesis testing with the lineup protocol requires human evaluation. The human judge is required to identify the most different plot among the field of plots. If this corresponds to the data plot – the test statistic – the null hypothesis is rejected. It means that the data plot is extreme relative to the reference distribution of null plots.

The null hypothesis is explicitly provided by the grammatical plot description. For example, if a histogram is the map type being used, the null might be that the underlying distribution of the data is a Gaussian. Null data would be generated by simulating from a normal model, with the same mean and standard deviation as the data. In practice, the null hypothesis used is generic, such as *there is NO structure or a pattern in the plot*, and contrasted to an alternative that there is structure.

The chance that an observer picks the data plot out of a lineup of size m plots accidentally, if the null hypothesis is true is $1/m$. With K observers, the probability of k randomly choosing the data plot, roughly follows a binomial distribution with $p = 1/m$. Figure 2 shows a lineup of the hexagon tile map, of size $m = 12$. Plot 3 is the data plot, and the remaining 11 are plots of null data. The supplementary materials contain all lineups used in the study, including the corresponding lineup to this one made using choropleth maps.

In order to determine the effectiveness of a type of display, this probability is less relevant than the overall proportion of observers who pick the data plot, k/K. The power of the test statistic (data plot) is provided by this proportion. Power in a statistical sense is the ability of the statistic to *produce a rejection* of the null hypothesis, if it is indeed *not true*. With the same data plotted using two different displays, the display with the highest proportion of people who choose the data plot would be considered to be the most powerful statistic.

There are several practical considerations when deploying the lineup protocol: (1) determining the appropriate null distribution to compute null samples, (2) employing independent observers to conduct the evaluations, (3) varying location of the data plot in a lineup, (4) how many null sets to include in the lineup, (5) construction of all lineups in the experiment (see VanderPlas et al. (2021)), (6) questions presented to participants to solicit evaluations, in addition to the usual experimental design issues. These are described in detail in the next section.

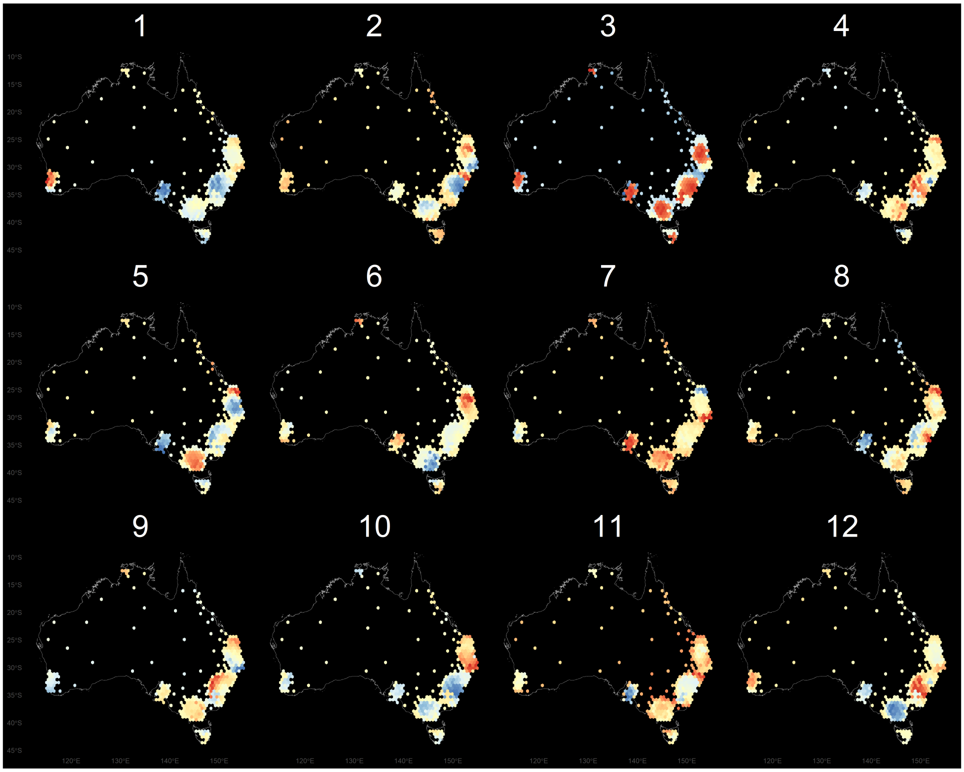

Figure 2. This lineup of twelve hexagon tile map displays contains one map with a real population related structure (location 3). The rest are null plots that contain only spatial dependence.

3. Methodology

This study aims to answer two key questions around the presentation of spatial distributions:

1. Are spatial disease trends that impact highly populated small areas detected with higher accuracy, when viewed in a hexagon tile map?
2. Are people faster in detecting spatial disease trends that impact highly populated small areas when using a hexagon tile map?

Additional considerations when completing this experimental task included the difficulty experienced by participants and the certainty they had in their decision.

Australia is used for the study, with Statistical Area 3 (SA3) (Australian Bureau of Statistics 2018) as the geographic units. The results should apply broadly to any other geographic areas of interest, if there are large differences in area and population size.

3.1. Experimental factors

The primary factor in the experiment is the map type. The secondary factor is a trend model. Three trend models were developed: one mirroring a large spatial trend for which the choropleth map would be expected to do well, and two with differing levels of inner city hot spots. These latter two reflect the structure seen in the thyroid cancer incidence data (Figure 1). This produces six treatment levels:

- Map type: *Choropleth, Hexagon tile*
- Trend:
  - *NW-SE*: Large spatial trend running diagonally across Australia
  - *Three Cities*: Locations in three population centres
  - *All Cities*: Locations in all state and territory capitals

Data is generated for each of the trend models, with four replicates, and each displayed both as a choropleth map and as a hexagon tile map, which yields 12 data sets, and 24 data plots. This set of displays is divided in half, providing two sets of 12 displays, Group A and Group B. Participants were randomly allocated to Group A or B. Participants saw a data set only once, either as a choropleth map or as a hexagon tile map. Figure 3 summarises the design and the allocation of the displays.

3.2. Generating null data

Null data needs to be data with no (interesting) structure. In most scenarios, permutation is the main approach for generating null plots. It is used to break association between variables, while maintaining marginal distributions. This is too simple for spatial data. In spatial data, a key feature is the spatial dependence or smoothness over the landscape. To do something simple, like permute the values relative to the geographic location would produce null plots which are too chaotic, and the data plot will be recognisable for its smoothness rather than any structure of interest.

For spatial data, null data is stationary data, where the mean, variance and spatial dependence are constant over the geographic units. Stationary data is specified by a variogram model (Matheron 1963). Simulating from a variogram model, where the spatial dependence is specified, generates the stationary spatial data used for the null plots. The parameters for the Gaussian model were sill=1, range=0.3 with the variance generated by a standard normal distribution.

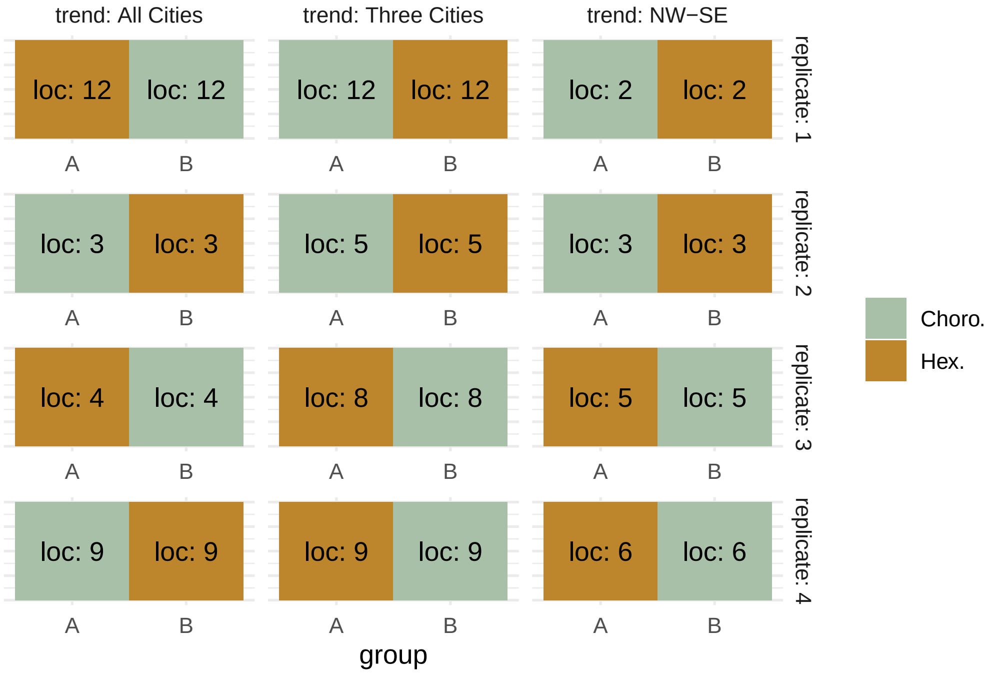

Figure 3. The experimental design used in the study. Participants are allocated to group A or B, to evaluate either the choropleth map or hexagon tile map lineup of each simulated data set. The text 'loc' refers to the location of the data plot in the lineup.

The R package gstat (Gräler, Pebesma & Heuvelink 2016) was used to simulate 144 null sets, 12 data sets for each plot in a lineup, and 12 sets for 12 lineups. Simulating spatial dependence is difficult as discussed by Beecham et al. (2017) as a blockage for using the lineup protocol for testing map displays. Because the result from gstat was inadequate to mirror the spatial dependence patterns results in the Cancer Council Queensland and Queensland University of Technology (2024), each null set was further smoothed. This was done by averaging a small number of spatial neighbours, approximating the methods described in Duncan et al. (2019).

3.3. Generating lineups

For each trend model, four real data displays were created by manipulating the centroid values of each of the SA3 geographic units. Each trend model is motivated by patterns observed in spatial data: North West to South East (NW-SE) is a basic spatial trend across the entire country, Three Cities is the existence of clusters of high values, and All Cities is also clusters, but more of them. We would expect that clusters pattern to be more visible with a hexagon tile map but the large spatial trend to be more visible in the choropleth map.

The NW-SE distribution was created using a linear equation of the centroid longitude and latitude values. The All Cities trend model was created using the distance from the centroid of each geographic unit to the closest capital city in Australia, calculated when creating the hexagon tile map produced by the sugarbag (Kobakian, Cook & Duncan 2023) package. Two-thirds of SA3s (201/336) were considered greater capital city areas, the values of these areas were increased to create red clusters. The amount was chosen to make clusters around the cities visible, even in the choropleth map with careful inspection. A similar selection process was applied to the Three Cities' trend model. However, for each of the four replicates for the Three Cities trend, a random sample of capital cities was taken from Sydney, Brisbane, Melbourne, Adelaide, Perth, and Hobart. Only values of the areas nearest to the three cities were increased to create clusters.

One of the plot locations (1-20) is chosen to embed the data plot, in each of the four replicates, for the three trend models. These locations were chosen by random sampling. Using random locations reduces the chance that participants might deduce the location coincidentally. Locations 1, 7, 10 and 11 were not in the sample. Yin et al. (2013) used the lineup protocol for a genomics study and similarly varied the the location of the data plot in replicates of treatments. Their results demonstrated that the actual location of the data plot didn't affect performance. So we don't expect that the location affects results but randomising locations among different lineups is to guard against participants expecting the data plot to be in a particular location.

The lineup locations were the same for both map types, because each set of lineup data was used to produce a choropleth map lineup and hexagon tile map lineup. This ensures that performance on the two map types can be directly compared. Lineups were grouped into A or B, so that a participant saw only one version. Participants were assigned to group A or B, randomly, and thus evaluated either the choropleth map or the hexagon tile map lineup. Because there were four replicates of each lineup, each participant evaluated two choropleth map and two hexagon tile map lineups, for each trend model. This design is illustrated in Figure 3.

For each of the 144 individual maps, the values for each geographic area were rescaled to create a similar colour scale from deep blue to dark red within each map. This meant at least one geographic unit was coloured dark blue, and at least one was red, in every map display of every lineup.

For the geographic NW-SE distribution, this resulted in the smallest values of the trend model (blue) occurring in Western Australia, the North West of Australia, and the largest values of the trend model (red) occurring in the South East. This resulted in Tasmania being coloured completely red. For the other two trend types, clusters localised in the cities appeared more red than the rest of Australia.

3.4. Web application to collect responses

The taipan (Kobakian & O'Hara-Wild 2018) package for R was used to create the survey web application. This structure was altered to collect responses regarding participants' demographics and their survey responses. The survey app contained three tabs. Participants were first asked for their demographics, their unique identifier and their consent to the responses being used for analysis. The demographics collected included participants' preferred pronoun, the highest level of education achieved, their age range and whether they had lived in Australia.

After submitting these responses, the survey application switched to the tab of lineups and associated questions. This allowed participants to easily move through the twelve displays and provide their choice, reason for their choice, and level of certainty.

When participants completed the twelve evaluations the survey application triggered a data analysis script. This created a data set with one row per evaluation. Containing the responses to the three questions. The script also added the title of the image, which indicated the type of map display, the type of distribution hidden in the lineup, and the location of the data plot. It also calculated the time taken by participant to view each lineup.

Each participant used the internet to access the survey, and data was transferred by secure link from the web app to a Google sheet using the `googlesheets` package tools (Bryan & Zhao 2018).

3.5. Participants

Participants were recruited from the Figure Eight crowd-sourcing platform (Figure Eight Inc 2019) to evaluate lineups. The lineup protocol expects that the participants are uninvolved judges with no prior knowledge of the data, to avoid inadvertently affecting results. Potential participants needed to have achieved level 2 or level 3 from prior work on the platform, ensuring only participants with a good record on prior tasks could provide evaluations. All participants were at least 18 years old.

Participants were allocated to either group A or group B when they proceeded to the survey web application. There were 92 participants involved in the study. All participants read introductory materials, and were provided with some training using using three simple lineups, to orient them to the evaluation task. All participants who completed the task were compensated $AUD5 for their time, via the Figure Eight payment system.

A pilot study was conducted in the working group of the Econometrics and Business Statistics Department of Monash University. This allowed us to estimate the effect size, and thus decide on number of participants to collect responses from.

3.6. Data collection

Each participant answered demographic questions and provided consent before evaluating the lineups.

Demographics were collected regarding the study participants:

- Gender (female / male / other),
- Education level achieved (high school / bachelors / masters / doctorate / other),
- Age range (18-24 / 25-34 / 35-44 / 45-54 / 55+ / other)
- Lived at least for one year in Australia (Yes / No)

Participants then moved to the evaluation phase. The set of images differed for Group A and Group B. After being allocated to a group, each individual was shown the 12 lineups in randomised order, and asked to report their responses to these three items:

- **Plot choice**: the number of the plot that they deemed to be most different from the others.
- **Reason**: one of “Clusters of colour”, “Colour trend across the areas”, “Big differences between neighbouring areas”, “All areas have similar colours” or “None of these reasons”. Note providing restricted list of reasons rather than free text encourages a response because it is easy. The list needs to contain the primary expected reasons and other potential reasons need to be added so that it does not bias the participants’ behaviour.
- **Certainty**: how certain that their choice is different from the others, on a scale of 1-5.

3.7. Analysis

3.7.1. Data Cleaning

Data is checked to ensure that survey responses collected for each participants were only included once. Technically it is possible to submit results more than once if the submit button is clicked multiple times in short sequence. Participants who did not finish the evaluation of all lineups or clicked through without providing their evaluation are removed.

3.7.2. Descriptive statistics

Basic descriptive statistics were computed for the different experimental treatments. Basic plots summarising detection rates by map type and trend model type, and feedback and demographic variables against the different experimental design elements are provided.

3.7.3. Modelling

The likelihood of detecting the data plot in the lineup can be modelled using a linear mixed effects model. The R `glmer()` function in the `lme4` (Bates et al. 2015) package implements generalised linear mixed effect models. The model used includes the two main effects map type and trend model, which gives the fixed effects model to be:

$$\hat{y}_{ijk} \sim Bernoulli(p_{ijk})$$

with

$$\text{logit}(p_{ijk}) = \mu_i + \tau_j + \delta_k + (\tau\delta)_{jk}$$

where $y_{ijk} = 0, 1$ represents whether subject i detected the data plot (1) or did not (0), $\mu_i,\ i = 1, \ldots n$ is the subject-specific random intercept, n is the number of subjects, $\tau_j,\ j = 1, 2$ is the map type effect, $\delta_k,\ k = 1, 2, 3$ is the trend model effect. The interaction between map type and trend model allows for any map type effect to differ between trend models. As each participant provides results from 12 lineups, this model can account for each individual participants' abilities with the subject-specific random intercept.

Table 1. Parameter estimates of the fitted model fit for detection rate. All terms are statistically significant ($^{**} = 0.01$, $^{***} = 0.001$).

Term	Est.	Std. Err.	P-val.	Sig.
Intercept	-1.27	0.19	0.00	***
Hex.	1.63	0.24	0.00	***
Three Cities	-2.07	0.43	0.00	***
All Cities	1.34	0.24	0.00	***
Hex:Three Cities	1.28	0.48	0.01	**
Hex:All Cities	-1.16	0.33	0.00	***

4. Results

A total of 1273 responses were collected from 97 participants. A small number of participants, 7, were removed because they did not provide at least 11 responses, or left more than 3 at the default value of 0. These are participants that stopped early or clicked through without doing the evaluation. Set A was evaluated by 39 participants, and 51 evaluated set B. This resulted in 1080 evaluations, corresponding to 90 subjects, each evaluating 12 lineups, that were analysed on accuracy and speed. The certainty and reasons of subjects in their answers is also examined.

4.1. Accuracy

Figure 4 displays the average detection rates for the two types of plot separately for each trend model. Each trend model was tested using four repetitions, evaluations on the same data set were seen as either choropleth maps or hexagon tile maps by each group as specified in Figure 3; the detection rates for each display are connected by a line segment. The Three Cities and All Cities trend models shown in the hexagon tile map allowed viewers to detect the data plot substantially more often than the choropleth map counterparts. One replicate for the All Cities group had similar detection rates for both map types - the rate of detection using the choropleth map was much higher than other replicates. Surprisingly, participants could also detect the gradual spatial trend in the NW-SE group from the hexagon tile map. We expected that the choropleth map would be superior for the type of spatial pattern, but the data suggests the hexagon tile map performs slightly better, or equally as well.

Table 1 presents a summary of the generalised linear mixed effects model, testing the effect of map type and trend model on the detection rate. The results support the summary from Figure 4. Overall, the hexagon tile map performs marginally better

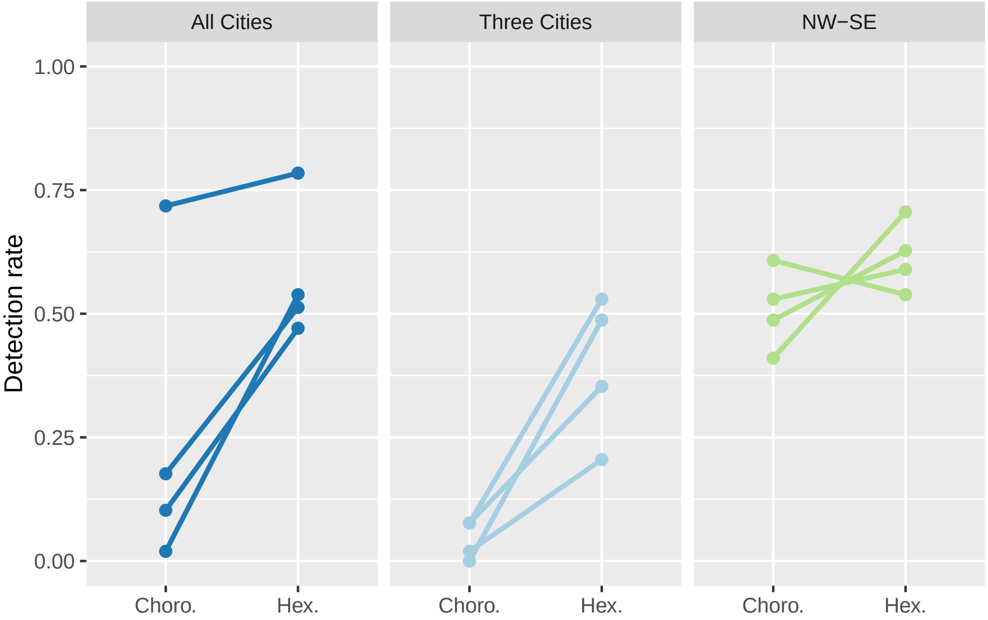

Figure 4. The detection rates achieved by participants are contrasted when viewing the four replicates of the three trend models. Each point shows the probability of detection for the lineup display, the facets separate the trend models hidden in the lineup. The points for the same data set shown in a choropleth or hexagon tile map display are linked to show the difference in the detection rate.

Table 2. Model estimates for the proportion of detection in each of the trend models (standard error). Note that selecting the data plot by chance would produce a detection rate of 0.083, for each lineup.

Map type	All Cities	Three Cities	NW-SE
Choro.	0.22 (0.03)	0.03 (0.01)	0.52 (0.04)
Hex.	0.59 (0.04)	0.39 (0.04)	0.63 (0.04)

than the choropleth map for all trend models, and with differing magnitudes of effects. The subject-specific random intercepts have mean 0.3 and standard deviation 0.55.

The estimated detection proportion from the model fit, computed using the `emmeans` package (Lenth 2025) are shown in Table 2. For the All Cities trend participants were about three times more likely to detect the cluster pattern with the hexagon tile map than the choropleth map. The Three Cities trend was practically not detectable in the choropleth map. The detection rates were more similar for the NW-SE trend, but slightly higher for the hexagon tile map, which was a surprise. Note that, these detection rates are all substantially higher than chance, except for the choropleth map on the Three Cities. For a single evaluation, the detection rate of the data plot selected by chance is $1/12 = 0.083$ because the lineups used in this experiment had 12 plots. The choice of 12 plots in the lineups instead of the usual 20, which would have produced the by chance detection rate of 0.05, is because reading a map is relatively complex, and pilot studies suggested that 12 was a reasonable cognitive load but 20 was not. When designing an experiment like this it is important to produce lineups that strike a balance between simple and hard, so that there is a chance of discovering the effect of interest. This experiment has managed to do this extremely well, which is a result of pilot studies, careful null data generation, and power calculations to determine appropriate sample size.

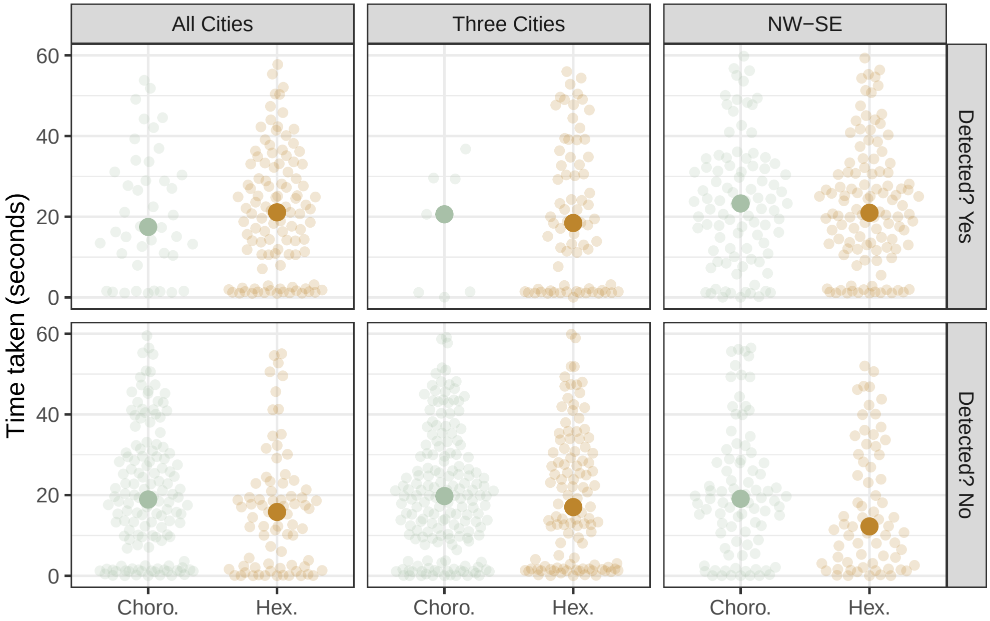

Figure 5. The distribution of the time taken (seconds) to submit a response for each combination of trend, whether the data plot was detected, and type of display, shown using a median value and horizontally jittered dotplots. There are only small differences in time taken between map types. Some participants take under a second per evaluation, and some take as much as 60 seconds, but this occurs with detection and non-detection.

4.2. Speed

Figure 5 shows horizontally jittered dot plots to contrast the time taken by participants to evaluate each lineup faceted by map type and trend model. Each dot is an evaluation. The time taken to complete an evaluation ranged from fractions of a second to 60 seconds. The average time taken for type of display is shown as a large coloured dot on each plot, and show there is little difference in the average time taken to read a lineup made with either a choropleth map or hexagon tile map.

That some evaluations occurred within milliseconds is a little surprising. Investigating whether this was related the 28 of 1080 evaluations where participants left the default choice of 0, and we find it is not; most of these people took the routine time to examine, and then left it at the default suggesting that they just could not pick one as different. This is the same as a non-detect. On a per participant basis, the average time per lineup over the 12 evaluations ranged between 3.91 and 40.42 seconds. The correlation between average detection rate and time taken across subjects 0.3 which is weakly positive. This is similar to what we have found in other studies, some subjects are especially fast and accurate in visual evaluation, and conversely some subjects are quite slow but inaccurate.

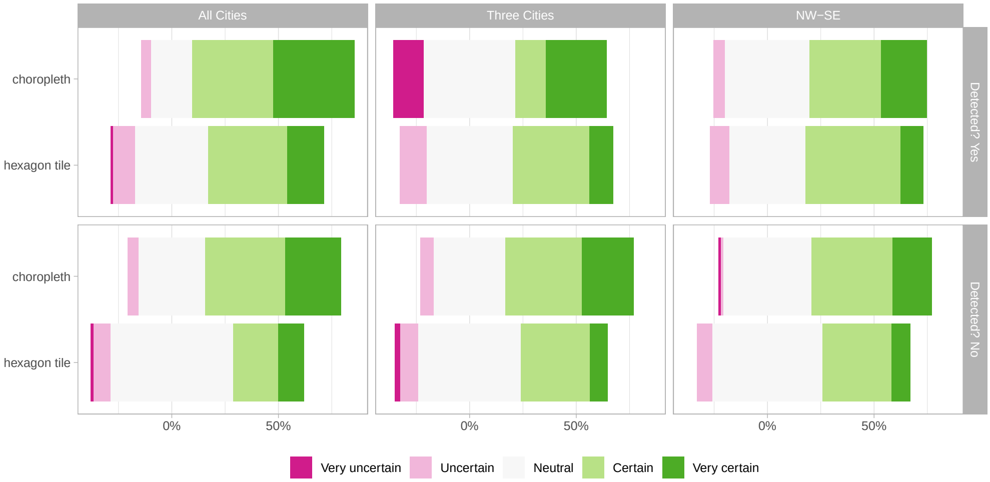

Figure 6. The distribution of certainty chosen by participants when viewing hexagon tile map or choropleth map displays, shown as centered bar plots, faceted by the trend model and whether the plot was detected or not. Participants tended to choose higher certainty when evaluating a choropleth map, on average, particularly when the data plot was not detected.

4.3. Certainty

Participants provided their level of certainty regarding their choice using a five point scale. Unlike the accuracy and speed of responses that were derived during the data processing phase, this was a subjective assessment by the participant prompted by the question: “How certain are you about your choice?”. Figure 6 shows centred bar charts summarising how participants reported their certainty with their decision. The sub-plots show each combination of trend models and whether the data plot was detected or not. Colour indicates the certainty in their decision. Participants tended to be slightly more certain when shown the choropleth map when the trend model was "All cities", and when the data plot was not detected for all three trend types.

4.4. Reason

Table 3 summarises the reasons that participants gave for their choices: "clusters" = "Clusters of colour", "trend" = "Colour trend across the areas", "consistent" = "All areas have similar colours", "hotspots" = "Big differences between neighbouring areas", "none" = "None of these reasons". These proportions are computed separately for each trend, map type and whether the participant detected the data plot. With All Cities and Three Cities, when correct, participants tended to select 'consistent' with the choropleth map, and 'clusters' with the hexagon tile map. With the NW-SE trend, 'trend' was primarily selected for the choropleth map, but 'clusters' was the primary reason for the hexagon tile map. The primary reasons are similar when the participant did not detect the data plot.

The results when the data plot was detected are as expected, but that they are similar to the not detected group is interesting. It suggests that with the choropleth map people are trying to read contiguous patterns, while the hexagon tile map is being read for pockets of differences. This may be due the hexagon tile map being non-contiguous.

4.5. Participant demographics

Of the 90 participants, 66 were male, and 24 female. Most participants (55) had a Bachelors degree, 13 had a Masters degree, and the remaining 22 had high school diplomas. The age distribution was 13, 36, 21, 11, 6 for age groups 18-24, 25-34, 35-44, 45-54, 55+, with 3 preferring not to answer. Only 1 reported having lived in Australia. Note that, the purpose of reporting these numbers is to illustrate the reasonable variety of demographic background of participants. However, 90 observations is insufficient to include this demographic information in the model. Majumder, Hofmann & Cook (2025) (published in 2025 but conducted in 2012) showed that there is little difference in performance on lineup experiments between demographic groups. VanderPlas & Hofmann (2016) found that visual aptitude for reading data plots was associated with mathematical skills, but this study was done on statistical plots of data, not maps. Assessing mathematical ability requires substantially more data collection, and does not necessarily correlate with education level.

Table 3. Proportion of reasons provided by participants for their plot choice, broken down by Trend, Map Type, and data plot detection. The primary reason when participants were evaluating the choropleth map was 'consistent' or 'trend', but for the hexagon tile map it was 'clusters', when they detected the data plot.

Trend	Detect	Type	clusters	trend	consistent	hotspots	none
All Cities	Yes	Choro.	0.33	0.17	0.33	0.07	0.10
All Cities	Yes	Hex.	0.42	0.33	0.04	0.13	0.08
Three Cities	Yes	Choro.	0.14	0.29	0.57	0.00	0.00
Three Cities	Yes	Hex.	0.49	0.32	0.00	0.14	0.06
NW-SE	Yes	Choro.	0.34	0.40	0.02	0.16	0.08
NW-SE	Yes	Hex.	0.50	0.29	0.11	0.05	0.05
All Cities	No	Choro.	0.36	0.41	0.09	0.08	0.07
All Cities	No	Hex.	0.36	0.29	0.08	0.07	0.20
Three Cities	No	Choro.	0.31	0.43	0.07	0.10	0.09
Three Cities	No	Hex.	0.39	0.31	0.06	0.10	0.14
NW-SE	No	Choro.	0.23	0.39	0.16	0.13	0.09
NW-SE	No	Hex.	0.40	0.26	0.04	0.07	0.22

Summary statistics (not included here, but available in the analysis code) show no differences in results between sets A and B in detection rate or time taken. Similarly, detection rates and time taken vary little across age, education and gender.

5. Discussion

This study provides evidence that the hexagon tile map is superior to a choropleth map for communicating population statistics, for Australia. While the cartogram has been established as better than a choropleth map, cartograms do not work for the vast disparity between population density and geographic area in Australia. The hexagon tile map was developed to provide a possible solution, and this study demonstrates that it has potential.

The R package `sugarbag` can be used to generate a hexagon tile map. It can be used for any spatial polygon data, so is applicable to other countries or geographic areas.

One of the strengths but potential limitations of the hexagon tile map is that it is non-contiguous; large rural areas are represented by isolated hexagons. This is why we expected that the hexagon tile map might not work well to detect large-scale spatial trend ("NW-SE"). The primary reason for producing the non-contiguous display was to preserve geography sufficiently for the reader to easily recognise the location. This is a strength, and allows a map of the country to be drawn underneath the hexagons. It appears to not inhibit reading of the spatial distribution based on this experiment. However, there is considerable room for improving the algorithm and exploring some variations. These might include increasing the size of isolated hexagons, or collecting multiple hexagons together.

The manner in which hexagons are exploded out from the city centres is another direction of research. Ideally, the location of the hexagons should be close to their original location but this is hard to control and measure. The current algorithm works sequentially to place hexagons, radially from a provided centre. There are likely better optimisation procedures that could improve the layout. For reading the spatial distribution, this is less important, but if the hexagon tile map is provided to users as an interactive tool, they will want to locate themselves in the plot. If the hexagon is not close to the true location it could be disconcerting.

This experiment focused on comparing a new display, the hexagon tile map, against the standard display, choropleth map. There are other options that could have been included in the study, such as the use of insets of dense population areas along with the choropleth map, or the use of interactive graphics linking statistical charts with the choropleth map. Keeping the scope of the study small was important to understand whether it was reasonable to recommend use of the hexagon tile map. Although we only tested on the Australian geography, the results should hold for other regions that have similarly disparities between population size and geographic size.

We would recommend doing follow-up studies that allow deeper understanding of how the different displays are read. For example, an eye-tracking experiment could help to understand the differences in how people read the choropleth map and the hexagon tile map as indicated by the different reasons given. Zhao et al. (2013) is an example of such an experiment where the manner in which people read lineups was examined.

While the significance of the difference in detection was the key focus of this experiment, the secondary focus was the time taken by participants. It was expected that the participants may take longer to consider the hexagon tile map distribution but would be able to detect the data plot in the lineup. The bimodal distributions seen in Figure 5 showed very little difference in the median evaluation times. As the maximum time of all of the distributions approached 60 seconds it cannot be said that the participants took longer to evaluate the hexagon tile map displays.

The responses to the questions asked of participants included the reason for their choice and the certainty around their choice. Figure 6 showed generally higher levels of certainty were chosen by participants when looking at the population distributions in a choropleth map display suggesting that they were more confident. This was especially the pattern when the data plot was not detected. The high levels of the mid-range value of “Neutral” could indicate that the participants did not want to provide a response, as this was the default value.

The colour scaling applied in Three cities and All cities displays resulted in the rural areas of the real data plot appearing more blue or yellow than the other plots in the lineups. Due to the consistent colouring of rural areas in a choropleth map display, the choice “All areas have similar colours” was most common reason for a participants choice. The All Cities displays coloured the inner-city areas of all capital cities more red, this was observable to participants and explains the equal choice of the city clusters or rural colour consistency. Choosing “Clusters of colour” was expected when participants viewed the Hexagon tile map display of the All Cities and Three Cities distributions. It was unexpected that it was also the most common reason for the NW-SE hexagon tile map displays. Due to the spatial covariance introduced in the smoothing, groups of similarly coloured hexagons were present in all of the hexagon tile map displays. All Cities and Three Cities distributions of real data trends had distinctly different patterns or red inner-city areas, while some of the plots in each lineup may have shared similar features.

6. Conclusion

The choropleth map display and the tessellated hexagon tile map have been contrasted using the lineup protocol. The hexagon tile map was significantly more effective for spotting a real population related data trend model hidden in a lineup.

The hexagon tile map display should be considered as an alternative visualisation method when communicating distributions that relate to the population across a set of geographic units. As an additional display to the familiar choropleth map, cancer atlas products may benefit from the opportunity to allow exploration via an alternative display. The spatial distributions used to test these displays were inspired by the real spatially smoothed estimates of the cancer burden on Australian communities. This technique may be useful for other population related distributions, such as other diseases, or election results or socioeconomic indicators.

The increasing population densities of capital cities despite large land area exacerbates the difference in the smallest and largest communities. The population density structure of Australia can be considered similar to that of Canada, New Zealand and many other countries. Therefore, this display is not only relevant to Australia, but all nations or population distributions that experience densely populated cities separated by vast rural expanses.

Acknowledgments

The authors would like to thank the Australian Cancer Atlas team for discussions regarding alternative spatial visualisations, and Professor Kerrie Mengersen and Dr Earl Duncan for regular meetings filled with suggestions and comments. Mitchell O'Hara-Wild was a co-developer of the taipan (Kobakian & O'Hara-Wild 2018) R package for image tagging, used as the base for the web app constructed to collect participant evaluations of lineups. We are thankful for the NUMBATs (Non-Uniform Monash Business Analytics Team) for participating in the pilot study that helped to assess the experimental design and determine an appropriate sample size for the study.

Supplementary materials and reproducibility

This document was written using quarto. This document contains the code to produce the summaries, plots and additional checks in the paper. All the code to reproduce the analysis, and do additional checks can be found at https://github.com/srkobakian/experiment. Supplementary materials have been included to discuss the survey procedures and the lineups that were used. The full set of images can be found here, too.

The supplementary material contains:

- Additional analysis of the experimental results
- Survey procedure including training materials for the participants
- 24 lineups as images, that were used in the experiment
- 12 data sets used to construct the lineups

The analysis of the work was completed in R (R Core Team 2019) with the use of the following packages:

- Document creation: quarto (Allaire et al. 2025), anzjs template (Tanaka 2024), knitr (Xie 2015).

- Lineup creation: nullabor (Wickham et al. 2018), gstat (Gräler, Pebesma & Heuvelink 2016).
- Data analysis: tidyverse (Wickham et al. 2019), ggthemes (Arnold 2019), RColorBrewer (Neuwirth 2014).
- Plots: ggplot2 (Wickham 2009), cowplot (Wilke 2019), png (Urbanek 2013), ggbeeswarm (Clarke, Sherrill-Mix & Dawson 2023), ggmosaic (Jeppson & Hofmann 2023).
- Modelling and summary presentation: lme4 (Bates et al. 2015), kableExtra (Zhu 2019).

Ethics Declaration

Ethics approval for the online survey was granted by QUT's Ethics Committee (Ethics Application Number: 1900000991). All applicants provided informed consent in line with QUT regulations prior to participating in this research.

References

Allaire, J., Teague, C., Scheidegger, C., Xie, Y., Dervieux, C. & Woodhull, G. (2025). Quarto. doi:10.5281/zenodo.5960048. URL https://github.com/quarto-dev/quarto-cli.

Arnold, J.B. (2019). *ggthemes: Extra Themes, Scales and Geoms for ggplot2*. URL https://CRAN.R-project.org/package=ggthemes. R package version 4.2.0.

Australian Bureau of Statistics (2018). Australian Statistical Geography Standard (ASGS). URL https://www.abs.gov.au/statistics/statistical-geography/australian-statistical-geography-standard-asgs.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* **67**, 1–48. URL https://doi.org/10.18637/jss.v067.i01.

Beecham, R., Dykes, J., Meulemans, W., Slingsby, A., Turkay, C. & Wood, J. (2017). Map LineUps: Effects of Spatial Structure on Graphical Inference. *IEEE Transactions on Visualization and Computer Graphics* **23**, 391–400. URL https://doi.org/10.1109/TVCG.2016.2598862.

Bryan, J. & Zhao, J. (2018). *googlesheets: Manage Google Spreadsheets from R*. URL https://CRAN.R-project.org/package=googlesheets. R package version 0.3.0.

Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E.K., Swayne, D.F. & Wickham, H. (2009). Statistical Inference for Exploratory Data Analysis and Model Diagnostics. *Philosophical Transactions of the Royal Society, A (Invited)* **367**, 4361–4383. URL https://doi.org/10.1098/rsta.2009.0120.

Cancer Council Queensland and Queensland University of Technology (2024). Australian Cancer Atlas 2.0, version 05-2024. Queensland University of Technology, Cooperative Research Centre for Spatial Information.

Clarke, E., Sherrill-Mix, S. & Dawson, C. (2023). *ggbeeswarm: Categorical Scatter (Violin Point) Plots*. URL https://doi.org/10.32614/CRAN.package.ggbeeswarm. R package version 0.7.2.

Dorling, D. (2011). Area Cartograms: Their Use and Creation. In *The Map Reader: Theories of Mapping Practice and Cartographic Representation*, eds. M. Dodge, R. Kitchin & C. Perkins. John Wiley & Sons, Ltd, pp. 252–260. URL https://doi.org/10.1002/9780470979587.ch33.

Duncan, E.W., Cramb, S.M., Aitken, J.F., Mengersen, K.L. & Baade, P.D. (2019). Development of the Australian Cancer Atlas: Spatial Modelling, Visualisation, and Reporting of Estimates. *International Journal of Health Geographics* **18**. URL https://doi.org/10.1186/s12942-019-0185-9.

Evershed, N. & Ball, A. (2025). Australian election 2025 live results: votes tracker and federal seat and senate counts. https://www.theguardian.com/australia-news/ng-interactive/2025/may/07/results-2025-federal-election-live.

Fieberg, J., Freeman, S. & Signer, J. (2024). Using Lineups to Evaluate Goodness of Fit of Animal Movement Models. *Methods in Ecology and Evolution* **15**, 1048–1059. URL https://doi.org/10.1111/2041-210X.14336.

Figure Eight Inc (2019). The Essential High-Quality Data Annotation Platform. URL https://www.figure-eight.com/.

Green, A. (2025). Australian Federal Election 2025 Live Results. https://www.abc.net.au/news/elections/federal/2025/results.

Green, J.A. (2021). Too Many Zeros and/or Highly Skewed? A Tutorial on Modelling Health Behaviour as Count Data with Poisson and Negative Binomial Regression. *Health Psychology and Behavioral Medicine* **9**, 436–455. doi:https://doi.org/10.1080/21642850.2021.1920416.

Gräler, B., Pebesma, E. & Heuvelink, G. (2016). Spatio-Temporal Interpolation using gstat. *The R Journal* **8**, 204–218. URL https://doi.org/10.32614/RJ-2016-014/.

Hofmann, H., Follett, L., Majumder, M. & Cook, D. (2012). Graphical Tests for Power Comparison of Competing Designs. *IEEE Transactions on Visualization and Computer Graphics* **18**, 2441–2448.

House, D.H. & Kocmoud, C.J. (1998). Continuous cartogram construction. In *Proceedings of the Conference on Visualization '98*. VIS '98, Washington, DC, USA: IEEE Computer Society Press, p. 197–204.

Jeppson, H. & Hofmann, H. (2023). Generalized Mosaic Plots in the ggplot2 Framework. *The R Journal* **14**, 50–78. URL https://doi.org/10.32614/RJ-2023-013.

Kaspar, S., Fabrikant, S.I. & Freckmann, P. (2011). Empirical Study of Cartograms. In *Proceedings of the 25th International Cartographic Conference*. Paris, France, pp. 1–8.

Kobakian, S., Cook, D. & Duncan, E. (2023). A Hexagon Tile Map Algorithm for Displaying Spatial Data. *The R Journal* **15**, 6–16. URL https://doi.org/10.32614/RJ-2023-021.

Kobakian, S., Cook, D. & Roberts, J. (2020). Mapping Cancer: the Potential of Cartograms and Alternative Map Displays. *Annals of Cancer Epidemiology* **4**. URL https://doi.org/10.21037/ace-19-31. Https://ace.amegroups.org/article/view/6040.

Kobakian, S. & O'Hara-Wild, M. (2018). *taipan: Tool for Annotating Images in Preparation for Analysis*. URL https://CRAN.R-project.org/package=taipan. R package version 0.1.2.

Kossmeier, M., Tran, U.S. & Voracek, M. (2019). Visual inference for the funnel plot in meta-analysis. *Hotspots in Psychology* **227**. URL https://doi.org/10.1027/2151-2604/a000358.

Lenth, R.V. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. URL https://CRAN.R-project.org/package=emmeans. R package version 1.11.1.

Li, W., Cook, D., Tanaka, E. & VanderPlas, S. (2024). A Plot is Worth a Thousand Tests: Assessing Residual Diagnostics with the Lineup Protocol. *Journal of Computational and Graphical Statistics* **33**, 1497–1511. URL https://doi.org/10.1080/10618600.2024.2344612.

Loy, A., Hofmann, H. & Cook, D. (2017). Model Choice and Diagnostics for Linear Mixed-Effects Models using Statistics on Street Corners. *Journal of Computational and Graphical Statistics* **26**, 478–492. URL https://doi.org/10.1080/10618600.2017.1330207.

Majumder, M., Hofmann, H. & Cook, D. (2013). Validation of Visual Statistical Inference, Applied to Linear Models. *Journal of the American Statistical Association* **108**, 942–956. URL https://doi.org/10.1080/01621459.2013.808157.

Majumder, M., Hofmann, H. & Cook, D. (2025). Effect of Human Factors on Visual Statistical Inference. *WIREs Computational Statistics* **17**, e70033. URL https://doi.org/10.1002/wics.70033.

Matheron, G. (1963). Principles of Geostatistics. *Economic Geology* **58**, 1246–1266. URL http://dx.doi.org/10.2113/gsecongeo.58.8.1246.

Monmonier, M. (2005). Cartography: Distortions, World-views and Creative Solutions. *Progress in Human Geography* **29**, 217–224. URL https://doi.org/10.1191/0309132505ph540pr.

Monmonier, M. (2018). *How to Lie with Maps*. Chicago: University of Chicago Press, 3rd edn.

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*. URL https://CRAN.R-project.org/package=RColorBrewer. R package version 1.1-2.

Olson, J.M. (1976). Noncontiguous Area Cartograms. *The Professional Geographer* **28**, 371–380. URL https://doi.org/10.1111/j.0033-0124.1976.00371.x.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raisz, E. (1963). Rectangular Statistical Cartograms of the World. *Journal of Geography* **35**, 8–10. doi:https://doi.org/10.1080/00221343608987880.

Reda, K. & Szafir, D.A. (2021). Rainbows Revisited: Modeling Effective Colormap Design for Graphical Inference. *IEEE Transactions on Visualization and Computer Graphics* **27**, 1032–1042. doi:https://doi.org/10.1109/TVCG.2020.3030439.

Skowronnek, A. (2016). Beyond Choropleth Maps – A Review of Techniques to Visualize Quantitative Areal Geodata. URL https://alsino.io/static/papers/BeyondChoropleths_AlsinoSkowronnek.pdf.

Tanaka, E. (2024). An unofficial quarto template for the australian and new zealand journal of statistics. https://github.com/emitanaka/quarto-anzjs.

Tufte, E.R. (1990). *Envisioning Information*. Graphics Press.

Urbanek, S. (2013). *png: Read and write PNG images*. URL https://CRAN.R-project.org/package=png. R package version 0.1-7.

VanderPlas, S. (2021). Designing Graphics Requires Useful Experimental Testing Frameworks and Graphics Derived From Empirical Results. *Harvard Data Science Review* **3**. URL https://doi.org/10.1162/99608f92.7d099fd0.

VanderPlas, S., Cook, D. & Hofmann, H. (2020). Testing Statistical Charts: What Makes a Good Graph? *Annual Review of Statistics and Its Application* **7**, 61–88. URL http://dx.doi.org/10.1146/annurev-statistics-031219-041252.

VanderPlas, S. & Hofmann, H. (2016). Spatial Reasoning and Data Displays. *IEEE Transactions on Visualization and Computer Graphics* **22**, 459–468. URL https://doi.org/10.1109/TVCG.2015.2469125.

VanderPlas, S., Röttger, C., Cook, D. & Hofmann, H. (2021). Statistical Significance Calculations for Scenarios in Visual Inference. *Stat* **10**, e337. doi:https://doi.org/10.1002/sta4.337.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer New York. URL http://had.co.nz/ggplot2/book.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686. doi:10.21105/joss.01686.

Wickham, H., Chowdhury, N.R., Cook, D. & Hofmann, H. (2018). *nullabor: Tools for Graphical Inference*. URL https://CRAN.R-project.org/package=nullabor. R package version 0.3.5.

Wickham, H., Cook, D., Hofmann, H. & Buja, A. (2010). Graphical Inference for Infovis. *IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis '10)* **16**, 973–979. URL http://dx.doi.org/10.1109/TVCG.2010.161.

Wilke, C.O. (2019). *cowplot: Streamlined Plot Theme and Plot Annotations for ggplot2*. URL https://CRAN.R-project.org/package=cowplot. R package version 1.0.0.

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Boca Raton, Florida: Chapman and Hall/CRC, 2nd edn. URL https://yihui.org/knitr/. ISBN 978-1498716963.

Yin, T., Majumder, M., Roy Chowdhury, N., Cook, D., Shoemaker, R. & Graham, M. (2013). Visual Mining Methods for RNA-Seq Data: Data Structure, Dispersion Estimation and Significance Testing. *Journal of Data Mining in Genomics & Proteomics* **4**. URL https://doi.org/10.4172/2153-0602.1000139.

Zhao, Y., Cook, D., Hofmann, H., Majumder, M. & Roy Chowdhury, N. (2013). Mind Reading: Using an Eye-Tracker to See How People are Looking at Lineups. *International Journal of Intelligent Technologies and Applied Statistics* **6**, 393–413. URL https://doi.org/10.6148/IJITAS.2013.0604.05.

Zhu, H. (2019). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. URL https://CRAN.R-project.org/package=kableExtra. R package version 1.1.0.