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

Stephanie Kobakian
Queensland University of Technology
Science and Engineering Faculty
Brisbane, Australia
stephanie.kobakian@qut.edu.au

Dianne Cook
Monash University
Econometrics and Business Statistics Faculty
Melbourne, Australia
dicook@monash.edu

***Abstract*—The choropleth map display is commonly used for communicating spatial distributions across geographic areas. However, when choropleths are used, the sizes of areas can lead to the misinterpretation of the distribution. The visualization method used to present geospatial data will influence the understanding of the distribution derived by map users. Choosing an effective alternative display could positively influence the communication of the spatial distribution. 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 design, and the choropleth is used as a comparison. The hexagon tile map display is also tested using a distribution that is directly related to the geography, with values monotonically decreasing from the North-West to South-East areas of Australia. This study finds the single map in a hexagon tile map lineup 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.**

***Index Terms*—statistics; visual inference; geospatial; population**

I. Introduction

This study compares the effectiveness of a new type of display, a hexagon tile map, against the standard, a choropleth map, for communicating information about disease statistics. The choropleth map is the traditional approach for visualizing aggregated statistics across administrative boundaries. A hexagon tile map forgoes the familiar boundaries, in favor of representing each geographic unit as an equally sized hexagon, placed approximately in the correct spatial location. The hexagon tile map builds on existing displays, such as the cartogram, and tesselated hexagon displays. It differs in that it relaxes the requirement to have connected hexagons, and allows sparsely located hexagons. The algorithm to construct a hexagon tile map is available in the R package sugarbag [1]. This type of display may be useful for other countries, and other purposes.

The hexagon tile map was designed for Australia, motivated by a need to display spatial statistics for a new 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 [2] 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 for the atlas.

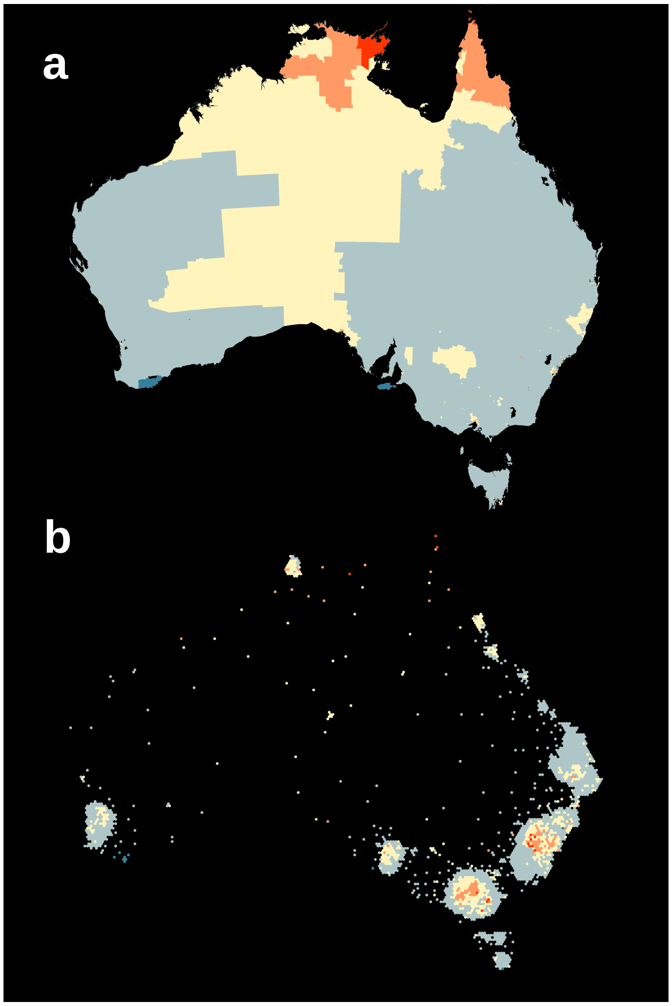
The experiment was conducted using the lineup protocol, a visual inference procedure [3], to objectively test the effectiveness of the two displays.

The paper is organised as follows. The next section discusses the background of geographic data display and visual inference procedures. The Methodology Section describes the methods for conducting the experiment and analysing the results. The results are summarized in Section Results.

II. Background

A. Spatial data displays

Spatial visualisations communicate the distribution of statistics over geographic landscapes. The choropleth map [4], [5] 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.



The choropleth map is effective 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. A choropleth map focuses attention on the geography, and for heterogeously sized areas it presents a biased view of the population related distribution of the statistic [6]. *Land doesn't get cancer, people do* – a more effective way to communicate the spatial distributions of cancer statistics is needed.

A cartogram is a general solution for better 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 [7], [6]; 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 [8] 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 [7] 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 [9], using rectangles that tile to align borders of neighbours [10]. A thorough review of the array of methods, as suitable for cancer atlas displays is available in XXX.

The hexagon tile map algorithm, autmatically 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 (XXX REF). Figure **??** shows the hexagon tile map, along with the choropleth map of liver cancer rates in Australia. Colour maps from substantially below average (blue) to substantially above average (red) rates. The inner city areas are expanded out, 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 oof interest to know how well the spatial distribution is perceived for this display, in comparison to the choropleth.

B. *Visual Inference*

We suggest the hexagon tile map as an alternative display. To test the effectiveness of these displays, there is a formal testing framework. Visual inference considers the communication of data or data summaries through visualizations. It considers these visualizations to be visual statistics, with observable features.

Classical statistical inference involves hypothesis testing, rejecting a null hypothesis in favor of an alternative. This approach requires data with the appropriate distributions and assumptions. The lineup protocol was formalized as a test for the effectiveness of a visualization. Rather than classical statistical tests, it tasks human participants with finding the display that shows structure between the variables. If they do not detect the structure, then it is not significantly different from the null data.

Visual inference uses the general hypothesis structure:

- Null hypothesis: There is **no relationship between the variables**
- **Alternative hypothesis: There is a relationship between the variables**

To test these hypotheses, the line up protocol randomly places a "guilty" data visualization in a lineup of "innocents". Where the guilty data set contains structure, and the innocents data sets do not. In a grid of visualizations, an observer is asked to pick the display that is most different, if they select the data set containing structure, they have identified the guilty hidden within the innocents [11]. The participants identify the guilty data as different from the innocent data with probability $1/m$, where m is the number of null plots plus 1 to account for the guilty data set. When the guilty data set is chosen we reject the null hypothesis that it was innocent is rejected with a $1/m$ chance or type I error of being wrong.

The lineup protocol can be used in a variety of testing scenarios for visualizations. [3] suggest the choropleth map for testing spatial structure in a data set. The lineup protocol allows for flexibility in the definition of null data. In the case of spatial visualizations, it is likely that neighbors are related. To account for this, null data sets can be generated by sampling from a known model that accounts for spatial covariance.

To contrast the effectiveness of two displays, we can produce null data plots with a hidden real data plot using the different plot designs. We can compare the time taken by participants to evaluate the same data in the different displays. We can also contrast the accuracy achieved by participants in detecting the real data plot.

III. Methodology

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

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

Additional considerations when completing this experimental task included exploration of the difficulty experienced by participants and the certainty they felt about their decision.

The mean of the detection rate for choropleth map, denoted as μ_C, and the hexagon tile maps, μ_H will be contrasted. This leads to the following one sided hypothesis:

$H_0 : \mu_H = \mu_C$ $H_a : \mu_H > \mu_C$

The detection rate $\hat{\pi}$ is calculated as the amount of people that made the choice of plot that contained the real data, out of the participants who saw data plot in the lineup of the null data plots.

A. Participants

We recruited participants for the survey style task to test the effectiveness of the hexagon tile map display. The lineup protocol expects that the participants are uninvolved judges that had no prior knowledge of the data to avoid discrimination or advantages. The Figure-Eight crowd source platform was used to advertise the survey to potential participants that had achieved level 2 or level 3 on the Figure-Eight Platform. The participants were chosen as they have experience with survey tasks that involve evaluating images. The participants were at least 18 years old, as it is a requirement of the platform.

Participants were able to choose to participate by selecting this task from the list of tasks available to them. Participants were allocated to either group A or group B when they proceeded to the survey web application, hosted externally from the Figure-Eight website. There were 95 participants involved in the study. All of these participants read the introductory materials and viewed example questions before proceeding to the survey. Each participant was trained using three test displays orienting them to the evaluation task. All participants who volunteered to take part were compensated for their time via the payment system of Figure-Eight.

TABLE I
The experimental design results in the following allocation of replicate data sets within each trend model to groups A and B.

Group	Trend	Choro.	Hex.
A	NW-SE	1, 2	3, 4
	Three Cities	1, 2	3, 4
	All Cities	2, 4	1, 3
B	NW-SE	3, 4	1, 2
	Three Cities	3, 4	1, 2
	All Cities	1, 3	2, 4

B. Variables

The variables changed between groups were the type of plot shown and the trend model.

A combination of map type and spatial trend model created the twenty-four lineup displays. We split this set of displays into a collection of twelve displays for Group A and twelve displays for Group B. The random allocation of participants to Group A or Group B resulted in 42 participants allocated to Group A and 53 participants allocated to Group B.

The levels of the factors measured in the experiment were:

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

Each group did not see the same data for both map types. We generated four simulated sets of data for each treatment. This will generate twenty-four lineups (twelve lineups of geographic maps and twelve lineups of hexagon tile maps). Participants evaluated twelve lineups, size of each map type. For each of the six geographic displays and six hexagon displays, two of each trend model were shown to the participants. We outline the design in Table. I.

The variables measured as a result of the changes were the probability of detection each display and the time taken to submit responses. To measure the accuracy of the detections, the plot chosen for each lineup evaluated was compared to the position of the real spatial trend plot in the lineup. A correct result occurs when the chosen plot matches the position of the real plot, this was recorded in an additional binary variable; 1 = correct; 0 = incorrect. High efficiency occurs when a small amount of time is taken to evaluate each lineup. This will be measured as the numeric variable measuring the length of time taken to submit the answers to the evaluation of each line up.

C. Simulation process

The lineup protocol allows a known model to be imposed on the null data set plots. The underlying spatial correlation model was created to provide spatial autocorrelation between neighboring areas using the longitude and latitude values for the Statistical Areas.

$$z = 1$$

$$locations = longitude + latitude$$

The null model imposed suggests that neighbors are related. The randomness imposed was smoothed to mirror the practice taken by the Australian Cancer Atlas. This smoothing allowed neighbors to be related to each other and show distributions similar to the Liver cancer distribution shown in **??**.

Spatially dependent data sets were simulated using the variogram model on the centroids of each geographic unit. Twelve simulations from the variogram model were created for each of the twelve lineups.

In these 12 sets of data, each of the 144 maps were smoothed several times to replicate the spatial autocorrelation seen in cancer data sets presented in the Australian Cancer Atlas.

For each of the 144 individual maps, the values attributed to each geographic area are rescaled to show a similar color scale from deep blue to dark red within each map.

A random location was selected for each set of lineup data. In this location, a trend model was overlaid on the null set of spatially correlated data. Each set of lineup data was used to produce a choropleth maps and hexagon tile maps. These matched pairs were split between Group A and Group B.

D. Experiment procudure and data collection

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

Demographics were collected regarding the study participants:

- Gender (female / male / other),
- Degree 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 evaulation phase. The set of images differed for Group A and Group B. After being allocated to a group, each individual was shown the 12 displays in randomised order.

Three questions were asked regarding each display:

- Plot choice
- Reason
- Difficulty

After completing the 12 evaluations, the participants were asked to submit their responses.

Data was collected through a web application containing the online survey. Each participant used the internet to access the survey. The data collection took place using a secure link between the survey web application and the googlesheet used to store results. The application would first connect to the googlesheet using the googlesheets [12] R package, and interacted again at the completion of the survey by adding the participant's responses to the 12 displays as 12 rows of data in the googlesheet.

E. Experimental design

The choropleth map was used as the comparative visualization for presenting the lineups [13] as this is the common display for spatial cancer data. Geographic distributions usually have some degree of spatial autcorrelation between neighbors. This feature was incorporated in all maps shown in the lineup displays, the map that contained the trend feature shown in only one set of data was also affected by spatial autocorrelation. A line up protocol was implemented to arrange N maps in each display. A reasonable amount of null plots $N-1$ in the lineup was chosen to ensure the real data map was well hidden. A reasonable number of plots to show in each lineup, $N = 12$ was chosen to not overwhelm participants due to the detailed choropleth maps of Australian SA3 areas.

The hypotheses for each lineup are H_0 : All plots look the same H_a : One plot looks different to the other plots

The same data was visualized on a choropleth map, and on a hexagon tile map, however in this study the participants only saw one of these two displays. The accuracy and times taken were aggregated for Groups and. Comparing the results of participants who see the choropleth to those who see a hexagon tile map will show that population related distributions are spotted more frequently in a hexagon tile map display.

Let n be the number of independent observers and x_i the number of observers who picked plot i, $i = \{1, ..., m\}$

Then $x_i, x_2, ..., x_m$ follows a multinomial distribution$Mult_{\pi_1,\pi_2,...,\pi_m}(x_i, x_2, ..., x_m)$ with $\sum_i \pi_i = 1$, where π_i is the probability that plot i is picked by an observer, which we can estimate as $\hat{\pi}_i = x_i/n$. The researchers compared the length of time taken, and the accuracy of the participants choices. The power of a lineup can therefore be estimated as the ratio of correct identifications x out of n viewings.

F. The methods of data analysis used

The data analysis methods used in order to analyse and collate the results included downloading the survey submissions and opening them into the analysis software R [14].

For each of the 12 lineup displays the researchers calculated:

- accuracy: the proportion of subjects who detected the data plot
- efficiency: average time taken to respond

1) visualizations: Side-by-side dot plots were made of accuracy (efficiency) against map type, facetted 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.

Plots will be made in R (R Core Team 2019), with the ggplot2 package (Wickham 2016).

2) Modeling: The likelihood of detecting the data plot in the lineup can be modelled using a linear mixed effects model. The R [14] `glmer()` function in the [15] 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:

$$\widehat{y_{ij}} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} + \epsilon_{i,j}, \quad i = 1, 2; \ j = 1, 2, 3$$

where $y_{ij} = 0, 1$ whether the subject detected the data plot, 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. Because the response is binary, a logistic model is used. This model can account for each individual participants' abilities as it includes a subject-specific random intercept. As each participant provides results from 12 lineups.

The model specifies a logistic link, this means the predicted values from the `glmer` model should be back-transformed to fit between 0 and 1. They are transformed with the link specified below:

$$\mu = \frac{e^{\eta}}{1 + e^{\eta}}$$

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

The feedback and demographic variables will possibly be incorporated as covariates.

Computation will be done using R [14], with the `lme4` package [15].

G. Limitations of the data collection

A pilot study was conducted to determine whether the lineups were appropriate. All lineups in the pilot study were deemed viable as at least one participant detected the real data plot in each lineup. The demographics of the participants showed a skew towards male participants. The randomness of the group allocation also resulted in more participants being allocated to Group B. Due to the allocation of lineup displays the participants all saw six Choropleth displays and six Hexagon Tile Map displays.

IV. Results

Responses from 95 participants were collected. Three participants provided no answers for any task, 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 are analysed on accuracy and speed. The certainty and reasons of subjects in their answers is also examined.

Fig. 1. 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 haxgon tile map display are linked to show the difference in the detection rate.

A. Participant demographics

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

B. Accuracy

Fig. 1 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 Table. I; 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 plot type. Surprisingly, participants could also detect the gradual spatial trend in the NW-SE group from the hexagon tile map. We expected that the choropleth map would be superior for the type of spatial pattern, but the data suggests the hexagon tile map performs slightly better, or equally as well.

Table. II shows the means and standard deviations of the detection rate for each type of plot and each trend model. This also gives the standard errors, the smallest standard deviation for all sets of replicates is the Three Cities trend model shown in a Choropleth display. This group of displays had a very small detection rate of 0.04. The mean detection rate for the Three Cities trend model shown as choropleth map lineups was also the smallest at 0.39. The North-West to South-East (NW-SE) trend model unexpectedly had a higher mean detection rate for the hexagon tile map displays, but the difference in the means of detection rate was only 0.10.

Table. III presents a summary of the generalised linear mixed effects model, testing the effect of plot type and trend model on the detection rate. The results support the summary

TABLE II
THE RATE OF DETECTION FOR EACH TREND MODEL HAS BEEN CALCULATE FOR THE CHOROPLETH AND HEXAGON TILE MAP DISPLAYS. THE ASSOCIATED STANDARD ERRORS ARE ALSO INCLUDED.

Type	NW-SE	Three Cities	All Cities
Choro.	0.51 (0.50)	0.04 (0.19)	0.23 (0.42)
Hex.	0.61 (0.49)	0.39 (0.49)	0.57 (0.50)

TABLE III
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	0.02		0.17	0.90
Hex.	0.46	*	0.22	0.04
Three Cities	-3.43	***	0.42	0.00
All Cities	-1.35	***	0.24	0.00
Hex:Three Cities	2.46	***	0.47	0.00
Hex:All Cities	1.17	***	0.33	0.00

from Fig. 1 and all terms are statistically significant despite the large standard deviations observed in Table. II. 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 choroleth map lineup, but increases for a hexagon tile map display. The log odds of detection show in Table. III 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 0.62 from 0.51 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 0.51. When the All Cities trend is presented in a choropleth display the predicted probability of detection is 0.21. The hexagon tile map has a substantially high detection rate for the display of a Three Cities trend 0.38 and All Cities trend 0.58.

C. Speed

Fig. 2 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,

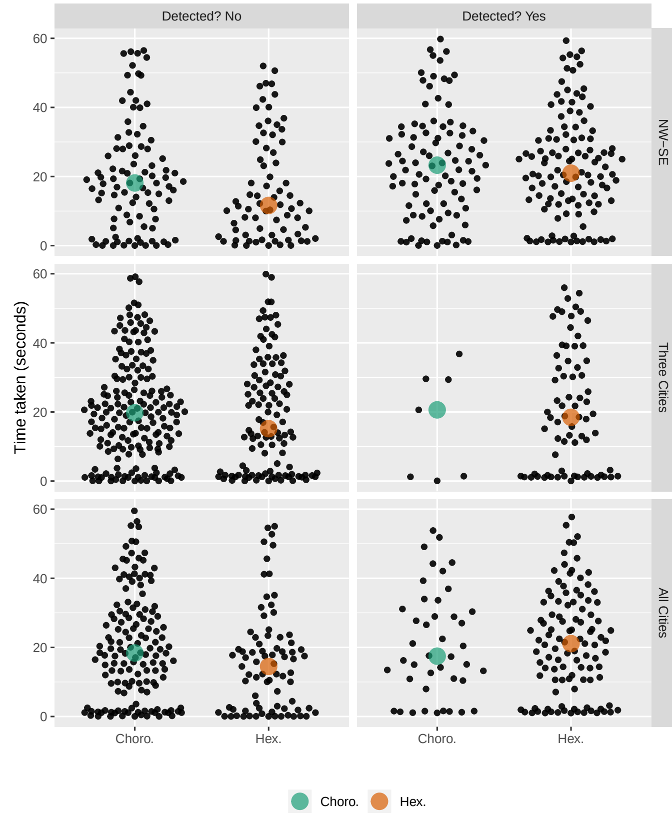


Fig. 2. The distribution of the time taken (seconds) to submit a response for each combination of trend, whether the data plot was detected, and type of display, shown using horizontally jittered dotplots. The colored point indicates average time taken for each plot type. Although some participants take just a few seconds per evaluation, and some take as much as 60 seconds, but there is very little difference in time taken between plot types.

there is little difference in the average time taken to read a choropleth or hexgon 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.

D. Certainty

Participants provided their level of certainty regarding their choice using a five point scale. Unlike the accuracy and speed of reponses 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. IV-D 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 chose 4 or 5 often when viewing the population related trends in the choropelth display, even though they were often incorrect when viewing an All Cities

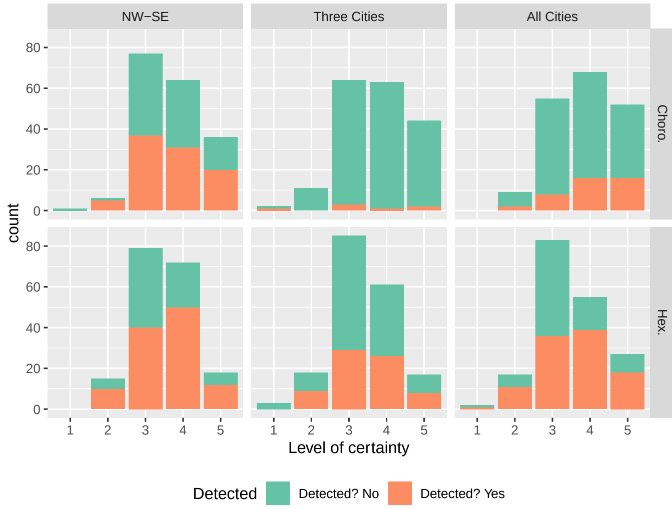


Fig. 3. The amount of times each level of certainty was chosen by participants when viewing hexagon tile map or choropleth displays. Participants were more likely to choose a high certainty when considering a Choropleth map. The mid value of 3 was the default certainty, it was chosen most for the Hexagon tile map displays.

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.

E. 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.

The reasons used when viewing choropleth displays varied more than the hexagon tile map reasons. 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.

V. 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.

TABLE IV
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.
NW-SE	No	trend	clusters
	Yes	trend	clusters
Three Cities	No	trend	clusters
	Yes	consistent	clusters
All Cities	No	trend	clusters
	Yes	clusters, consistent	clusters

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 compare detect plot and Table. II are reflected in the model results. Surprisingly the difference scene 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 disciplines 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 the Fig. 2 display showed very little difference in the mean evaluation times. As the ranges 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. Fig. **??** 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 wasy 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 consisent 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.

VI. 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 exaserbates 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 others. 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.

VII. Acknowledgment

The authors would like to thank the Australian Cancer Atlas team for discussions regarding alternative spatial visualizations. They would also like to thank Kerrie Mengersen and Dr. Earl Duncan for suggestions and comments.

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.

VIII. References

IX. Acknowledgment

The authors would like to thank the Australian Cancer Atlas team for discussions regarding alternative spatial visualizations. They would also like to thank Distinguished Professor Kerrie Mengersen and Doctor Earl Duncan for suggestions and comments.

The source code to produce this document can be found on GitHub. Supplementary materials have been included to discuss the survey procedures,

The analysis of the work was completed in R [14] with the use of the following packages:

- For document creation: rmarkdown [16], rticles [17], knitr [18].
- For lineup creation and data analysis: tidyverse [19], nullabor [20], ggthemes [21], RColorBrewer [22].
- For image displays: cowplot [23], png [24], grid [25].
- For modeling and presentation of models: lme4 [15], kableExtra [26].

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.

X. References

1. Kobakian S, Cook D (2019) Sugarbag: Create tessellated hexagon maps.

2. Queensland CC Australian Cancer Atlas (https://atlas.cancer.org.au).

3. 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

4. Tufte ER (1990) Envisioning Information. Graphics Press

5. Skowronnek A (2016) Beyond Choropleth Maps – A Review of Techniques to Visualize Quantitative Areal Geodata. In: Infovis Reading Group WS 2015/16. https://alsino.io/static/papers/BeyondChoropleths_AlsinoSkowronnek.pdf.

6. Kocmoud C, House D (1998) A Constraint-based Approach to Constructing Continuous Cartograms. In: Proc. Symp. Spatial data handling. pp 236–246

7. Dorling D (2011) Area Cartograms: Their Use and Creation. In: Concepts and techniques in modern geography (catmog). pp 252–260

8. Olson JM (1976) Noncontiguous Area Cartograms. The Professional Geographer 28:371–380

9. Raisz E (1963) Rectangular Statistical Cartograms of the World. Journal of Geography 35:8–10

10. Monmonier M (2005) Cartography: Distortions, Worldviews and Creative Solutions. Progress in Human Geography 29:217–224

11. 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

12. Bryan J, Zhao J (2018) Googlesheets: Manage google spreadsheets from r.

13. Majumder M, Hofmann H, Cook D (2013) Validation of visual statistical inference, applied to linear models. Journal of the American Statistical Association 108:942–956

14. R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

15. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48

16. Xie Y, Allaire JJ, Grolemund G (2018) R markdown: The definitive guide. Chapman; Hall/CRC, Boca Raton, Florida

17. Allaire J, Xie Y, R Foundation, et al (2019) Rticles: Article formats for r markdown.

18. Xie Y (2014) Knitr: A comprehensive tool for reproducible research in R. Implementing reproducible computational research

19. Wickham H, Averick M, Bryan J, et al (2019) Welcome to the tidyverse. Journal of Open Source Software 4:1686

20. Wickham H, Chowdhury NR, Cook D, Hofmann H (2018) Nullabor: Tools for graphical inference.

21. Arnold JB (2019) Ggthemes: Extra themes, scales and geoms for 'ggplot2'.

22. Neuwirth E (2014) RColorBrewer: ColorBrewer palettes.

23. Wilke CO (2019) Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'.

24. Urbanek S (2013) Png: Read and write png images.

25. R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

26. Zhu H (2019) KableExtra: Construct complex table with 'kable' and pipe syntax.