A protocol to develop visual geographic models of socio-economic and public health data

Joshua Wells

School of Life Sciences, Pharmacy and Chemistry, Kingston University London; Kingston upon Thames

https://orcid.org/0000-0003-2021-6055

Robert Grant

Faculty of Health, Social Care & Education, Kingston and St George's, University of London; London

https://orcid.org/0000-0001-9563-9824

John Chang

Croydon University Hospital; Croydon; London

Reem Kayyali (✉ r.kayyali@kingston.ac.uk)

School of Life Sciences, Pharmacy and Chemistry, Kingston University London; Kingston upon Thames

https://orcid.org/0000-0002-7300-8738

Method Article

Abstract

Understanding the impact of socio-economic inequality on health outcomes is arguably more relevant than ever before given the global repercussions of COVID-19. With limited resources, innovative methods to track disease, population needs, and current health and social service provision are essential. One approach of interest is the implementation and integration of mapping software. This approach develops multi-layered interactive maps to demonstrate relationships between socio-economic and health data. Paediatric vaccination uptake and readmission rates are used as an example; however, the wider application of this method is not limited to these sets of health data. This approach aims to improve the interpretation and accessibility of data by creating a simple visual model with little need for additional technical experience beyond the guidance of the protocol.

Introduction

The Townsend Score[1] and Index of Multiple Deprivation[2] (IMD) are examples of data tools used to correlate the overall wellbeing and socioeconomic status (SES) of a population. Factors may include education, employment, health deprivation and disability. However, conveying this data in a meaningful and accessible manner that can also be interpreted by a layperson and scientist alike can be particularly challenging. Utilising visual geographic models to map, evaluate and communicate socio-economic and health data has proven to be particularly effective in identifying correlations, such as contrasting regional diabetes prevalence vs available community resources or tracking seasonal variation in serum vitamin D levels within a community.[3,4] This approach can highlight areas or populations that might be considered as "high risk-low resource" (HRLR), which in turn may support the development or optimisation of service allocation to tackle the specific needs of a population. Despite the increasing reliance on and access to smart technology, the uptake of visual data models is poor. Cited issues include poor usability, a lack of support to implement such models and inadequate knowledge of how to fully utilise visual tools to communicate and interpret data.[5] This protocol details a straightforward stepwise methodology to create and develop a visual geographic model of socio-economic and health data using 3 case studies with example data from Croydon University Hospital (CUH).

Reagents

Equipment

Procedure

1) Map Layer 1 (Base)

Select an area to map. This example selected the Borough of Croydon in London.

Access census data via GOV.UK[6] to collect the IMD values for each ward within the Borough of Croydon.

Access census data via GOV.UK[6] to download the available .shp files for the Borough's ward boundaries.

Download Mapshaper[7] software to convert the ward .shp file into a JSON (JavaScript Object Notation) file.

Download JSON editor to match the ward JSON files with their matching IMD values.

Open JSON editor and insert the relevant code e.g. "IMD":22.003" to the matching ward's file to correlate the

ward boundary data with the IMD value for that ward.

Download the Mapbox[8] software

Upload the edited JSON ward files to Mapbox[8] using the online tutorial to produce an outline of the ward boundary, which will contain the JSON edited IMD value for that ward.

Ward IMD values act as a defining variable to identify differences between wards in the Borough of Croydon.

Group wards based on their IMD values as follows <10, >10/<20, >20/<30, >30/<40, >40.

Assign a colour code to identify variations in ward value by IMD based on the groups above e.g. green and red corresponding to the least and most deprived areas respectively (Figure 1).

2) Map Layer 2 (Case Study 1)

Select a comparator data set to match which contains geographical data e.g. postcode. This example uses the addresses of children aged 5 who have received an Mumps, Measles & Rubella (MMR) vaccination. This data was provided by CUH as part of an approved research study with Kingston University.

If using postcode data, convert to longitudinal and latitudinal values using UK Grid Reference Finder[9]

Note at this stage, the data are non-anonymised in terms of addresses for study participants. If required to anonymise longitudinal/latitudinal data, upload the data to an excel file format.

Truncate the long/lat values using the =TRUNC function e.g.

=TRUNC (12.12345,3) = 12.123

Save the long/lat value as a CSV (Comma-Separated Values) file.

Upload the CSV file to Mapbox[8] to create an overlay to the base map layer that demonstrates vaccination uptake vs ward IMD (Figure 2).

3) Map Layer 3 (Case Study 2)

Data were provided by CUH identifying addresses of the patients with the highest readmission rates within 30 days of discharge for the following conditions: congestive heart failure, rheumatoid arthritis and falls. These data were compared with the distribution of civil service organisations in the Borough of Croydon.

If using postcode data, convert to longitudinal and latitudinal values using UK Grid Reference Finder.[9]

Note at this stage, the data are non-anonymised in terms of addresses for study participants. If required to anonymise longitudinal/latitudinal data, upload the data to an excel file format.

Truncate the long/lat values using the =TRUNC function e.g.

=TRUNC (12.12345,3) = 12.123

Save the long/lat value as a CSV (Comma-Separated Values) file.

Upload the CSV file to Mapbox[8] to create an overlay to the base map layer.

Download the postcodes of registered civil service organisation in the Borough of Croydon.

Convert to longitudinal and latitudinal values using UK Grid Reference Finder.[9]

Truncation is not required for this data set as the postcodes are not confidential and available in a public domain.

Save the long/lat value as a CSV (Comma-Separated Values) file.

Upload the CSV file to Mapbox[8] to create an overlay to the base map layer and health data for readmissions vs distribution of civil service organisation (Figure 3).

4) Map Layer 4 (Case Study 3)

To optimise the visualisation of paediatric MMR vaccination uptake by ward within the Borough of Croydon a simple calculation was performed using data provided by CUH.

(Number of children aged 5 who received an MMR vaccination in ward X) - (Total number of children aged 5 eligible to receive an MMR vaccination in ward X) = Number of unvaccinated children in ward X

Using the above calculation, a numerical figure for unvaccinated children can be assigned for each ward as a means to visually identify variance across the Borough of Croydon.

Open JSON editor and insert the relevant code e.g. "vaccination":50" to the matching ward's file to correlate the ward boundary data with the vaccination value for that ward, which will replace the previously edited IMD value.

Group wards based on their vaccination values to create distinct categories.

Assign a colour code to identify variations in ward by vaccination value e.g green and red corresponding to the most and least vaccinated areas respectively.

Download the postcodes of registered GP surgeries in the Borough of Croydon

Convert to longitudinal and latitudinal values using UK Grid Reference Finder.[9]

Truncation is not required for this data set as the postcodes are not confidential and available in a public domain.

Save the long/lat value as a CSV (Comma-Separated Values) file.

Upload the CSV file to Mapbox[8] to create an overlay to the base map layer (vaccination values by ward) vs distribution of GP surgeries in the Borough of Croydon (Figure 4)

Troubleshooting

Each step of this protocol outlines the manipulation of a specific data set relevant to the examples given in the case studies. When working with a unique data set, the authors suggest referring to the guidelines on the Mapbox[8] software which provides various guides to developing a mapping model. This protocol should be tackled in a stepwise approach, with each data layer developed individually starting with the base layer. Therefore, if errors occur, the developer can refer back to the previous step to ensure the correct methodology has been followed as well as identify the subject for troubleshooting using any search engine e.g., if using confidential long/lat data and the =TRUNC function is not working, the users can refer to the guide to identify that the following term should be used for troubleshooting "How to use the truncate function on excel?". Although this troubleshooting guidance may seem rudimentary, the protocol was used by several researchers in the original study with little to no experience of mapping software or coding. By following this simple troubleshooting method, easily overlooked errors were quickly identified.

Time Taken

The protocol and time taken will be completely reliant upon the type of data sets used to develop the visual models. However, the authors note that the time-limiting step for this process was becoming familiar with the Mapbox[8] software and would advise any protocol user to familiarise themselves with the software guides online before attempting to create a visual model, especially if the user lacks experience with this type of data visualisation method.

Anticipated Results

References

1. Townsend P, Phillimore P, Beattie A. Health and deprivation: inequality and the North: Routledge; (1988).

2. Jordan H, Roderick P, Martin D. The Index of Multiple Deprivation 2000 and accessibility effects on health. Journal of Epidemiology & Community Health. Mar 1;58(3):250-7. (2004)

3. Curtis AB, Kothari C, Paul R, Connors E. Using GIS and secondary data to target diabetes-related public health efforts. Public Health Rep May-Jun;128(3):212-220. (2013) https://doi.org/10.1177/003335491312800311

4. Laird E, Shannon T, Crowley VE, Healy M. The benefits of utilising geo-mapping for visualising the vitamin D status of Dublin city and the surrounding urban districts. Irish Journal of Medical Science. Nov 1;186(4):807-13. (2017)

5. Carroll LN, Au AP, Detwiler LT, Fu TC, Painter IS, Abernethy NF. Visualization and analytics tools for infectious disease epidemiology: a systematic review. J Biomed Inform Oct;51:287-298. (2014) https://doi.org/10.1016/j.jbi.2014.04.006

6. Department for Communities and Local Government. English Indices of Deprivation 2015 (https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015). September (GOV.UK). (2015)

7. MapShaper. Available at https://mapshaper.org/.

8. Mapbox, 2019. Available at: https://www.mapbox.com/

9. UK Grid Reference Finder. Postcode Batch Converter. 2014; Available at: https://gridreferencefinder.com/postcodeBatchConverter/

Acknowledgements

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Figure1.docx
- Figure2.docx
- Figure3.docx
- Figure4.docx