# Visual Inference Test of the Hexagon Tile Map for Spatial Distributions

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 abstract goes here. On multiple lines eventually.**

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

## INTRODUCTION

Geo-spatial statistics are often presented on the geographic map base. To present geo-spatial population statistics, information for individuals within each geographic region are often aggregated. A choropleth map is the common display to present aggreagated statistics for geographic units, and they are often used to present statistics regarding the population. This visualisation method involves drawing the administrative boundaries and filling them with colour to communicate the value of the statistic. In Australia, there are many sets of administrative boundaries that define subdivisions of the population at various granularities.

When presenting population statistics on a geographic map base, the size of the regions can allow errornous conclusions to be drawn about the state of the statistic over the entire population. This occurs as large regions filled with a consistent colour or pettern can draw the attention of map readers, and small regions are not paid equal attention.

## BACKGROUND AND MOTIVATION

### *Australian Cancer Atlas*

- Communicating spatial distributions
- Trend over geographic space
- Trend over communities and populations

The Australian Cancer Atlas explores the burden of cancer on Australian communities. There are many cancer types presented, and they can be explored on an individual or aggregate level. The Australian communities examined are Statistical Areas at Level 2 (SA2)("Australian Statistical Geography Standard (ASGS)" 2018) used by the Australian Bureau of Statistics. Bayesian spatial smoothing has been applied to incorporate the statistics of neighbouring areas, for both privacy and stability of the estimates. The statistics that can be mapped are the diagnoses (Standardised Incidence Rates) and excess deaths for each SA2, communicated as the difference from the Australian average of the statistics. The values of the statistic for each are communicated using a diverging colour scheme. Dark blue represents areas with values much less than the Australian average, and represents areas much greater than the Australian average.

### *Visual Inference*

- Communicating data through visualisations
- Effective displays for types of data
- Testing the effectiveness

Classical statistical inferences involves hypothesis testing, the process of rejecting a null hypothesis in favour of an alternative. This approach relies on data, the appropriate distributions and their assumptions.

### *Line up protocol*

The lineup protocol presents a method for visual inference testing.

> "In this framework, plots take on the role of test statistics, and human cognition the role of statistical tests." Buja et al. (2009)

The line up protocol involves placing a "guilty" data visualisation in a lineup of "innocents". Where the guilty data set contains structure, and the innocents are equivalent to a null data set. In a grid of visualisations, 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 group innocents. The guilty data is identified as different from the innocent data with probability $1/m$, where $m$ is the number of null plots plus 1 to account account for the guilty data set. When the guilty data set is chosen, 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 tsting scenarios. The choropleth map is best used for testing spatial structure in a data set.

### *Population focussed displays*

Map creators have the ability to present spatial statistics in alternative displays that can highlight the population. This work aims to show that a hexagon tile map display is a viable alternative to the geographic map base for presenting population statistics. The same data were shown on a choropleth map, and on a hexagon tile map. 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.

## Study Design

### *Aims*

This study aims to answer several 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?
3. Do people find hexagon tile maps more difficult to read than choropleth maps?
4. Are the reasons for choosing a plot different depending on the type of display?
5. Does an Australian resident find the choropleth map easier to read than the hexagon tile map?

### *Methodology*

A survey was created to test the effectiveness of the hexagon tile map display.

The online crowdsource platform Figure-Eight was used to recruit participants. A line up protocol was implemented to arrange 12 maps in each display. Individual displays were created by a combination of plot type, and spatial trend model.

The researchers contrasted the different plot designs, as hexagon tilemap and geography in the lineups were created using the same data, and same null positions within the lineup.

The researchers compared the length of time taken, and the accuracy of the participants choices.

### *The participants*

There were 95 participants in the study. We recruited participants using the Figure-Eight crowd source platform by advertising this survey to participants that fulfilled the following crtieria:

• level 2 or level 3 on the Figure-Eight Platform. • at least 18 years old

Participants then selected our task from the list of tasks available to them.

### *The variables being manipulated and measured?*

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

Each participant was randomly allocated to either group A or group B when they begun the survey. This 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: • Plot type: *Geography, Hexagons* • Trend: *Locations in two population centres, Locations in multiple population centres, South-East to North-West*

Factor combinations examined by each participant amount to 6 (2x3) lineup displays. A participant did see the same data for both plot types. Four simulated sets of data were generated for each treatment. This will generate 24 lineups (12 were geographic maps, and 12 were hexagon tile maps). Participants will evaluate 12 lineups, 6 of each plot type. Appendix A shows the experimental design visually. For each of the six geographic displays and six hexagon displays, two of each trend model were shown to participants.

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.

### *Participant training*

Each participant saw three test displays orienting them to the task. Participants proceeded to the survey, this involved evaluating 12 displays.

### *Experiment procudure and data collection*

The participant submitted 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 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 (Bryan and Zhao 2018) 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.

### *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 (**???**).

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

*Visualisations:* Side-by-side dot plots were made of accuracy (efficiency) against plot 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).

*Modeling:* The results will be analysed using a generalised linear model, with a subject random effect. There will be two main effects: plot type and trend model, which gives the fixed effects part of the model to be

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

where $y_{ij} = 0, 1$ whether the subject detected the data plot, $\mu$ is the overall mean, $\tau_i, i = 1, 2$ is the plot type effect, $\delta_j$ is the trend model effect. We are allowing for an interaction between plot type and trend model. Because the response is binary, a logistic model is used.

A similar model will be constructed for the efficiency, using a log time, and normal errors.

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

Computation will be done using R (**???**), with the `lme4` package (**???**).

### *Limitations of the data collection*

This required internet connection for participants to access the survey

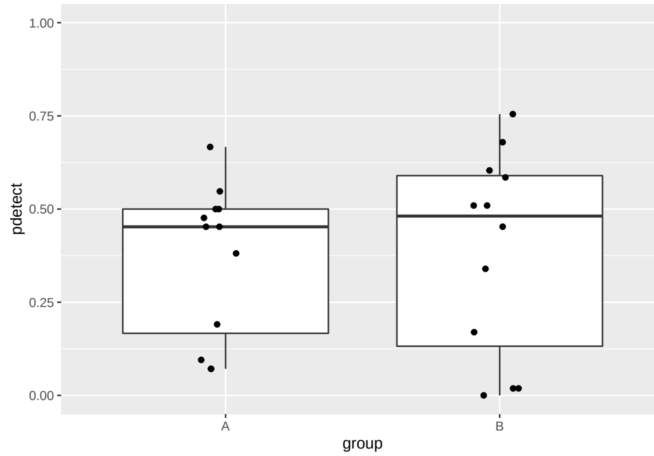
Ethics approval for the online survey was granted by QUT's Ethics Committee (Ethics Application Number: 1900000991). All applicants provided informed consent in line with QUT regulations prior to participating in this research.

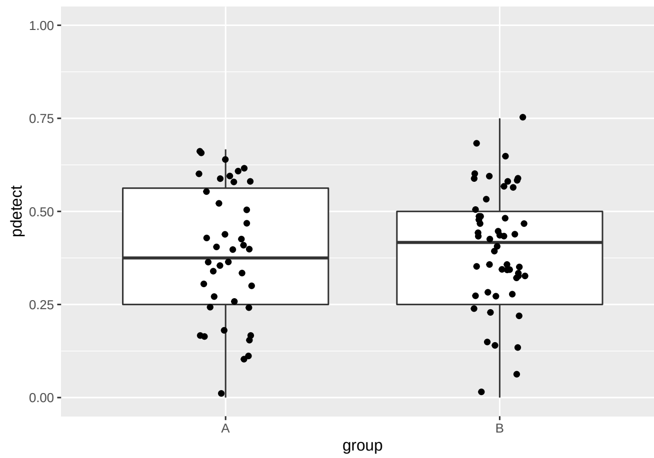
## Results

1. Type

2. Trend
3. Interaction of type and trend
4. Demographics of participants

The survey responses from participants were kept only if the participant submitted answers for all 12 displays. This resulted in 95 surveyed.



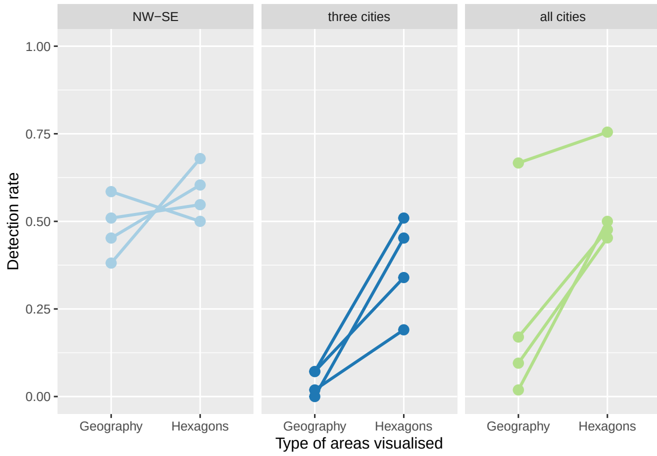



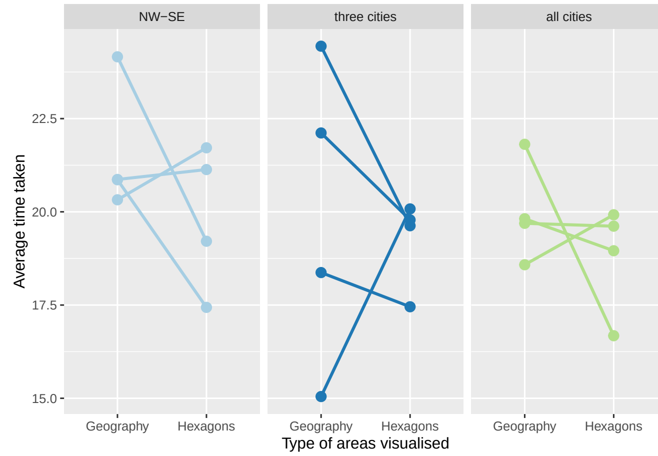

```
## # A tibble: 2 x 2
##   gender     n
##   <chr>  <int>
## 1 He        70
## 2 She       25
```

```
## # A tibble: 6 x 2
##   age         n
##   <chr>   <int>
## 1 18 - 24    15
## 2 25 - 34    37
## 3 35 - 44    23
## 4 45 - 54    11
## 5 55+         6
## 6 NA          3
```

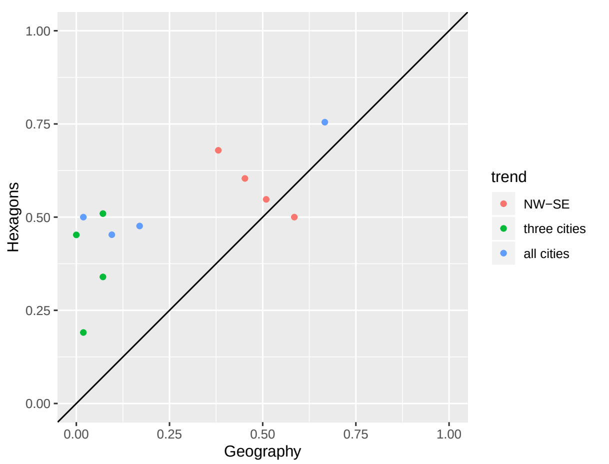
```
## # A tibble: 3 x 2
##   education       n
##   <chr>       <int>
## 1 Bach           56
## 2 High School    25
## 3 Masters        14
```

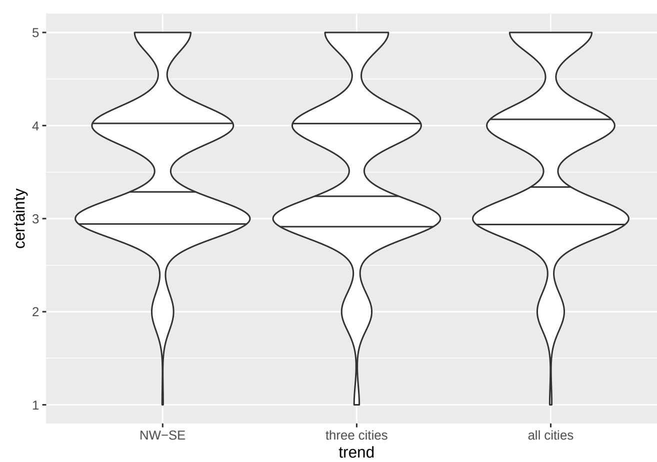
```
## # A tibble: 2 x 2
##   australia     n
##   <chr>     <int>
## 1 No           93
## 2 Yes           2
```



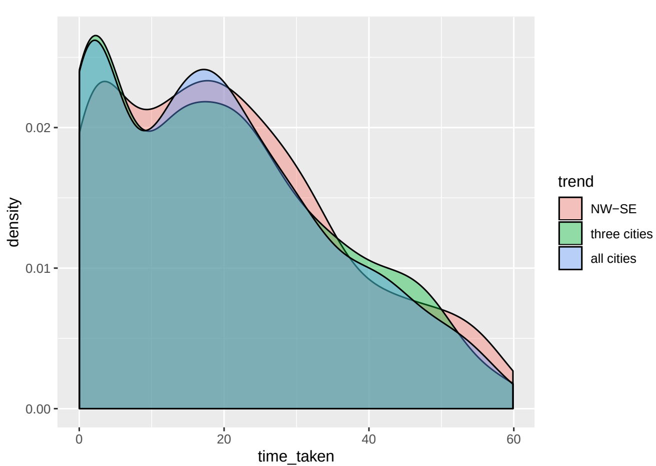



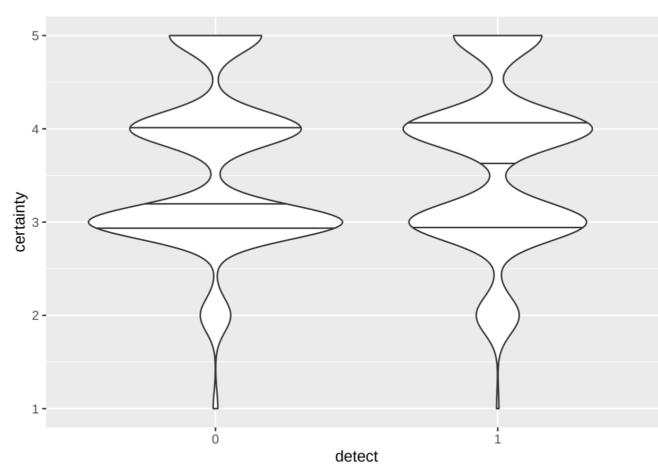

11 of the 12 real distribution plots were found more often in the hexagon display. This was better than expected. As even geographic distributions were spotted in the hexagon lineup.

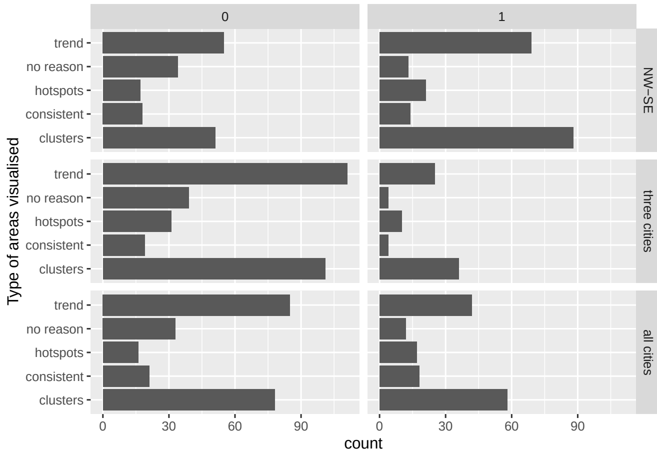



```
## [1] 0.004079185
```






*Probability of detection:*

*Time taken*

*Modeling*

## Discussion

## Conclusion

The conclusion goes here.

## Acknowledgment

The authors would like to thank…

## Bibliography styles

## References

"Australian Statistical Geography Standard (ASGS)." 2018. *Australian Bureau of Statistics*. Australian Government. {https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian Statistical Geography Standard (ASGS)}.

Bryan, Jennifer, and Joanna Zhao. 2018. *Googlesheets: Manage Google Spreadsheets from R*. https://CRAN.R-project.org/package=googlesheets.

Buja, Andreas, Dianne Cook, Heike Hofmann, Michael Lawrence, Eun-Kyung Lee, Deborah F. Swayne, and Hadley Wlckham. 2009. "Statistical Inference for Exploratory Data Analysis and Model Diagnostics." *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 367 (1906): 4361–83. http://www.jstor.org/stable/40485732.