Project Protocol

Contents

1 Title

Assessing the effectiveness of hexagon tile maps for communicating spatial distributions of disease for Australia.

2 Project team roles & responsibilities

Researcher	Role	Affiliation
Stephanie Kobakian	Masters Student	Faculty of Science and Engineering (QUT)
Prof. Kerrie Mengersen	Principal Supervisor	Faculty of Science and Engineering (QUT)
Dr. Earl Duncan	Associate Supervisor	Faculty of Science and Engineering (QUT)
Prof. Dianne Cook	External Supervisor	Department of Econometrics and Business Statistics, Monash University

3 Background information

3.1 Project outline

This project will test the effectiveness of two types of spatial displays: a choropleth map, and a hexagon map. It will contrast the effectiveness of communicating information when each geographic region is represented by a hexagon, rather than its geographical shape. The purpose of these displays is to convey the spatial distribution of the disease occurrence, or incidence. This can allow for detecting hot spots corresponding to outbreaks, or spatial trends, for example, indicating occurrence is related to rural vs urban differences or localised to a single population dense region. Effectiveness of the display will be measured by accurate and efficient detection of these patterns.

3.2 Introduction/background information

Dorling (2011) suggests alternative map displays that aim for 'equitable representation design' present each member of a society equally. They are considered more 'socially just' but also help to present an honest display of the experiences of the whole population (Dent 1972). This work aims to determine if using a coloured hexagon display is a more effective method of communicating the spatial relationship between a set of areas on a map.

3.3 Rationale/justification

This work is motivated by the newly developed Australian Cancer Atlas where information regarding cancer incidence and mortality is provided to the community through a web page. We are interested to know whether providing the spatial distribution in a hexagon map is more effective for communication than the traditional choropleth map. The responses of participants will be contrasted when shown the same information in both a geographic and hexagon map display and asked to find the same real distribution in a collection of maps arranged in a line up. The rate of participants who pick the maps in the line ups that display the true spatial relationship will be used to contrast the effectiveness of the displays. We hope that this will show a hexagon map is an effective method for communicating spatial relationships.

4 Study objectives

4.1 Hypotheses

A hexagon tile map display allows people to detect the impact of a disease in highly populated small areas with higher accuracy and efficiency, in comparison to the accuracy and efficiency of detections made when viewing a choropleth map.

4.2 Research questions/ aims

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?

4.3 Outcome measures

For each plot evaluation the subject will provide these responses:

- Their choice of plot from a lineup
- Reason for choice of plot
- Time taken to respond
- Perceived difficulty in making a choice

5 Study design

Each participant will first see three test displays, they will then undertake an online survey linked to them through the online crowdsource platform Figure-Eight. Within the experiment: A line up protocol will be used to arrange 12 maps into one display. Each display will be created by a combination of plot type, spatial error distribution and spatial trend model. We will contrast the sets of different plot designs, hexagon tilemap and geography in the lineups created using the same data, and same null positions within the lineup. We will compare the length of time taken and the accuracy with which participants report their choice.

The levels of the factors are:

- Plot type: Geography, Hexagons
- Trend: Locations in two population centres, Locations in multiple population centres, South-East to North-West

Factor combinations to be examined by each participant amount to 6 (2x3) lineup displays. A participant cannot see the same data for both plot types. Four simulated sets of data will be generated for each treatment. This will generate 24 lineups (12 will be geographic maps, and 12 will be hexagon tile maps). Participants will evaluate 12 lineups, 6 of each plot type. Appendix A shows the experimental design visually.

6 Study population

6.1 Participants

We plan to collect responses from 100 participants.

6.2 Inclusions and exclusion criteria

The platform will advertise this survey to participants that fulfill the following:

- level 2 or level 3 on the Figure-Eight Platform.
- at least 18 years old
- trained and qualified to annotate images.

6.3 Recruitment strategies, timeframe

Recruitment will take place via advertising on the Figure-Eight site. On this site the available tasks like this survey, are advertised to participants by the addition to the list of tasks available to contributors on the platform. Participants must select our task from the list to read the description of the task, and what is expected of them to be able to participate.

Removed: [Recruitment will take place via advertising on the Figure-Eight dashboards. The Figure-Eight platform will advertise this as a "job" to potential participants, they will not be under any obligation to complete this task. We keep advertising until enough participants have completed the survey, then remove the advertisements and close the survey.]

6.4 Consent approaches

Participants will be shown instructions which will provide sufficient information about the task when they select to do the survey. These instructions will explain the survey to participants, and give them an example of the questions they will answer. Contact information for the researchers will be made available to the participants. Participants will be asked to provide consent via a checkbox question. They will then continue to the survey, and complete the test questions. This page will also explain the use and storage of the data they contribute. The data will be the answers they provide, which will be publicly available on github at https://github.com/srkobakian/experiment.

6.5 Participant withdrawal

Participants can withdraw from the experiment at any time by quitting the survey, they with also be withdrawn if they do not provide a choice for more than 50% of the displays.

7 Procedures

7.1 Screening of participants

Participants will need to have achieved contributor level two on the Figure-Eight platofrm, this is a smaller group of more experienced, higher accuracy contributors.

Test questions are the most important quality control mechanism in the Figure-Eight Platform. Participants will complete them to practice and prepare for the experiment questions. in our analysis stage, to use the responses from a participant they will need to reach an accuracy threshold of at least one correct choice.

Removed: [Test questions are the most important quality control mechanism in the Figure-Eight Platform. Participants will complete them to practice and prepare for the experiment questions, they will need to reach an accuracy threshold we will set, to be allowed to participate in answering the experiment survey questions.

The Figure Eight procedures are followed: *Contributors that fail quiz mode are not paid and are permanently disqualified from working on your job. Participants will need to achieve a pre-determined accuracy threshold to be allowed to answer the questions in our survey.*]

7.2 Data collection

As the survey is conducted through the Figure-Eight platform, which will record IP Address of computer used to access the survey, Contributor ID, Trust Score and Channel. In addition, we will record the responses of participants:

- Choice of plot in each line up

- Reason for choice of plot
- Time taken to answer
- Perceived difficulty of question

and participants will be asked provide this demographic information:

- 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)

The experimental design variables, Plot Type, Trend Model and simulation number will also be saved for each display shown to participants.

7.3 Data collection/gathering techniques

Data will be collected through survey distribution via a web application. Information and illustrative displays regarding the survey have been included in the word document 01_O_ETH_Survey_Data_Collection_Tool_V1_20191023. Participants will access the application via a link from the Figure-Eight task site. This will lead them to the web page that currently hosts the pilot study: https://ebsmonash.shinyapps.io/dingo/

7.4 Impact of, and response to, missing data

Participants have the option to not respond to any question.

If there are 50% or more of unanswered survey questions (excluding demographic information) we will remove this participant's responses before data analysis.

8 Outcome measures

8.1 Accuracy

To measure the accuracy of the detections, the plot chosen for each lineup evaluated will be 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 will be recorded in an additional binary variable; $1 = correct$; $0 = incorrect$.

8.2 Efficiency

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.

9 Statistical plan

The design of this study is based on work by Hofmann, Follett, Majumder and Cook (2012).

9.1 Sample size determination and power

To be able to effectively evaluate the use of a hexagon tilemap we will contrast the proportion of times the plot of the data was chosen from the field of null plots.

The same data will be used to create 11 null data plots and one data plot. This data will be displayed to users in both geographic maps and hexagon tile maps.

We will have aim to collect results from 100 participants ($n = 100$), who been asked to choose the plot that is most different from the others. Each participant will evaluate 12 lineups.

By chance, the data plot will be detected with probability, $1/12 = 0.083$. Assuming a binomial distribution we can determine the critical value that would indicate the number of detections from n evaluations, corresponding to a rejection of $H_0 : p = 0.083$. For example, with 50 evaluations, with significance level 0.05, if there are 8 or more detections, it is better than chance, say C.

This critical value can be used to estimate the sample size needed, for the study, by computing the power, the probability of rejecting $H_0 : 0.083$, for alternative values of $p = p_A$. If we expect the data to be detected twice as often as by chance, $p_A = 1/6$. To determine the probability of observing C detections from n evaluations, for one lineup, to calculate the power:

$$P(X \geq C | p = p_A, n) = \sum_{j=C}^{n} \binom{n}{j} p_A^j (1 - p_A)^{n-j}$$
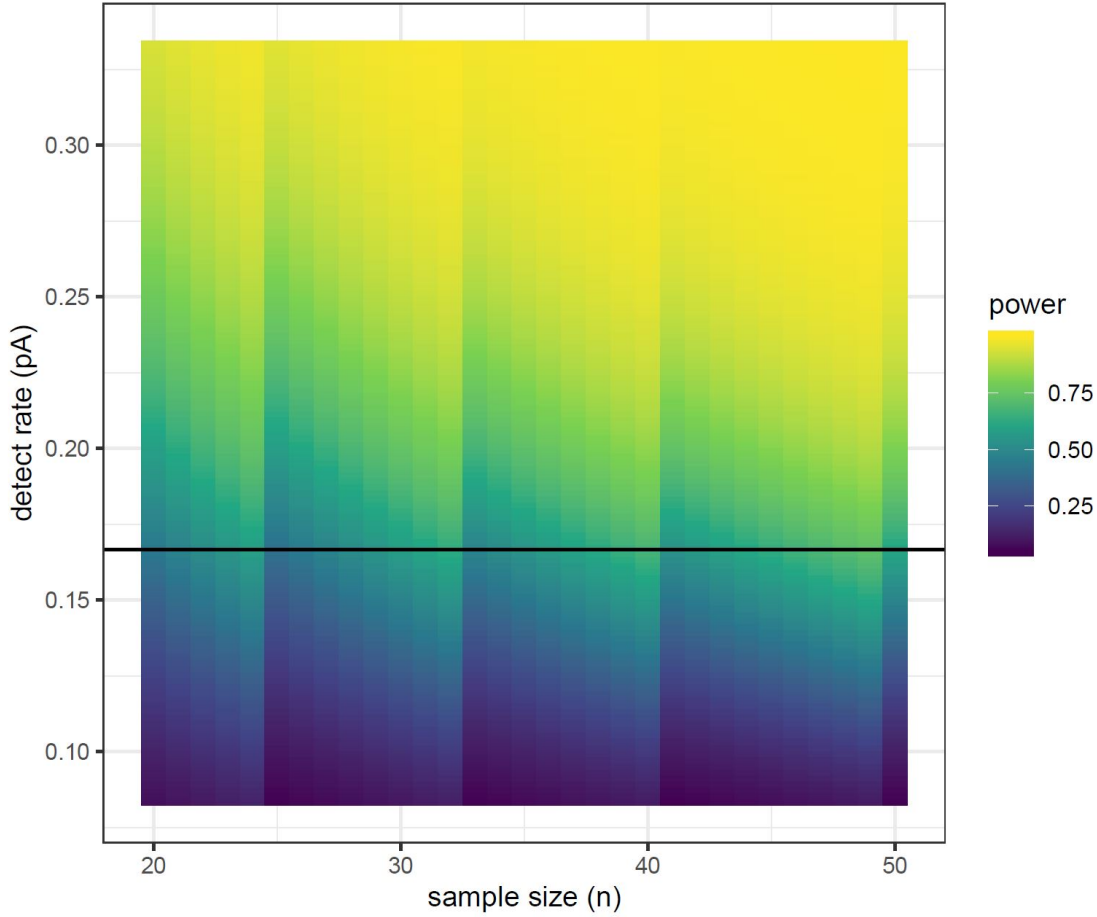
n	k	pA	prob
48	7	0.167	0.713
49	7	0.167	0.734

The table above shows the sample sizes yielding above 70% chance of detection (p_d) in a single lineup assuming

6

$p_A = 1/6 = 0.167$ we would need approximately 50 evaluations for each lineup.

If 4 lineups are evaluated for each trend model, the chance that none produce a detection is $1 - (1 - p_d)^4$, which is effectively 0. With 100 participants, giving 50 evaluations per lineup, if there is an effect size twice as high as chance, then we should see it in the results.

9.2 Data analysis and statistical methods

9.2.1 Graphical analysis

For each lineup we will compute

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

Side-by-side dot plots will be made of accuracy (efficiency) against plot type, facetted by trend model type.

Similar plots will be 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).

9.2.2 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, μ is the overall mean, $\tau_i, i = 1, 2$ is the plot type effect, δ_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 (R Core Team 2019), with the `lme4` package (Bates et al. 2015).

10 Data management and record keeping

Research expectations today require that publications use reproducible research protocols, requiring data to be available. The external survey data will be made available via QUT's Research Data Finder, on github, and, if required by the publisher, once the results are published also available on the journal article web site.

10.1 Confidentially and privacy

Datasets produced from the project and that will be available for public access on github will be de-identified and only contain the responses to survey, and information regarding the participant demographics: age range, education level, gender, and whether they have lived in Australia. If there are participants with unique sets of demographics, their data will be combined with larger categories to minimise the risks of reidentification.

Responses to the survey will be stored securely using a google sheet owned by Stephanie's Monash University google account drive, accessible only to the researchers during the study through their password protected google drive accounts. This will contain the demographic information, at the coarse resolution specified in the data collection.

Removed: [Datasets produced from the project and that will not be available for public access will contain original demographic information about the participants and only the anonymous Figure-Eight contributor ID.

Datasets produced from the project and that will be available for public access on github will be de-identified and only contain the responses to survey. If there are participants with unique sets of demographics, their data will be combined with larger categories to minimise the risks of reidentification.]

10.2 Data security

Responses to the survey will be stored securely using a private google drive accessible only to the researchers during the study through their password protected google drive accounts. This will contain the demographic information, at the coarse resolution specified in the data collection.

Removed: [Responses to the survey will be stored securely using a private google drive accessible only to the researchers during the study through their password protected google drive accounts. This will contain the demographic information, at the coarse resolution specified in the data collection.]

10.3 Record retention

To ensure records are retained during and following the study the responses will be stored on the google sheet immediately following the submission by participants. At the conclusion of the survey, the data will remain in the password protected google drive account.

Immediately following the submission by participants. At the conclusion of the survey, the data will remain in the password protected google drive account and on github During the analysis phase the researchers Stephanie and Dianne will make copies of the survey responses to their work devices (laptop computers), that are password protected.

10.4 Secondary use

Research expectations today require that publications use reproducible research protocols, requiring data to be available. The external survey data will be made available will be made available via QUT's Research Data Finder, on github, and, if required by the publisher, once the results are published also available on the journal article web site.

The analysis scripts and the data will be available for secondary use by other researchers to reproduce these results and also use in future studies. Participants will be asked to provide consent for this usage in the consent form.

11 Resources

Each participant will earn $AUD5 for their work.

12 Results, outcomes and future plans

12.1 Plans for return of results of research to participants

There is no intention to provide the results to participants. This might compromise their ability to participate in future experiments.

12.2 Publication plan

We intend to publish this work as a research article in a top journal, along with de-identified data as required by the journal's reproducibility policy.

12.3   Other potential uses of the data at the end of the project

We do not intend to re-use this data, we don't see any major reasons for others to use this data. It may be possible that the data may be aggregated with many other studies to examine effects of education, age and gender on reading data plots.

12.4   Project closure processes

The project will be closed once the journal article is published.

## 12.5   Plans for sharing and/or future use of data and/or follow-up research

Details are provided in the Data Management Plan. The results from this study might inform design of future studies.

13   Appendices

13.1   Appendix A

The experimental design plan shows the distribution of the null data sets to each type of trend model that will be tested. It also states the location of the trend model in the lineup of 12 plots in a three row, four column grid.
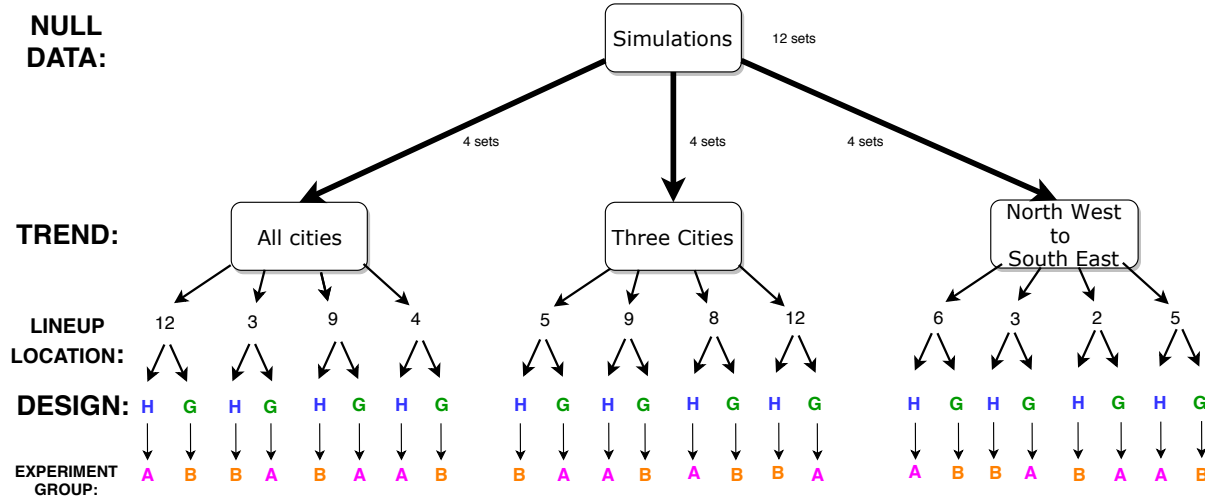
References

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Dent, Borden D. 1972. "A Note on the Importance of Shape in Cartogram Communication." *Journal of Geography* 71 (7): 393–401. https://doi.org/10.1080/00221347208981697.

Dorling, Daniel. 2011. "Area Cartograms: Their Use and Creation." In *Concepts and Techniques in Modern Geography (Catmog)*, 59:252–60. https://doi.org/10.1002/9780470979587.ch33.

Hofmann, H., L. Follett, M. Majumder, and D. Cook. 2012. "Graphical Tests for Power Comparison of Competing Designs." *IEEE Transactions on Visualization and Computer Graphics* 18 (12): 2441–8. https://doi.org/10.1109/TVCG.2012.230.

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

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.