Chapter 1: Introduction

Stephanie

Introduction

There are many visualisation methods used to present geospatial data. The design of the visualisation chosen can hinder or improve the communication of the spatial distribution. A choropleth map is the most common display used to present geographical data.Maps contribute to understanding spatial distributions of disease occurrence, and locating disease clusters. Disease data is often aggregated by political areas. One reason for this is privacy, another the responsibility on the political entity to respond. The typical visualisation for aggregated spatial data is a choropleth map, where areas are coloured by the numerical value.

Choropleth maps do a disservice to the map reader, as the attention of the map user is distributed according to the size of the area. Using a choropleth map to get a broad perspective of Australia can be misleading, when the use of geographical areas misrepresents the spatial distribution of a dataset. This is not practical if each area is considered equally important. In Australia, the population is not equally dispersed across the geographic map base. Instead, the communities are densely populated in the inner city areas, especially around the capital cities. There are several visualisation methods that have been developed to emphasise the population dense areas. These alternatives should be considered when planning the communication of geospatial statistics. Visualisations should be chosen to best represent the spatial distribution, The work is motivated by the Cancer Atlas of Australia, which presents the spatial patterns of many cancers in Australia. The aim of this thesis is to contribute an algorithm that creates effective visualisations for the communication of geospatial population statistics.

The Australian Cancer Atlas

This work was motivated by the Australian Cancer Atlas. An online, interactive web tool for exploring the impact of cancer on Australian communities. The prominent display used by the Australian Cancer Atlas is a visualisation of Incidence Rates or Excess Death Rates. The set of geographic units used is Australian Statisical Areas, at Level 2. There are almost 2,200 individual areas, ranging in size from

The choropleth map display used in the Australian Cancer Atlas is familiar to the general public of Australia. It is appropriate to use this display as users can orient themselves on the map base and find geographic areas relevant to them. However, when the intention of the map user is to convey the whole spatial distribution the information derived visually from the colours can be misleading. The rural areas are over emphasised, and the densely populated inner city areas are not given enough attention.

Visual Inference

Visual inference will be used to determine if the communication of population geospatial statistics is more effective when using an alternative display. Buja et al [@GIIV] provide the 'lineup' protocol as a formal framework for testing visual statistical methods. Implementing this framework will allow new alternative visualisation method to be tested.

The lineup protocol will be used to test if a visualisation is effective, a visualisation displaying a real population based distribution can be hidden in a collection of visualisations that display null distributions [@chowd]. It takes inspiration from a police lineup. The witness in this regard is a participant recruited from a crowdsource platform, such as Figure-Eight. The visualisation containing a real distribution is considered the 'accused'. It is put in a lineup of innocent displays that do not show a real population based distribution. If the ‘witness’ chooses the ‘accussed’ as different from the innocent plots, it can be considered that there is a specific pattern displayed that is not present in the others. In this protocol, the null hypothesis can be rejected in favour of the alternative when it is chosen in the lineup. The null hypothesis fails to be rejected when it is not selected in the linup.

Aims and Objectives

This work aims to provide a solution to presenting geospatial data regarding populations. It considers the visualisation methods developed over the past two centuries that shift the focus from the geographic map base.

1. *Algorithm for creating hexagon tile maps of Australia:* The algorithm will take geospatial areas and create an alternative visualisation of the spatially distribution.
2. *Test the effectiveness of the hexagon tile map relative to the choropleth map:* The hexagon tile map produced by the algorithm will be contrasted with the traditional choropleth map, applying the same colour methods to represent the data. The maps will be used in an experiment to test the effectiveness by asking for users to spot spatial distributions.
3. *Communicating the relationship between the hexmap and choropleth map through animation:* Maximise the benefits of both displays when communicating to the public. The use of animations may control how people follow a recognisable map of Australia into an alternative visualisation for inference.

Research Contributions

This research contributes a new algorithm for creating hexagon tile map displays. It contributes an R [@R] package which implements the algorithm and allows R users to create their own visualisations. It presents a case study that contributes to a growing field of visual inference studies, applied to spatial data by comparing a choropleth map to a hexagon tile map display. It also shows how it can be used in practice to effectively communicate cancer distributions.

Thesis Structure

The thesis is structured as follows: Chapter two contains a literature review. The literature reviews considers the current peer reviewed literature and published books that explore spatial distributions of cancer across the globe. It also considers how to evaluate the visualisation methods used for spatial data.

Chapter three explores the algorithm to create hexagon tile maps and the code used to create a small example of Tasmania in Australia. Chapter four is a visual inference study that contains the methods and results that compare the use of a choropleth map and a hexagon tile map on the same data sets. Chapter five provides a conclusion of the results of the visual inference study and how the hexagon tile map may be used in practice.