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 subdivisions of the geography, 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 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. XXX

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 can be explored on an individual or aggregate level. The Australian communities examined are Statistical Areas at Level 2 (SA2) 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 will be 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.

Methodology

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

Experimental Design: Three trends were tested, in each of the two displays. For each of the

Line up protocol:

- Experimental design
  - The overall structure and operation of the experiment or observational experience.
  - The subject populations: The groups studied in the research including the size of each group and any features of the subjects which may be relevant to the topic being researched.
  - The variables that were changed between groups and the variables measured as a result of the changes.
  - The conditions under which the research was undertaken and any factors or variations in conditions which may have an impact on the results.
  - The methods of data analysis used in order to analyse and collate the results.
  - Any limitations of the data collected

Results

1. Type

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

Probability of detection:

Time taken

Modeling

Discussion

Conclusion

The conclusion goes here.

Acknowledgment

The authors would like to thank...

Bibliography styles

Here are two sample references: (**???**; **???**).

References

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.