Comparing the Effectiveness of the Choropleth Map with a Hexagon Tile Map for Communicating Cancer Statistics

Stephanie Kobakian[1] and Dianne Cook[2]

Queensland University of Technology and Monash University

Summary

The choropleth map display is commonly used for communicating spatial distributions across geographic areas. However, when choropleths are used the size of the geographic units will influence the understanding of the distribution derived by map users. The hexagon tile map is presented as an alternative display for visualizing population related distributions effectively. Visual inference is used to measure the power of the hexagon tile map design, and the choropleth is used as a comparison. The hexagon tile map display is tested using a distribution that is directly related to the geography, with values monotonically increasing from the North-West to South-East areas of Australia. This study finds in a hexagon tile map lineup the single map that contains a population related distribution is detected with greater probability than the same data displayed in a choropleth map. These findings should encourage map creators to implement alternative displays and consider a hexagon tile map when presenting spatial distributions of heterogeneous areas.

Key words: statistics; visual inference; geospatial; population

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. 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 favor of representing each geographic unit as an

---

[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:

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 useful for other countries, and other 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 2018) 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 color scheme to draw attention to relationships between neighboring 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 (Wickham et al. 2010), to objectively test the effectiveness of the two displays.

The paper is organised as follows. The next section discusses the background of geographic data display and visual inference procedures. The Section 3 describes the methods for conducting the experiment and analysing the results. The results are summarized 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 trend it may imply that the disease is spreading from one location to another. If there is a cluster, or an anomaly, there may be a localized outbreak of the disease. Aggregating the statistic on administrative units, provides a level of privacy to individuals, while allowing the impact of the disease on the community to be analyzed.
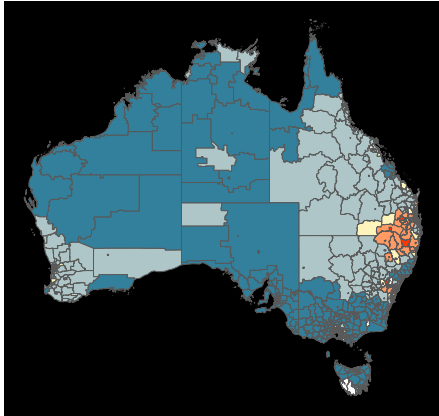
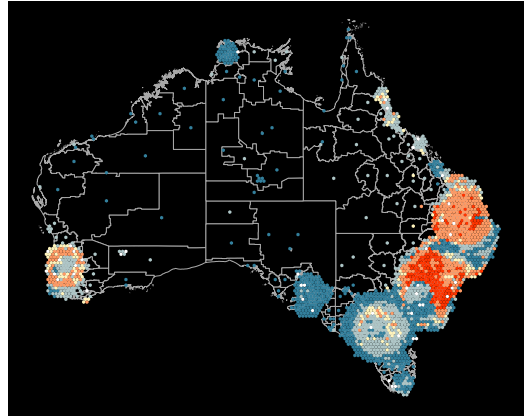a. choropleth map   b. hexagon tile map

Figure 1. Thyroid incidence among females across the Statistical Areas of Australia at Level 2, displayed using a choropleth (a) and a hexagon tile map (b). Blue indicates lower than average, and red indicates higher than average incidence. The choropleth 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 a shows the choropleth map of thyroid cancer 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 (Kocmoud & House 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, Kocmoud & House (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 recognizable.

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 (Kobakian, Cook & Roberts 2020, Skowronnek (2016)).

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 b shows the hexagon tile map, where the map is coloured from substantially below average (blue) to substantially above average (red) rates. 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 is perceived for this display, in comparison to the choropleth.

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 designed, outlined in Hofmann et al. (2012), are followed.

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.

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.

3. Methodology

This study aims to answer two key questions around the presentation of spatial distributions:
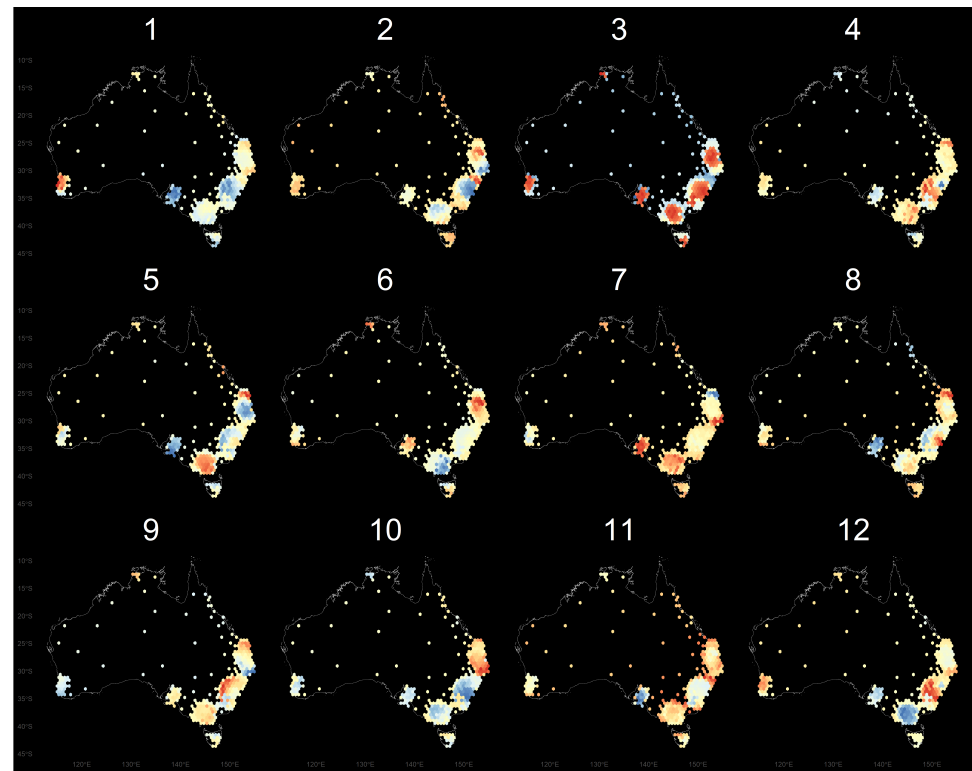
Figure 2. This lineup of twelve hexagon tile map displays contains one map with a real population related structure. The rest are null plots that contain only spatial dependence.

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 area of interest.

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 would be expected to do well, and two with differing level of inner city hot spots. These latter two reflect the structure seen in the thyroid cancer data (Figure 1). This produces six treatment levels:

- Map type: *Choropleth, Hexagon tile*
- Trend: *South-East to North-West; Locations in three population centres; Locations in multiple population centres,*

Data is generated for each of the trend models, with four replicates, and each displayed both as a choropleth 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 or as a hexagon tile map. Figure 3 summarises the design and the allocation of the displays.
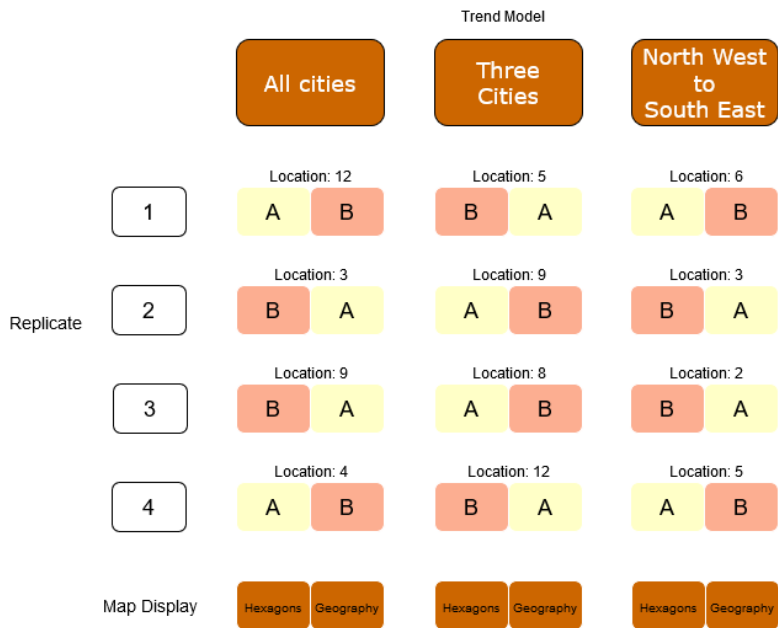

Figure 3. The experimental design used in the study. Participants are allocated to with group A or B, to evaluate either the choropleth or hexagon tile map lineup of each simulated data set.

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.

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. Each null set was further processed to produce maps with similar spatial dependence as those produced by the methods in Duncan et al. (2019), by averaging a small number of spatial neighbours.

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 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 lineup locations was chosen to embed the real trend model map, in each of the four replicates, for the three trend models. The locations were chosen by random sampling. Using random locations reduces the chance that participants might inducing the location coincidentally. Locations 1, 7, 10 and 11 were not in the sample.

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. 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 or the hexagon tile map lineup. Because there were four replicates of each lineup, each participant evaluated two choropleth 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 color 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 colored completely red.

For the population related displays, the clusters in the cities appeared more red than the rest of Australia.

3.4. Analysis

3.4.1. Data Cleaning

The first step in the data cleaning process involved checking that survey responses collected for each participants were only included once in the data set. The data

cleaning process also involved filtering out participants' who did not provide at least three unique choices when considering each of the twelve lineups. These participants achieved a detection rate of 0. If participants had made various plot choices for the 12 displays they saw they were still included in the dataset.

3.4.2. Descriptive statistics

Basic descriptive statistics were used to contrast the detection rate for the two types of displays. Comparison was also made across the trend models, contrasting the mean and standard detection rate for each group, who had seen the different map display type for each replicate.

Side-by-side dot plots were made of accuracy (efficiency) against map type, faceted by trend model type.

Similar plots were made of the feedback and demographic variables - reason for choice, reported difficulty, gender, age, education, having lived in Australia - against the design variables.

3.4.3. 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:

$$\hat{y}_{ij} \sim Bernoulli(p_i j)$$

with

$$\text{logit}(p_{ij}) = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} \quad i = 1, 2; \;\; j = 1, 2, 3$$

where $y_{ij} = 0, 1$ represents whether the subject detected the data plot (1) or did not (0), μ is the overall mean, $\tau_i, i = 1, 2$ is the map type effect, δ_j is the trend model effect. We are allowing for an interaction between map type and trend model as the response is binary, so a logistic model was used. As each participant provides results from 12 lineups, this model can account for each individual participants' abilities as it includes a subject-specific random intercept.

The model specifies a logit link, this means the predicted values from the `glmer` model should be back-transformed to fit between 0 and 1. The predictions $\widehat{p}(\eta)$ are transformed to be probabilities between 0 and 1 with the link specified below:

$$\widehat{p}(\eta) = \frac{e^{\eta}}{1 + e^{\eta}}$$

$$\eta = f(\tau_i, \delta_j)$$

3.5. 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. The data transfer from the web application to the data set took place using a secure link to the googlesheet used to store results. The application connected to the googlesheet using the googlesheets (Bryan & Zhao 2018) R package when participants opened the application, and interacted again when participants chose to submit the survey. At this time it added the participant's responses to the twelve lineup displays as twelve rows of data in the googlesheet.

3.6. Participants

Participants were recruited from the Figure Eight crowdsourcing 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. 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 trained using three test displays, 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.

3.7. Demographic 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:

- 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".

- Certainty: how certain that their choice is different from the others, on a scale of 1-5.

on each of lineup. These responses are saved after entering, but only recorded in the results sheet when the participants clicked the submit button on the final page.

4. Results

Responses from 92 participants were collected. Five participants did not provide more than three unique choices for the twelve lineups, and their data was removed. Set A was evaluated by 42 participants, and 53 evaluated set B. This resulted in 1104 evaluations, corresponding to 92 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. Participant demographics

Of the 92 participants, 67 were male, and 25 female. Most participants (56) had a Bachelors degree, 13 had a Masters degree, and the remaining 23 had high school diplomas.

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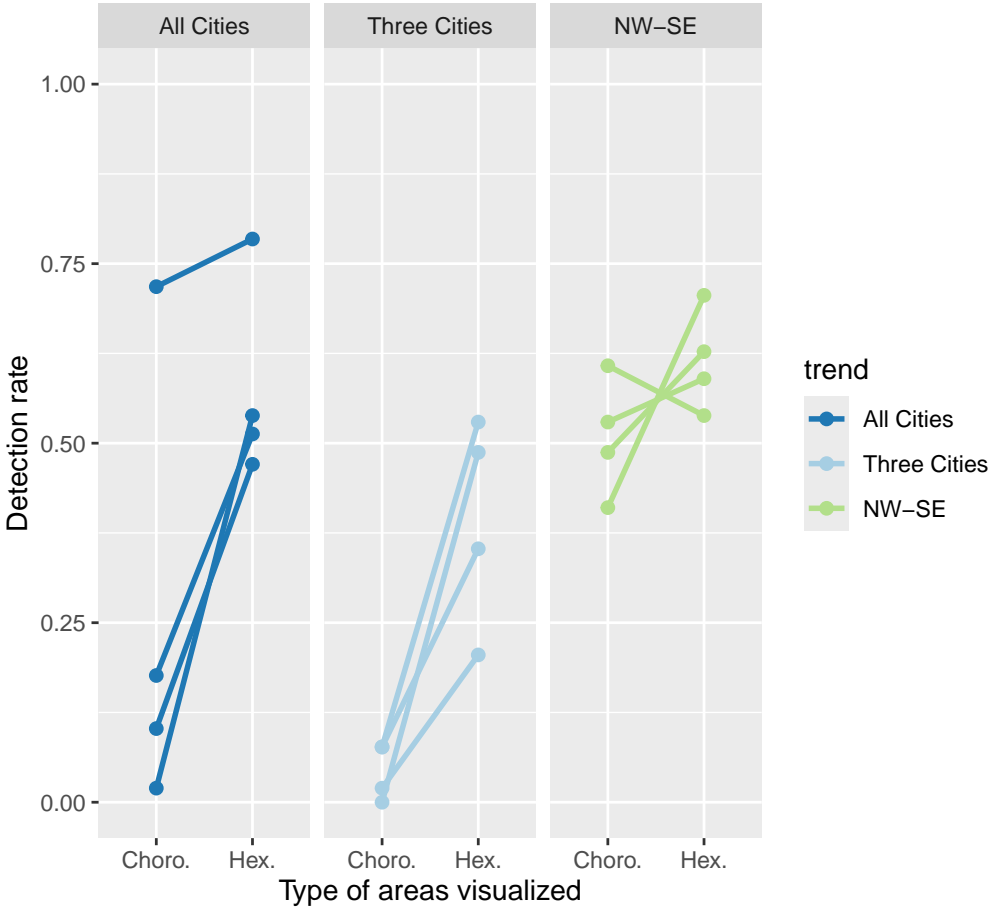
Figure 4. The detection rates achieved by participants are contrasted when viewing the four replicates of the three trend models. Each point shows the probability of detection for the lineup display, the facets separate the trend models hidden in the lineup. The points for the same data set shown in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

332   Table 1 shows the means and standard deviations of the detection rate for each type
333   of plot and each trend model. This also gives the standard deviations, the smallest
334   standard deviation for all sets of replicates was the Three Cities trend model shown
335   in a Choropleth display. This group of displays had a very small detection rate of
336   0.04. The mean detection rate for the Three Cities trend model shown as choropleth
337   map lineups was also the smallest at 0.40. The North-West to South-East (NW-SE)
338   trend model unexpectedly had a higher mean detection rate for the hexagon tile map
339   displays, but the difference in the means of detection rate was only 0.10.

Table 1

Table 2. The mean and standard deviation of the rate of detection for each trend model, calculated for the choropleth and hexagon tile map displays.

Type	NW-SE	Three Cities	All Cities
Choro.	0.52	0.04	0.23
	(0.50)	(0.19)	(0.42)
Hex.	0.62	0.40	0.58
	(0.49)	(0.49)	(0.49)

Table 3 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 and all parameters are statistically significant despite the large standard deviations observed in Table 1. Overall, the hexagon tile map performs marginally better than the choropleth for all trend models, which is a pleasant surprise. Allowing for the interaction effect, the difference in detection rate decreases for population related displays for a choropleth map lineup, but increases for a hexagon tile map display. The log odds of detection show in Table 3 can be back transformed after taking the sum of all terms for the trend and type of display that are of interest. For the NW-SE distribution, the predicted detection rate for the hexagon tile map display increases the predicted probability of detection to 0.63 from 0.52 for choropleths, this is almost exactly the difference seen in the table of means and is significant only at the 0.05 level.

When a choropleth map display is used, the predicted detection rate for the Three Cities trend, 0.03; this is extremely low, especially compared to the NW-SE trend of 0.52. When the All Cities trend is presented in a choropleth display the predicted probability of detection is 0.22. The hexagon tile map has a substantially high detection rate for the display of a Three Cities trend 0.39 and All Cities trend 0.59.

4.3. Speed

Figure 5 shows horizontally jittered dot plots to contrast the time taken by participants to evaluate each lineup when viewing each type of display. The time are also separated by trend model and whether the data plot was detected or not detected. The time taken to complete an evaluation ranged from milliseconds to 60 seconds. The average time taken for type of display is shown as a large colored dot on each plot. when considering the heights of the green and orange dots, there is little difference in the

Table 3

Table 4. The model output for the generalised linear mixed effect model for detection rate. This model considers the type of display, the trend model hidden in the data plot, and accounts for contributor performance.

Term	Est.	Sig.	Std. Error	P val
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

average time taken to read a choropleth or hexagon tile map. Comparing the same colored dot across each trend model row, there is a slight increase in the time taken to correctly detected the data plot in the hexagon tile map lineup, but little difference in evaluation time for the choropleth display. However, there were substantially less correct detections for choropleth lineups for the Three cities and All Cities trends.

4.4. 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 the amount of times participants provided each level of certainty. This was separated for each combination of trend models and display type, and colored depending on whether a participant correctly detected the data plot in the lineup. Participants often chose 4 or 5 when viewing the population related trends in the choropelth display, even though they were often incorrect when viewing an All Cities trend and overwhelmingly incorrect for the Three Cities trend. This shows overconfidence in their detection ability when using a choropleth map display. Participants were less likely to be certain when their choice was incorrect and they were viewing a hexagon tile map. For each trend model, participants were more likely to doubt their choice and choose 1 or 2 in the hexagon tile map displays, even though many had made the correct choice.
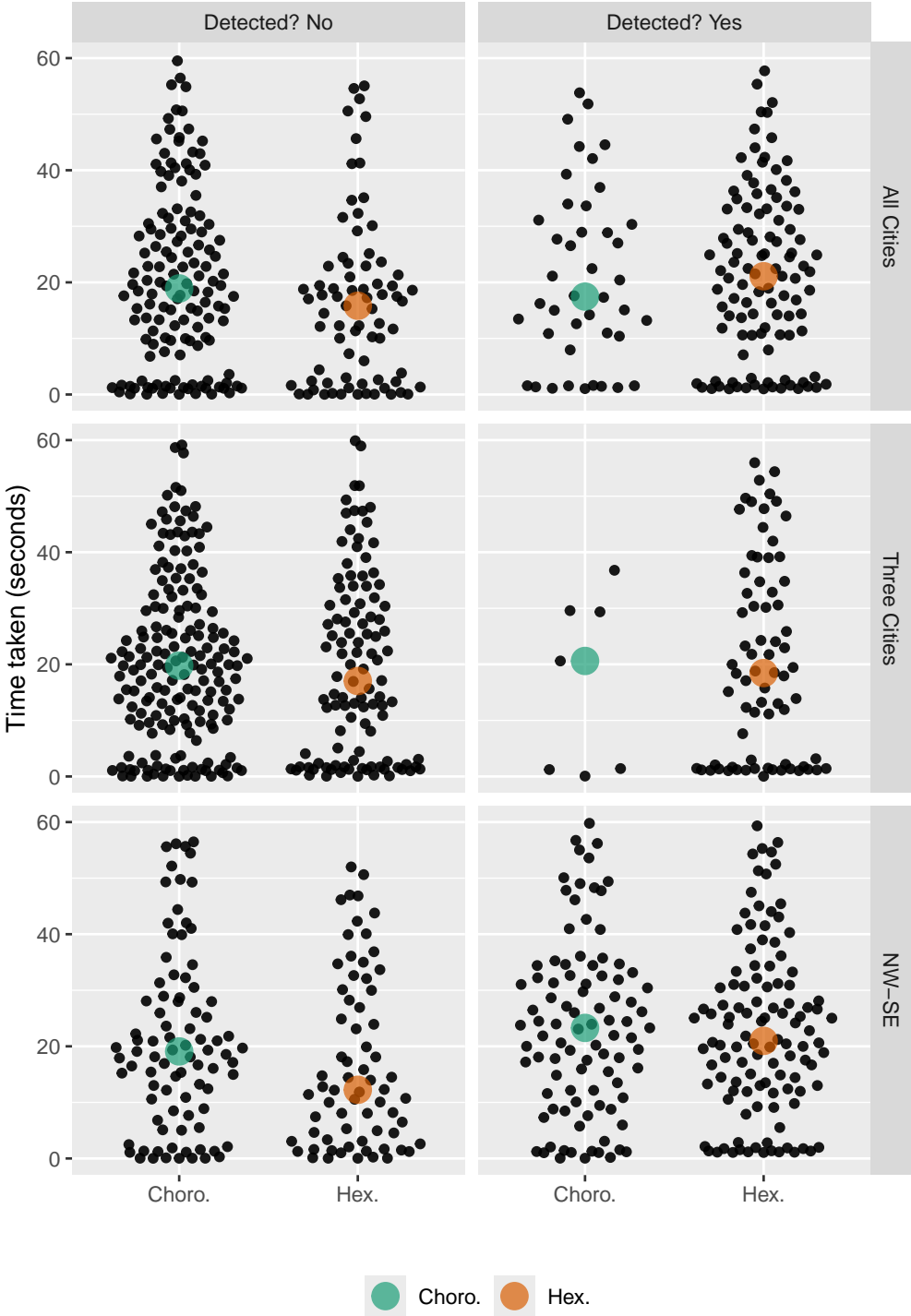
Figure 5. The distribution of the time taken (seconds) to submit a response for each combination of trend, whether the data plot was detected, and type of display, shown using horizontally jittered dotplots. The colored point indicates average time taken for each map type. Although some participants take just a few seconds per evaluation, and some take as much as much as 60 seconds, but there is very little difference in time taken between map types.
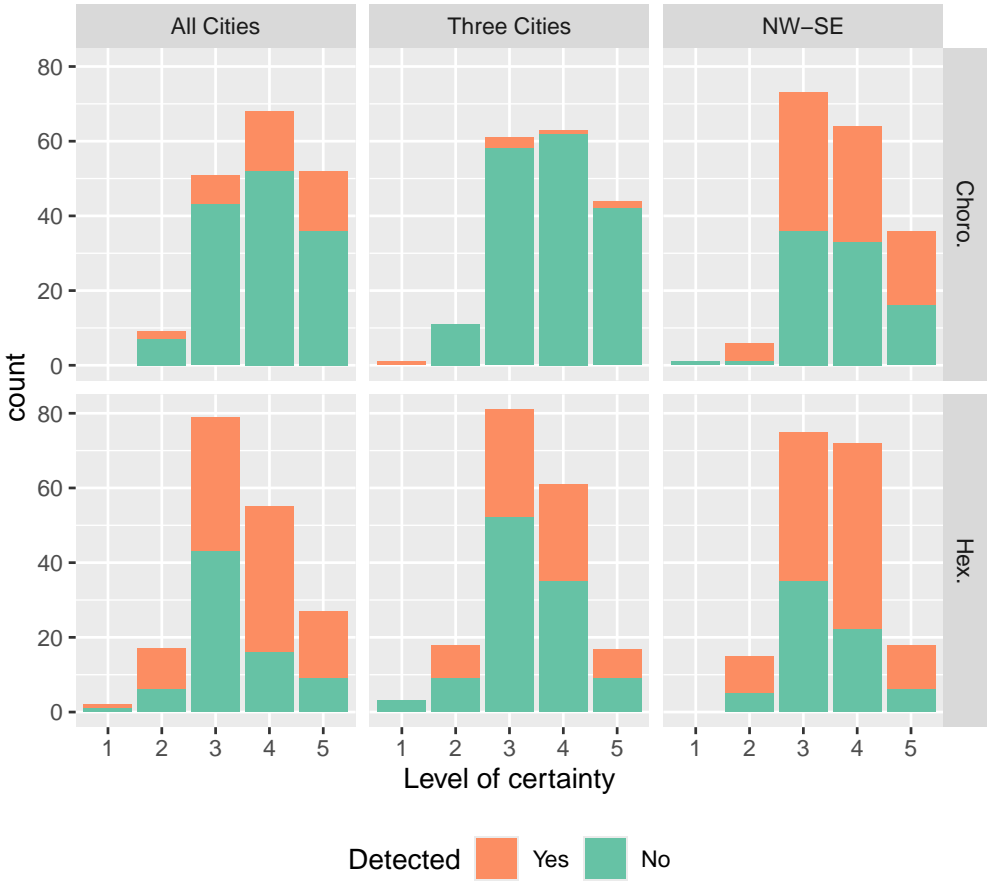
Figure 6. The amount of times each level of certainty was chosen by participants when viewing hexagon tile map or choropleth displays. Participants were more likely to choose a high certainty when considering a Choropleth map. The mid value of 3 was the default certainty, it was chosen most for the Hexagon tile map displays.

4.5. Reason

Participants were asked why they had made their plot choice and were able to select from a set of suggested reasons. "Color trend across the areas" was the most common selection for NW-SE trend displays (Table 5).

The reasons chosen by participants from the list provided to them varied more when viewing choropleth displays than the hexagon tile map. The hexagon tile map displays resulted in "Clusters of color" as the most common choice made by participants.

The choice "None of these reasons" was used as the default value to minimise noise from participants who did not select a response.

Table 5

Table 6. The amount of participants that selected each reason for their choice of plot when looking at each trend model shown in Choropleth and Hexagon Tile maps. The facets show whether or not the choice was correct.

Trend	Detected	Choro.	Hex.
All Cities	No	trend	clusters
	Yes	clusters, consistent	clusters
Three Cities	No	trend	clusters
	Yes	consistent	clusters
NW-SE	No	trend	clusters
	Yes	trend	clusters

5. Discussion

The intention of this study was to contrast the use of the choropleth map and the hexagon tile map. The visual inference lineup protocol was employed to contrast the effectiveness of the displays. The results have shown that overall the use of the hexagon tile map display allows participants to find the data plot in the lineup more often. Using the visual inference protocol this result can be extended to show that it is a valid alternative display to communicate spatial distributions of population related data.

We expected that the choropleth map would be superior for communicating the spatial pattern of geographic distributions. The data suggest that the participants perform slightly better or equally as well for each replicate in each trend model across the two displays. Table II shows that the difference in the mean detection rate for the two trend models was 0.10.

The differences seen in the Figure 4 plot and Table 1 are reflected in the model results. Surprisingly the difference for the geographic distribution was significant at the 0.05 level. It also showed that the hexagon tile map display performs marginally better than the choropleth for all trend models. Unexpectedly the detection rate suffers when using a choropleth map to display population related distributions.

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 mean 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. Table 5 shows high levels of certainty of 4 and 5 were chosen by participants when looking at the population distributions in a choropleth map display show that they were over confident when attempting to find the real data plot in the choropleth map displays. Participants performed better on the NW-SE distribution shown in the choropleth display and were reasonably confident about their decisions. The high levels of the mid range value of 3 could indicate that the participant did not want to provide a response, as this was the default value. Those who chose level 4 or 5 were equally likely to be correct for the three cities lineups, but more likely to be correct than incorrect for the other two trend models.

The color 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 coloring of rural areas in a choropleth display, the choice "All areas have similar colors" was most common reason for a participants choice. The All Cities displays colored 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 color 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 colored 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 visualization 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. However, this technique may be extended to other population related distributions, such as other diseases.

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 source code to produce this document can be found on GitHub. 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), grid (Murrell 2002).
- Modelling and summary presentation: lme4 (Bates et al. 2015), kableExtra (Zhu 2019).

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. doi:10.18637/jss.v067.i01.

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. doi:10.1098/rsta.2009.0120.

CANCER COUNCIL QUEENSLAND (2018). Australian Cancer Atlas, publisher = Queensland University of Technology, Cooperative Research Centre for Spatial Information, issue = Version 09, url=https://atlas.cancer.org.au, accessed = Jan 12 2020.

DORLING, D. (2011). *Area Cartograms: Their Use and Creation*, vol. 59, chap. 3.7. John Wiley & Sons, Ltd, pp. 252–260. doi: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.

FIGURE EIGHT INC (2019). The essential high-quality data annotation platform. URL https://www.figure-eight.com/.

GRÄLER, B., PEBESMA, E. & HEUVELINK, G. (2016). Spatio-temporal interpolation using gstat. *The R Journal* **8**, 204–218. URL https://journal.r-project.org/archive/2016/RJ-2016-014/index.html.

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.

KOBAKIAN, S., COOK, D. & DUNCAN, E. (2023). A hexagon tile map algorithm for displaying spatial data. *The R Journal* **15**, 6–16. doi:10.32614/RJ-2023-021. 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://ace.amegroups.com/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.

KOCMOUD, C. & HOUSE, D. (1998). A Constraint-based Approach to Constructing Continuous Cartograms. In *Proc. Symp. Spatial Data Handling*. pp. 236–246.

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. doi:10.1191/0309132505ph540pr. URL https://doi.org/10.1191/0309132505ph540pr. https://doi.org/10.1191/0309132505ph540pr.

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

MURRELL, P. (2002). The grid graphics package. *R News* **2**, 14–19. Https://journal.r-project.org/articles/RN-2002-010/.

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. doi:10.1111/j.0033-0124.1976.00371.x. URL https://doi.org/10.1111/j.0033-0124.1976.00371.x. 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:10.1080/00221343608987880.

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.

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.

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.

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.