Meeting Client Needs while Maintaining Spatial Integrity:   
A Spatial Project on Digitalization in Farming

By

Mia Thuge Kuntz

202006657

hand-in due on

7/6 – 2023

*Abstract*

The paper explores a project conducted by researchers at Aarhus University that focused on digitalization in farming and the adoption of smarter agricultural practices. The aim of this paper is to highlight the workflow of this project, and offer insight into some of the decisions and compromises faced by a spatial analyst throughout the process. The research involved the collection and visualisation of spatial data, resulting in two types of maps – choropleth and centroid. These maps visualise the count of articles related to the research in smarter farming for each country. The paper discusses the importance of collaboration between the client and analyst, emphasizing the importance of addressing client requirements and incorporating these into the maps while also adhering to spatial standards, such as a title and colour legend, and keeping potential biases in mind. The paper and its results show the significance of communication and collaboration in order to achieve a successful outcome, and conclude by stressing the importance of maintaining transparency and generating user-friendly visualisations with respect to its audience.

*Keywords*

Centroid map, choropleth map, client needs, smarter farming, spatial biases, spatial integrity, visualisations

*GitHub*

The data and script for my project can be found in the following GitHub repository:

<https://github.com/MiaKuntz/spatial_final_proj.git>

HUSKELISTE

* OPDATER METHODS
* 4-6 SIDER

Table of Contents

[1. Introduction 2](#_Toc136781370)

[2. Smarter Farming Project 2](#_Toc136781371)

[3. Methods and Sources 3](#_Toc136781372)

[3.1 Hardware and Prerequisites 3](#_Toc136781373)

[3.2 Data 4](#_Toc136781374)

[4. Results and Critical Evaluation 5](#_Toc136781375)

[4.1 Choropleth maps 5](#_Toc136781376)

[4.2 Centroid map 7](#_Toc136781377)

[5. Conclusions 9](#_Toc136781378)

[Acknowledgements 9](#_Toc136781379)

[References 9](#_Toc136781380)

[Required Metadata 12](#_Toc136781381)

[Table 1 – Software metadata 12](#_Toc136781382)

[Table 2 – Data metadata 13](#_Toc136781383)

# 1. Introduction

When working in any type of specialised field, catering to client needs is always of the utmost importance. However, meeting these requirements can sometimes pose a challenge for specialists, as this may compromise their expertise and understanding of standard practices.

When given the opportunity to use my spatial skills on a research project, I saw a chance to utilize this paper to showcase the progress of generating and visualising spatial data while collaborating with non-spatial analysts on a spatial project. The paper explores the challenges and results of this assignment, while also presenting the workflow and essential elements intrinsic to any data science project.

# 2. Smarter Farming Project

In recent years, a significant increase in studies and research focused on smart farming technologies has occurred. With the introduction of the fourth industrial revolution and Agriculture 4.0 in the 2010s, a revolution in the technologies available to farmers has taken place, with growing opportunities in efficiency and sustainability (*Y. Liu, X. Ma, L. Shu, G. P. Hancke and A. M. Abu-Mahfouz. 2021*). Achieving this is made possible by adopting both old and new emerging technologies such as localisation technology, connectivity and sensors. In order to do so farmers as well as other interested parties need to understand what factors influence why some farmers adopt these technologies and why some don’t (*Commission.europa.eu - Industry 4.0 in agriculture: Focus on IoT aspects. 2017*).

A study conducted by researchers at Aarhus University in the spring of 2023 sought to recognize these factors, both by looking at previous research done in the field as well as using the obtained knowledge to prepare for their study to be done in the fall of 2023. [[1]](#footnote-1) The client responsible for this project required a spatial analyst to collect and visualise where the data required originated from, and I was approached due to earlier collaboration with the client, who was aware of my acquired spatial skills. The client required a number of articles concerning research in smart farming technologies and adoption to be collected, as well as an interactive map to be generated, which should depict where previous research on the subject has been done while being easy for people with no previous knowledge of the subject or spatial theory to understand. [[2]](#footnote-2) That implied that the map I had to generate didn’t necessarily have to adhere to the best spatial practices, as long as it lived up to these specifications. And yet, I set out to create a map that would adhere both to the client and their vision, as well as implement elements that are key to any map.

# 3. Methods and Sources

## 3.1 Hardware and Prerequisites

The code written for this project was done locally on a five-year-old 13-inch MacBook Pro (2017), which has 8 GB RAM and the macOS Ventura (version 13.3.1 a) operating system. It was developed and run in the desktop version of R (4.2.2) and RStudio (2023.03.1 Build 446).

The code was done in an R Markdown file, as the opportunity for knitting to an HTML file was needed for client needs, as it facilitates visualisations of the results. To run the code, please clone the repository to your device and ensure that you have both the desktop version of R and RStudio installed. You should also ensure that all packages used in the code are installed on your device. For more information on necessary packages please see the “ReadMe.md” file in the repository.

## 3.2 Data

The repository includes a “data” folder that contains “articles.xlsx” and “countries.geojson” files. The “articles.xlsx” file is included in the “data” folder for offline work should that be needed. The code will pull the data from an online Google Spreadsheet, and a link will be provided here as well as in the code:  
<https://docs.google.com/spreadsheets/d/1NFlbYvgNJCsr0uW5uyhvtZPhoj4UlpB0D9DB7rlhu6g/edit?usp=sharing>

The collection of articles was a collaborative effort with researchers at Aarhus University and could be updated even after the hand-in of this project. Please be aware that this may affect the outcome of the code should it be run after the hand-in date. The criteria for which articles should be included in the file were made in collaboration with the client. Different search words were defined, with the primary being “digital”, “farm\*”, “study”, and “survey”, and the search was primarily done on literary databases such as scopus.com, library.au.dk and sciencedirect.com. Since not all articles found via these portals and with these search words met the client’s requirements, a manual review through all considered articles for the file was done before those that met the criteria were added to the file.

After adding the articles to the file, their information was divided into 9 different columns based once again on client needs:

|  |  |
| --- | --- |
| **Column name** | **Description** |
| Author | The author(s) of the article |
| Year | The year the article was published |
| Title | The title of the article |
| Journal | The literary journal that published the article |
| Abstract | A short abstract of the article |
| Keywords | A number of keywords to describe the articles |
| Country | The country of origin of the article |
| Questionnaire/Interview | A binary variable for whether or not the article has a questionnaire or interview included in its study |
| Citations | The number of times the articles have been cited |
| URL | The URL of the article |

The “countries.geojson” file was pulled from an online GitHub repository (*Github.com – geo-countries*. 2015). The GeoJSON file contains polygons for all the world’s countries, and is essential for visualising the “articles” data, as the client’s request was to show the country of origin of all articles in an appropriate manner. Polygons are a type of vector data that connects several lines to shape the area of the data (*Manuel Gimond. 2023*). That the “countries” file is a Multipolygon means that it contain several polygons which contains the area of several entities – In this case the countries of the world. The data uses the WGS84 geodetic CRS, which is quite common when working with maps on a global scale, exactly as this project does (*Manuel Gimond. 2023*).

The file contains three columns:

|  |  |
| --- | --- |
| **Column name** | **Description** |
| ADMIN | The common name of the country |
| ISO\_A3 | A three-letter ISO code for the country |
| geometry | Multipolygon geometry of the country |

The main issue with pulling a GeoJSON file from another source was that I had to make sure that there was equivalence between the country names of this file and the “articles” file. Luckily, no issues in regards to country names not coinciding appeared when developing and running the code.

# 4. Results and Critical Evaluation

## 4.1 Choropleth maps

Based on the map described by the client’s vision, I wished to generate a choropleth map (*r-graph-gallery.com – Interactive choropleth map with R and leaflet. 2018*). The initial map approved by the client lacked several features, which I decided to include in the second and final map approved by the client. This map, including its design and features, was designed in collaboration with the client so that the additions wouldn’t go against any requirements. The client’s primary requirement was that the map should be interactive and capable of displaying the count, that is the number of this that country appeared in the “country” column in the “articles” data, for all countries. A snippet of the initial map first described by the client is featured below:  
A map of the world

Description automatically generated

Figure 1 – Choropleth Map 1

This map shows the world and not much else except different colours filling the area of several countries. Since the map is interactive, the count for each country will appear as the mouse hovers over the different countries, and while it’s also possible to zoom in and out, there is nothing more to the map than these select features. Without an explanation for which colours correspond to which values, it could be interpreted as if large-area countries carry more importance than small-area countries, even though some small area countries have a higher count than some large area countries. This is known as map area bias, and cannot be avoided when working with choropleth maps (*Jack Dougherty & Ilya Ilyankou. 2023*). It is therefore important to be aware of this when interpreting the map by just looking at it without using the interactive features. It is therefore clear that this map needs improvement.

Since both maps were generated using the `leaflet` package, utilizing cartographic practices when possible was key when generating the second map. Firstly, I found it important to include a legend to show which colour corresponds to which value since the client preferred colours that weren’t significantly different from each other for aesthetic purposes. I presented the client with a few options and we agreed that while the chosen palette fit the requirement, I was allowed to exclude the first four colours since they were almost transparent and would almost erase the countries with those values. In addition to the count bar, the map now includes a title and a count for the top five countries, while retaining the interactive feature of displaying the name and count of the country when the mouse hovers over it. A snippet of the second and final map is featured below:  
A map of the world

Description automatically generated

Figure 2 – Choropleth Map 2

On this map, the world and several other features are shown, and it has now gotten much easier to interpret the message with this map. From both the title, count bar and top five countries box we get several impressions of what type of information this map can provide, as well as the count for each country as the mouse hovers over them.

## 4.2 Centroid map

The other type of map generated exclusively for this paper was a centroid map – Also known as a bubble map (*r-graph-gallery.com – Bubble map with ggplot2. 2018*). Using the `ggplot2` package, I could create a map showing the centroids for each country, where the placement of the centroids is based on the country’s centre, and the size of the centroid is based on the count for that specific colour. This map differs from the choropleth maps in several ways and was generated to display how various maps can display the same value in different ways. A snippet of the map is featured below:  
A map of the world with purple dots

Description automatically generated with low confidence

Figure 3 – Centroid Map

Firstly, the map isn’t interactive, and it is therefore not possible to hover the mouse over a country to see its name and count. Secondly, it doesn’t include a top five countries count, as this map was only generated for the use of this paper and not for the client. Lastly, the colour palette of this map differs from the choropleth maps, due to the centroids of especially Europe overlapping which created a need for more contrasting colours to distinguish between the different centroids.

It’s important to note that while the centroid map provides a more “accurate” way of showing which countries have the highest count, both types of maps fall under the Mercator projection, and therefore display countries as distorted compared to their true size (*Fletcher, S. J. 2023*). As a consequence, both types of maps suffer from projection bias, where countries closer to the Equator are diminished in size, while countries further away are inflated (*Jack Dougherty & Ilya Ilyankou. 2023*). This enhances the area bias the choropleth maps suffer.

By including both types of maps to present the same data, this paper showcases the importance in choosing the right visualisation to interpret the data, while also being aware of how different biases can or can’t be avoided when utilizing different types of maps. Overall, the process of developing these maps in collaboration with the client resulted in a more informative and user-friendly visualisation of the data.

# 5. Conclusions

This paper aimed to use a project concerning the growing development in smart farming technologies conducted by researchers at Aarhus University as its frame for developing visualisations of spatial data in R while offering insight into the discussions, decisions and compromises faced by a spatial analyst.

In close collaboration with the project’s client, the generated maps sought to not only visualise the data in one way or another but also to provide meaningful information to aid in interpretation for individuals both familiar and unfamiliar with the subject. Both types of maps visualises the count of different countries appearing in the collection of articles on smart farming technologies, one in the way of a choropleth map, where interactive features and colouring of country areas showed the count, another in the way of a centroid map, where the size of the centroid was an indicator for the count.

Both types of maps presented their challenges and biases, which once again highlights the importance of being in communication with the client when developing, as it is ultimately their call which map will be used, although expert advice often is appreciated. It’s important to acknowledge as a spatial analyst, that staying true to the field and maintaining integrity sometimes has to be put aside in order to cater to client needs, and not all clients, unlike the client of this project, are willing to implement the suggestions put forward by the spatial analyst. In this case, the process and collaboration between analyst and client have resulted in all parties being satisfied with the result.

# Acknowledgements

Thank you to the people at Aarhus University working on the smart farming project for allowing me to use their research as a framework for this paper, and for taking the time to collaborate on and discuss the progress and direction of this paper. Since the project is still in such an early state, all work done in this paper has been welcomed with gratitude by the researchers to aid in their further research.

# References

*Commission.europa.eu - Industry 4.0 in agriculture: Focus on IoT aspects*. Available:

<https://ati.ec.europa.eu/sites/default/files/2020-07/Industry%204.0%20in%20Agriculture%20-%20Focus%20on%20IoT%20aspects%20%28v1%29.pdf>

[2017, 01.06.2023]

Fletcher, S. J. (2023). Data assimilation for the geosciences – From theory to application: Chapter 12: Numerical Modeling on the Sphere. Second edition. *Elsevier*. [online].

*Github.com – geo-countries*. Available:

<https://github.com/datasets/geo-countries.git>

[2015, 23.05.2023]

Jack Dougherty & Ilya Ilyankou. (2023). Hands-On Data Visualization: Interactive Storytelling from Spreadsheets to Code – Chapter 14: Recognize and Reduce Spatial Bias. [online]. Available:

<https://handsondataviz.org/spatial-bias.html>

[2023, 02.06.2023]

Manuel Gimond. (2023). Intro to GIS and Spatial Analysis – Chapter 2: Feature Representation. [online]. Available:

<https://mgimond.github.io/Spatial/chp02_0.html>

[2023, 26.05.2023]

Manuel Gimond. (2023). Intro to GIS and Spatial Analysis – Chapter 9: Coordinate Systems. [online]. Available:

<https://mgimond.github.io/Spatial/chp02_0.html>

[2023, 26.05.2023]

*r-graph-gallery.com – Interactive choropleth map with R and leaflet*. Available:

<https://github.com/datasets/geo-countries.git>

[2018, 20.05.2023]

*r-graph-gallery.com – Bubble map with ggplot2*. Available:

<https://github.com/datasets/geo-countries.git>

[2018, 27.05.2023]

Y. Liu, X. Ma, L. Shu, G. P. Hancke and A. M. Abu-Mahfouz. (2021).From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges. *IEEE Transactions on Industrial Informatics*. [online] vol. 17, no. 6, pp. 4322-4334.

# Required Metadata

Please fill in the right column with the correct information about your digital resources, and leave the left columns as they are

## Table 1 – Software metadata

|  |  |  |
| --- | --- | --- |
| **Nr** | **Software metadata description** |  |
| S1 | Current software version | R (4.2.2)  RStudio (2023.03.1 Build 446) |
| S2 | Permanent link to your code in your Github repository | *https://github.com/MiaKuntz/spatial\_final\_proj.git* |
| S3 | Legal Software License | *MIT License (see: https://github.com/MiaKuntz/spatial\_final\_proj/blob/main/LICENSE* |
| S4 | Computing platform / Operating System | macOS Ventura (version 13.3.1 a) |
| S5 | Installation requirements & dependencies for software not used in class | *googlesheets4*  *leaflet*  *sf*  *RColorBrewer*  *dplyr*  *htmltools*  *tidyverse*  *ggplot2*  *ggnewscale* |
| S6 | If available Link to software documentation for special software |  |
| S6 | Support email for questions | [202006657@post.au.dk](mailto:202006657@post.au.dk) |

## Table 2 – Data metadata

|  |  |  |
| --- | --- | --- |
| **Nr** | **Metadata description** | ***License*** |
| D1 | articles | *MIT License (see: https://github.com/MiaKuntz/spatial\_final\_proj/blob/main/LICENSE)* |
| D2 | geo-countries | *Open Data Commons Public Domain Dedication and License* |

1. As the study is ongoing a proper source cannot be provided. Please note, that all references to the study are on the parts where I am responsible. [↑](#footnote-ref-1)
2. For a more detailed description of the data collected, please see section 3.2 [↑](#footnote-ref-2)