

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

4 Stephanie Kobakian¹ and Dianne Cook²

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

8 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.

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

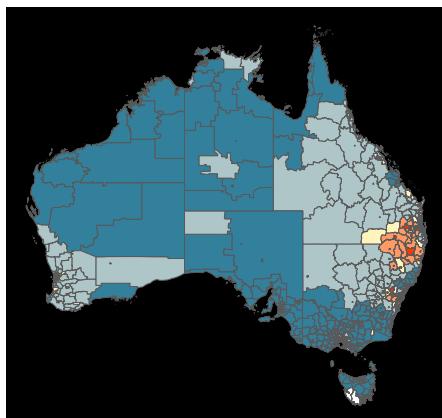
53

2. Background

54 2.1. Spatial data displays

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

a. choropleth map



b. hexagon tile map

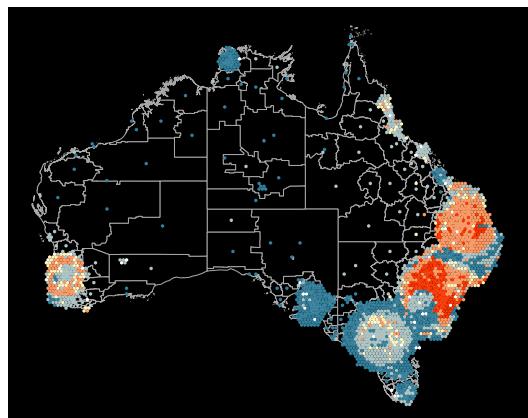


Figure 1. Thyroid cancer incidence among females across the Statistical Areas of Australia at Level 2, displayed using a choropleth map (a) and a hexagon tile map (b). Blue indicates lower than average, and red indicates higher than average incidence. The choropleth map suggests high incidence is clustered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

- 65 The choropleth map is an effective spatial display if the size of the geographic units
66 is relatively uniform. This is not the case for most countries. Size heterogeneity in
67 administrative units is particularly extreme in Australia: most of the landscape
68 of Australia is sparsely settled, with the population densely clustered into the
69 narrow coastal strips. Figure 1 shows the choropleth map of thyroid cancer incidence
70 rates in Australia. The choropleth map focuses attention on the geography, and
71 for heterogeneously sized areas it presents a biased view of the population related
72 distribution of the statistic ([House & Kocmoud 1998](#)). *Land does not get cancer, people
73 do – a more effective way to communicate the spatial distributions of cancer statistics
74 is needed (sentiment motivated by [Monmonier \(2018\)](#)).*
- 75 A cartogram is a general solution for adequately displaying a population-based statistic.
76 It transforms the geographic map base to reflect the population in the geographic
77 region, while preserving some aspects of the geographic location. There are several
78 cartogram algorithms ([Dorling 2011](#); [House & Kocmoud 1998](#)); each involves shifting
79 the boundaries of geographic units, using the value of the statistic to increase or
80 decrease the area taken by the geographic unit on the map. The changes to the
81 boundaries result in cartograms that accurately communicate population by map
82 area for each of the geographic units but can result in losing the familiar geographic
83 information. For Australia, the transformations warp the country so that it is no longer
84 recognisable (see [Kobakian, Cook & Roberts \(2020\)](#) for details).
85 Alternative algorithms make various trade-offs between familiar shapes and
86 representation of geographic units. The non-contiguous cartogram method ([Olson 1976](#))
87 keeps the shapes of geographic units intact, and changes the size of the shape. This
88 method disconnects areas creating empty space on the display losing the continuity of
89 the spatial display of the statistic. The Dorling cartogram ([Dorling 2011](#)) represents
90 each unit as a circle, sized according to the value of the statistic. The neighbour
91 relationships are mostly maintained by how the circles touch. A similar approach was
92 pioneered by [Raisz \(1963\)](#), using rectangles that tile to align borders of neighbours
93 ([Monmonier 2005](#)). There have been thorough reviews of the array of methods, as
94 suitable for cancer atlas displays (e.g. [Kobakian, Cook & Roberts 2020](#); [Skowronnek
95 2016](#)), and experiments demonstrating cartograms to be more effective than choropleth
96 maps ([Kaspar, Fabrikant & Freckmann 2011](#)).
97 The hexagon tile map algorithm, automatically matches spatial regions to their nearest
98 hexagon tile, from a grid of tiles. It has the effect of spreading out the inner city areas
99 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 & Szafrań 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

134 statistic. With the same data, two different plots can be considered to be competing
135 statistics, one possibly a more powerful statistic than the other.

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

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

146 The chance that an observer picks the data plot out of a lineup of size m plots
147 accidentally, if the null hypothesis is true is $1/m$. With K observers, the probability
148 of k randomly choosing the data plot, roughly follows a binomial distribution with
149 $p = 1/m$. Figure 2 shows a lineup of the hexagon tile map, of size $m = 12$. Plot 3 is the
150 data plot, and the remaining 11 are plots of null data. The supplementary materials
151 contain all lineups used in the study, including the corresponding lineup to this one
152 made using choropleth maps.

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

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

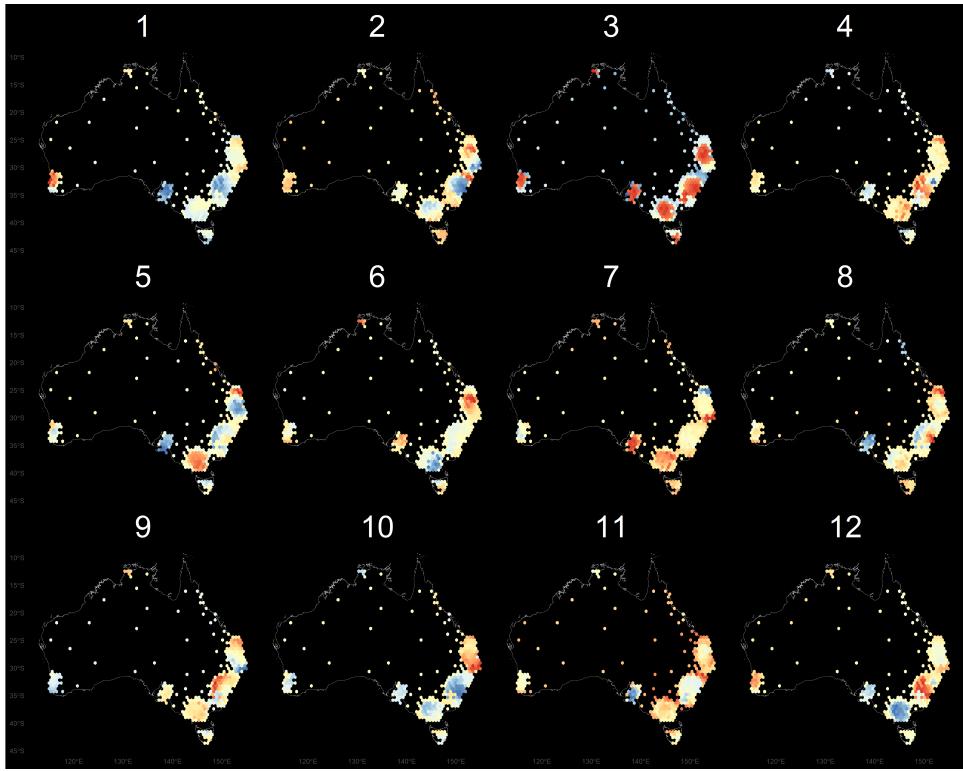


Figure 2. This lineup of twelve hexagon tile map displays contains one map with a real population related structure (location 3). The rest are null plots that contain only spatial dependence.

167

3. Methodology

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

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

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

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

179 **3.1. Experimental factors**

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

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

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

196 **3.2. Generating null data**

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

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

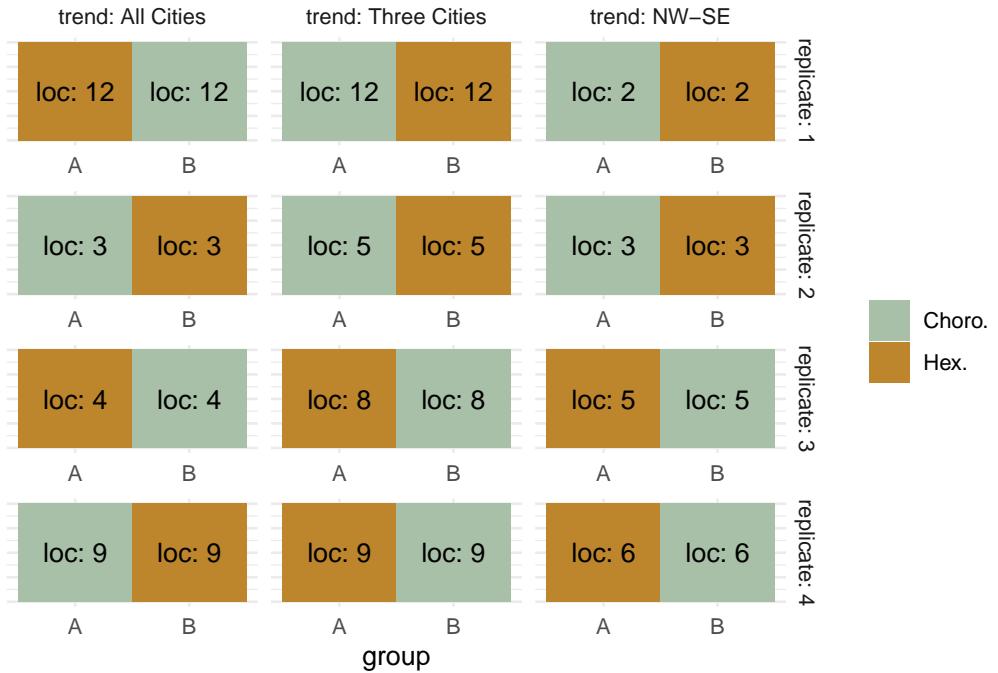


Figure 3. The experimental design used in the study. Participants are allocated to group A or B, to evaluate either the choropleth map or hexagon tile map lineup of each simulated data set. The text ‘loc’ refers to the location of the data plot in the lineup.

211 The R package `gstat` (Gräler, Pebesma & Heuvelink 2016) was used to simulate
 212 144 null sets, 12 data sets for each plot in a lineup, and 12 sets for 12 lineups.
 213 Simulating spatial dependence is difficult as discussed by Beecham et al. (2017) as
 214 a blockage for using the lineup protocol for testing map displays. Because the result
 215 from `gstat` was inadequate to mirror the spatial dependence patterns results in the
 216 Cancer Council Queensland and Queensland University of Technology (2024), each
 217 null set was further smoothed. This was done by averaging a small number of spatial
 218 neighbours, approximating the methods described in Duncan et al. (2019).

219 3.3. Generating lineups

220 For each trend model, four real data displays were created by manipulating the centroid
 221 values of each of the SA3 geographic units. Each trend model is motivated by patterns
 222 observed in spatial data: North West to South East (NW–SE) is a basic spatial trend
 223 across the entire country, Three Cities is the existence of clusters of high values, and
 224 All Cities is also clusters, but more of them. We would expect that clusters pattern to

225 be more visible with a hexagon tile map but the large spatial trend to be more visible
226 in the choropleth map.

227 The NW-SE distribution was created using a linear equation of the centroid longitude
228 and latitude values. The All Cities trend model was created using the distance from
229 the centroid of each geographic unit to the closest capital city in Australia, calculated
230 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
232 city areas, the values of these areas were increased to create red clusters. The amount
233 was chosen to make clusters around the cities visible, even in the choropleth map with
234 careful inspection. A similar selection process was applied to the Three Cities' trend
235 model. However, for each of the four replicates for the Three Cities trend, a random
236 sample of capital cities was taken from Sydney, Brisbane, Melbourne, Adelaide, Perth,
237 and Hobart. Only values of the areas nearest to the three cities were increased to
238 create clusters.

239 One of the plot locations (1-20) is chosen to embed the data plot, in each of the
240 four replicates, for the three trend models. These locations were chosen by random
241 sampling. Using random locations reduces the chance that participants might deduce
242 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
244 the location of the data plot in replicates of treatments. Their results demonstrated
245 that the actual location of the data plot didn't affect performance. So we don't expect
246 that the location affects results but randomising locations among different lineups is
247 to guard against participants expecting the data plot to be in a particular location.

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

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

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

265 **3.4. Web application to collect responses**

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

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

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

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

285 **3.5. Participants**

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

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

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

301 3.6. Data collection

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

304 Demographics were collected regarding the study participants:

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

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

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

322 **3.7. Analysis**

323 **3.7.1. Data Cleaning**

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

329 **3.7.2. Descriptive statistics**

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

334 **3.7.3. Modelling**

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

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

339 with

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

340 where $y_{ijk} = 0, 1$ represents whether subject i detected the data plot (1) or did not (0),
 341 μ_i , $i = 1, \dots, n$ is the subject-specific random intercept, n is the number of subjects,
 342 τ_j , $j = 1, 2$ is the map type effect, δ_k , $k = 1, 2, 3$ is the trend model effect. The
 343 interaction between map type and trend model allows for any map type effect to differ
 344 between trend models. As each participant provides results from 12 lineups, this model
 345 can account for each individual participants' abilities with the subject-specific random
 346 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	***

347

4. Results

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

355 **4.1. Accuracy**

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

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

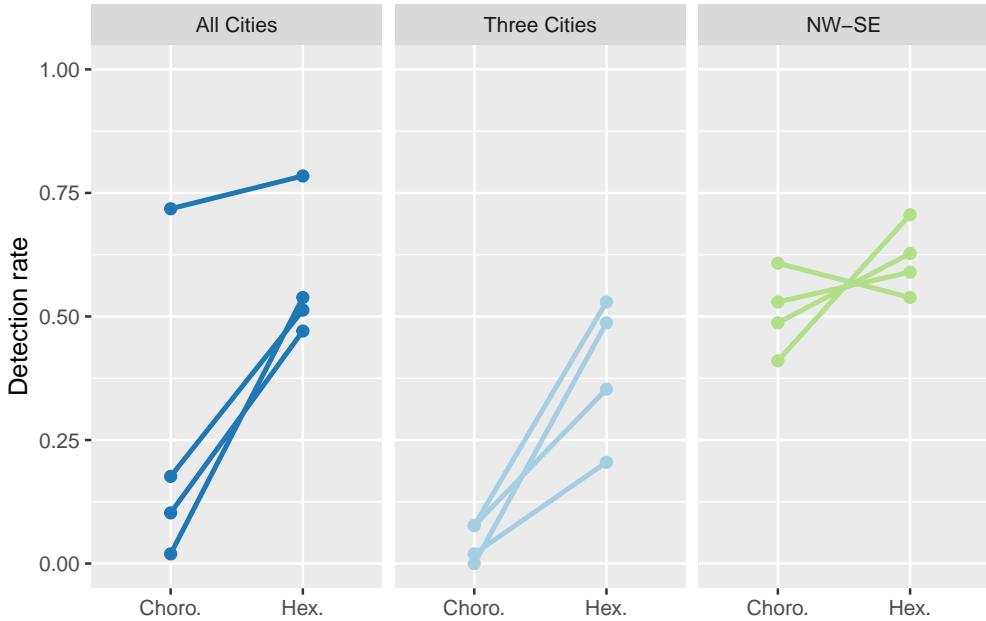


Figure 4. The detection rates achieved by participants are contrasted when viewing the four replicates of the three trend models. Each point shows the probability of detection for the lineup display, the facets separate the trend models hidden in the lineup. The points for the same data set shown in a choroleth 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)

- 371 than the choropleth map for all trend models, and with differing magnitudes of effects.
 372 The subject-specific random intercepts have mean 0.3 and standard deviation 0.55.
 373 The estimated detection proportion from the model fit, computed using the `emmeans`
 374 package (Lenth 2025) are shown in Table 2. For the All Cities trend participants were
 375 about three times more likely to detect the cluster pattern with the hexagon tile map
 376 than the choropleth map. The Three Cities trend was practically not detectable in
 377 the choropleth map. The detection rates were more similar for the NW-SE trend,
 378 but slightly higher for the hexagon tile map, which was a surprise. Note that, these

379 detection rates are all substantially higher than chance, except for the choropleth
 380 map on the Three Cities. For a single evaluation, the detection rate of the data plot
 381 selected by chance is $1/12 = 0.083$ because the lineups used in this experiment had 12
 382 plots. The choice of 12 plots in the lineups instead of the usual 20, which would have
 383 produced the by chance detection rate of 0.05, is because reading a map is relatively
 384 complex, and pilot studies suggested that 12 was a reasonable cognitive load but 20
 385 was not. When designing an experiment like this it is important to produce lineups
 386 that strike a balance between simple and hard, so that there is a chance of discovering
 387 the effect of interest. This experiment has managed to do this extremely well, which
 388 is a result of pilot studies, careful null data generation, and power calculations to
 389 determine appropriate sample size.

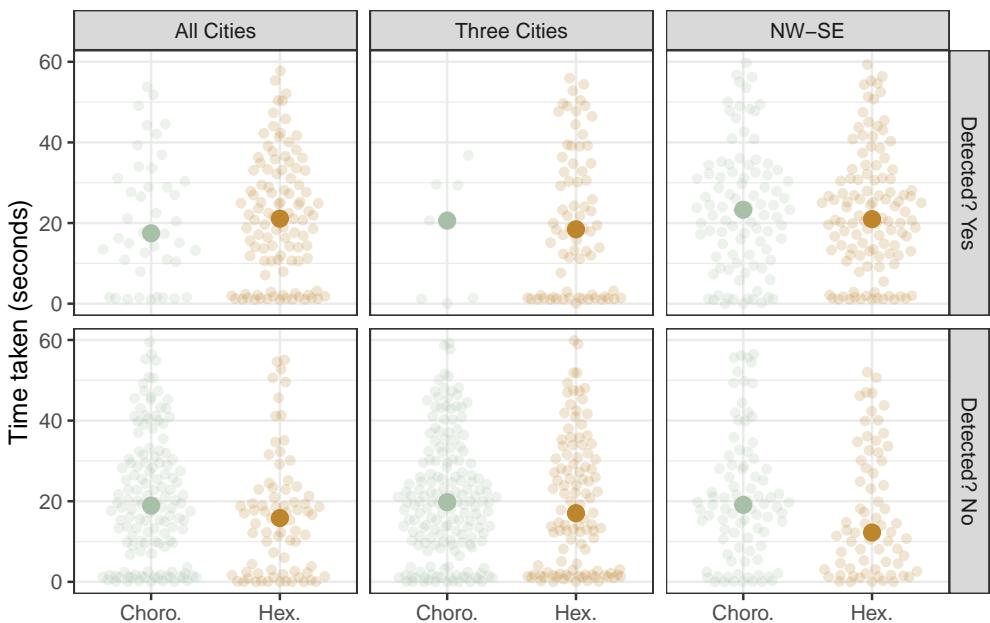


Figure 5. The distribution of the time taken (seconds) to submit a response for each combination of trend, whether the data plot was detected, and type of display, shown using a median value and horizontally jittered dotplots. There are only small differences in time taken between map types. Some participants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

390 4.2. Speed

391 Figure 5 shows horizontally jittered dot plots to contrast the time taken by participants
 392 to evaluate each lineup faceted by map type and trend model. Each dot is an evaluation.
 393 The time taken to complete an evaluation ranged from fractions of a second to 60
 394 seconds. The average time taken for type of display is shown as a large coloured dot

395 on each plot, and show there is little difference in the average time taken to read a
 396 lineup made with either a choropleth map or hexagon tile map.

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

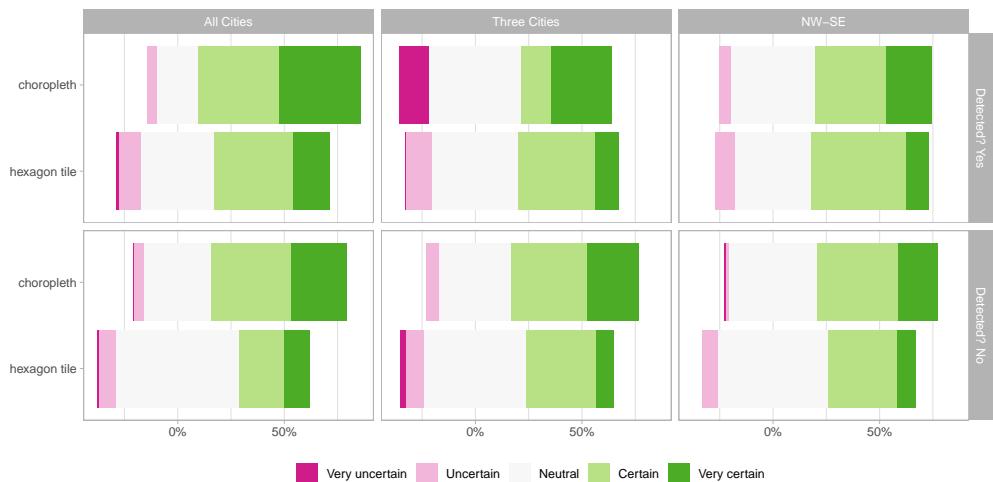


Figure 6. The distribution of certainty chosen by participants when viewing hexagon tile map or choropleth map displays, shown as centered bar plots, faceted by the trend model and whether the plot was detected or not. Participants tended to choose higher certainty when evaluating a choropleth map, on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

407 4.3. Certainty

408 Participants provided their level of certainty regarding their choice using a five point
 409 scale. Unlike the accuracy and speed of responses that were derived during the data
 410 processing phase, this was a subjective assessment by the participant prompted by
 411 the question: “How certain are you about your choice?”. Figure 6 shows centred bar
 412 charts summarising how participants reported their certainty with their decision. The
 413 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

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

449

5. Discussion

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

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

458 One of the strengths but potential limitations of the hexagon tile map is that it is
 459 non-contiguous; large rural areas are represented by isolated hexagons. This is why we
 460 expected that the hexagon tile map might not work well to detect large-scale spatial
 461 trend (“NW-SE”). The primary reason for producing the non-contiguous display was
 462 to preserve geography sufficiently for the reader to easily recognise the location. This

463 is a strength, and allows a map of the country to be drawn underneath the hexagons.
464 It appears to not inhibit reading of the spatial distribution based on this experiment.
465 However, there is considerable room for improving the algorithm and exploring some
466 variations. These might include increasing the size of isolated hexagons, or collecting
467 multiple hexagons together.

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

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

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

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

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

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

517 6. Conclusion

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

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

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

535 Acknowledgments

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

544 Supplementary materials and reproducibility

545 This document was written using quarto. This document contains the code to produce
546 the summaries, plots and additional checks in the paper. All the code to reproduce the
547 analysis, and do additional checks can be found on [GitHub](#). Supplementary materials
548 have been included to discuss the survey procedures and the lineups that were used.
549 The full set of images can be found here, too.

550 The supplementary material contains:

- 551 • Additional analysis of the experimental results
- 552 • Survey procedure including training materials for the participants
- 553 • 24 lineups as images, that were used in the experiment
- 554 • 12 data sets used to construct the lineups

555 The analysis of the work was completed in R ([R Core Team 2019](#)) with the use of the
556 following packages:

- 557 • Document creation: quarto ([Allaire et al. 2025](#)), anzjs template ([Tanaka 2024](#)),
558 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., SCHEIDECKER, 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.ggbbeeswarm>. R package version 0.7.2.

- 599 DORLING, D. (2011). *Area Cartograms: Their Use and Creation*. pp. 252–260.
- 600 DUNCAN, E.W., CRAMB, S.M.,AITKEN, J.F.,MENGERSEN, K.L. & BAADE, P.D. (2019).
- 601 Development of the Australian Cancer Atlas: Spatial Modelling, Visualisation, and Reporting
- 602 of Estimates. *International Journal of Health Geographics* **18**. URL <https://doi.org/10.1186/s12942-019-0185-9>.
- 604 EVERSHED, N. & BALL, A. (2025). Australian election 2025 live results: votes tracker
- 605 and federal seat and senate counts. <https://www.theguardian.com/australia-news/ng-interactive/2025/may/07/results-2025-federal-election-live>.
- 607 FIEBERG, J., FREEMAN, S. & SIGNER, J. (2024). Using Lineups to Evaluate Goodness of Fit
- 608 of Animal Movement Models. *Methods in Ecology and Evolution* **15**, 1048–1059. URL
- 609 <https://doi.org/10.1111/2041-210X.14336>.
- 610 FIGURE EIGHT INC (2019). The Essential High-Quality Data Annotation Platform. URL
- 611 <https://www.figure-eight.com/>.
- 612 GREEN, A. (2025). Australian Federal Election 2025 Live Results.
- 613 <https://www.abc.net.au/news/elections/federal/2025/results>.
- 614 GREEN, J.A. (2021). Too Many Zeros and/or Highly Skewed? A Tutorial on Modelling Health
- 615 Behaviour as Count Data with Poisson and Negative Binomial Regression. *Health Psychology*
- 616 and Behavioral Medicine
- 617 GRÄLER, B., PEBESMA, E. & HEUVELINK, G. (2016). Spatio-Temporal Interpolation using gstat.
- 618 *The R Journal* **8**, 204–218. URL <https://doi.org/10.32614/RJ-2016-014>.
- 619 HOFMANN, H., FOLLETT, L., MAJUMDER, M. & COOK, D. (2012). Graphical Tests for Power
- 620 Comparison of Competing Designs. *IEEE Transactions on Visualization and Computer*
- 621 Graphics **18**, 2441–2448.
- 622 HOUSE, D.H. & KOCHMOUD, C.J. (1998). Continuous cartogram construction. In *Proceedings of*
- 623 *the Conference on Visualization '98*. VIS '98, Washington, DC, USA: IEEE Computer Society
- 624 Press, p. 197–204.
- 625 JEPPSON, H. & HOFMANN, H. (2023). Generalized Mosaic Plots in the ggplot2 Framework. *The*
- 626 *R Journal* **14**, 50–78. URL <https://doi.org/10.32614/RJ-2023-013>.
- 627 KASPAR, S., FABRIKANT, S.I. & FRECKMANN, P. (2011). Empirical Study of Cartograms. In
- 628 *Proceedings of the 25th International Cartographic Conference*. Paris, France, pp. 1–8.
- 629 KOBAKIAN, S., COOK, D. & DUNCAN, E. (2023). A Hexagon Tile Map Algorithm for Displaying
- 630 Spatial Data. *The R Journal* **15**, 6–16. URL <https://doi.org/10.32614/RJ-2023-021>.
- 631 KOBAKIAN, S., COOK, D. & ROBERTS, J. (2020). Mapping Cancer: the Potential of Cartograms
- 632 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](https://ace.amegroups.org/article/view/6040).
- 633 KOBAKIAN, S. & O'HARA-WILD, M. (2018). *taipan: Tool for Annotating Images in Preparation*
- 634 *for Analysis*. URL <https://CRAN.R-project.org/package=taipan>. R package version 0.1.2.
- 635 KOSSMEIER, M., TRAN, U.S. & VORACEK, M. (2019). Visual inference for the funnel plot in meta-
- 636 analysis. *Hotspots in Psychology* **227**. URL <https://doi.org/10.1027/2151-2604/a000358>.
- 637 LENTH, R.V. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. URL
- 638 <https://CRAN.R-project.org/package=emmeans>. R package version 1.11.1.
- 639 LI, W., COOK, D., TANAKA, E. & VANDERPLAS, S. (2024). A Plot is Worth a Thousand Tests:
- 640 Assessing Residual Diagnostics with the Lineup Protocol. *Journal of Computational and*
- 641 *Graphical Statistics* **33**, 1497–1511. URL <https://doi.org/10.1080/10618600.2024.2344612>.
- 642 LOY, A., HOFMANN, H. & COOK, D. (2017). Model Choice and Diagnostics for Linear Mixed-
- 643 Effects Models using Statistics on Street Corners. *Journal of Computational and Graphical*
- 644 *Statistics* **26**, 478–492. URL <https://doi.org/10.1080/10618600.2017.1330207>.
- 645 MAJUMDER, M., HOFMANN, H. & COOK, D. (2013). Validation of Visual Statistical Inference,
- 646 Applied to Linear Models. *Journal of the American Statistical Association* **108**, 942–956.

- 648 URL <https://doi.org/10.1080/01621459.2013.808157>.
- 649 MAJUMDER, M., HOFMANN, H. & COOK, D. (2025). Effect of Human Factors on Visual Statistical
650 Inference. *WIREs Computational Statistics* **17**, e70033. URL <https://doi.org/10.1002/wics.70033>.
- 651 MATHERON, G. (1963). Principles of Geostatistics. *Economic Geology* **58**, 1246–1266. URL
652 <http://dx.doi.org/10.2113/gsecongeo.58.8.1246>.
- 653 MONMONIER, M. (2005). Cartography: Distortions, World-views and Creative Solutions. *Progress
654 in Human Geography* **29**, 217–224. URL <https://doi.org/10.1191/0309132505ph540pr>.
- 655 MONMONIER, M. (2018). *How to Lie with Maps*. Chicago: University of Chicago Press, 3rd edn.
- 656 NEUWIRTH, E. (2014). *RColorBrewer: ColorBrewer Palettes*. URL <https://CRAN.R-project.org/package=RColorBrewer>. R package version 1.1-2.
- 657 OLSON, J.M. (1976). Noncontiguous Area Cartograms. *The Professional Geographer* **28**, 371–380.
658 URL <https://doi.org/10.1111/j.0033-0124.1976.00371.x>.
- 659 R CORE TEAM (2019). *R: A Language and Environment for Statistical Computing*. R Foundation
660 for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 661 RAISZ, E. (1963). Rectangular Statistical Cartograms of the World. *Journal of Geography* **35**,
662 8–10. doi:<https://doi.org/10.1080/00221343608987880>.
- 663 REDA, K. & SZAFIR, D.A. (2021). Rainbows Revisited: Modeling Effective Colormap Design
664 for Graphical Inference. *IEEE Transactions on Visualization and Computer Graphics* **27**,
665 1032–1042. doi:<https://doi.org/10.1109/TVCG.2020.3030439>.
- 666 SKOWRONNEK, A. (2016). Beyond Choropleth Maps – A Review of Techniques to Visualize
667 Quantitative Areal Geodata. URL https://alsino.io/static/papers/BeyondChoropleths_Alsi_noSkowronnek.pdf.
- 668 TANAKA, E. (2024). An unofficial quarto template for the australian and new zealand journal of
669 statistics. <https://github.com/emitanaka/quarto-anzjs>.
- 670 TUFTE, E.R. (1990). *Envisioning Information*. Graphics Press.
- 671 URBANEK, S. (2013). *png: Read and write PNG images*. URL <https://CRAN.R-project.org/package=png>. R package version 0.1-7.
- 672 VANDERPLAS, S. (2021). Designing Graphics Requires Useful Experimental Testing Frameworks
673 and Graphics Derived From Empirical Results. *Harvard Data Science Review* **3**. URL
674 <https://doi.org/10.1162/99608f92.7d099fd0>.
- 675 VANDERPLAS, S., COOK, D. & HOFMANN, H. (2020). Testing Statistical Charts: What
676 Makes a Good Graph? *Annual Review of Statistics and Its Application* **7**, 61–88. URL
677 <http://dx.doi.org/10.1146/annurev-statistics-031219-041252>.
- 678 VANDERPLAS, S. & HOFMANN, H. (2016). Spatial Reasoning and Data Displays. *IEEE Transactions
679 on Visualization and Computer Graphics* **22**, 459–468. URL <https://doi.org/10.1109/TVCG.2015.2469125>.
- 680 VANDERPLAS, S., RÖTTGER, C., COOK, D. & HOFMANN, H. (2021). Statistical Significance
681 Calculations for Scenarios in Visual Inference. *Stat* **10**, e337. doi:<https://doi.org/10.1002/sta4.337>.
- 682 WICKHAM, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer New York. URL
683 <http://had.co.nz/ggplot2/book>.
- 684 WICKHAM, H., AVERICK, M., BRYAN, J., CHANG, W., McGOWAN, L.D., FRANÇOIS, R.,
685 GROLEMUND, G., HAYES, A., HENRY, L., HESTER, J., KUHN, M., PEDERSEN, T.L., MILLER,
686 E., BACHE, S.M., MÜLLER, K., OOMS, J., ROBINSON, D., SEIDEL, D.P., SPINU, V.,
687 TAKAHASHI, K., VAUGHAN, D., WILKE, C., WOO, K. & YUTANI, H. (2019). Welcome to
688 the tidyverse. *Journal of Open Source Software* **4**, 1686. doi:[10.21105/joss.01686](https://doi.org/10.21105/joss.01686).
- 689 WICKHAM, H., CHOWDHURY, N.R., COOK, D. & HOFMANN, H. (2018). *nullabor: Tools for
690 Graphical Inference*. URL <https://CRAN.R-project.org/package=nullabor>. R package

- 697 version 0.3.5.
- 698 WICKHAM, H., COOK, D., HOFMANN, H. & BUJA, A. (2010). Graphical Inference for Infovis.
699 *IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis '10)* **16**, 973–979.
700 URL <http://dx.doi.org/10.1109/TVCG.2010.161>.
- 701 WILKE, C.O. (2019). *cowplot: Streamlined Plot Theme and Plot Annotations for ggplot2*. URL
702 <https://CRAN.R-project.org/package=cowplot>. R package version 1.0.0.
- 703 XIE, Y. (2015). *Dynamic Documents with R and knitr*. Boca Raton, Florida: Chapman and
704 Hall/CRC, 2nd edn. URL <https://yihui.org/knitr/>. ISBN 978-1498716963.
- 705 YIN, T., MAJUMDER, M., ROY CHOWDHURY, N., COOK, D., SHOEMAKER, R. & GRAHAM, M.
706 (2013). Visual Mining Methods for RNA-Seq Data: Data Structure, Dispersion Estimation
707 and Significance Testing. *Journal of Data Mining in Genomics & Proteomics* **4**. URL
708 <https://doi.org/10.4172/2153-0602.1000139>.
- 709 ZHAO, Y., COOK, D., HOFMANN, H., MAJUMDER, M. & ROY CHOWDHURY, N. (2013). Mind
710 Reading: Using an Eye-Tracker to See How People are Looking at Lineups. *International
711 Journal of Intelligent Technologies and Applied Statistics* **6**, 393–413. URL <https://doi.org/10.6148/IJITAS.2013.0604.05>.
- 713 ZHU, H. (2019). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. URL
714 <https://CRAN.R-project.org/package=kableExtra>. R package version 1.1.0.