Aalto University
School of Engineering

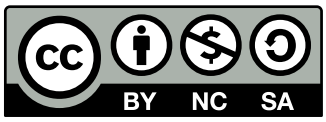
Master's Programme in Geoinformatics

Impact of web map survey design on survey quality

A quantitative exploration of public participation geographic information system survey design's impact on survey submission rate and spatial data quality

Tuukka Korpela

Master's Thesis
2025



Aalto University
School of Engineering

Author Tuukka Korpela

Title Impact of web map survey design on survey quality — A quantitative exploration of public participation geographic information system survey design's impact on survey submission rate and spatial data quality

Degree programme Geoinformatics

Supervisor D.Sc. Jussi Nikander

Advisor M.Sc. Sakari Ellonen

Collaborative partner Mapita Oy

Date 17.11.2025 **Number of pages** 69+16 **Language** English

Abstract

This thesis investigates how the design of web-based PPGIS surveys influences both survey submission rates and spatial data quality. Using a dataset of 216 surveys from Maptionnaire, a platform for participatory mapping, the study quantitatively analyzes relationships between survey design features and measures of survey quality through correlation and regression methods. The research identifies design factors, such as the length of a survey, background maps, and visual complexity, that affect survey completion and data precision. The results suggest that long surveys with relatively many popup windows tend to decrease submission rates, while spatial data quality is lowered by the use of multiple different background maps and visually complex and busy survey pages. Even though statistically significant, the results should be understood as indicative. Beyond survey design, there are multiple other factors, which were left out of this study, that can have a more significant effect on survey quality. The findings contribute empirical evidence to an underexplored area in PPGIS research and provide scientific suggestions for practical survey design guidelines.

Keywords PPGIS, spatial data quality, survey design, exploratory analysis, quantitative analysis, multiple regression

Aalto University
School of Engineering

Tekijä Tuukka Korpela

Työn nimi PPGIS-kyselyn muotoilun vaikutus kyselyn laatuun — Kvantitatiivinen tarkastelu verkkopohjaisen karttakyselyn muotoilun vaikutuksesta vastausprosenttiin ja paikkatiedon laatuun

Koulutusohjelma Geoinformatiikan maisteriohjelma

Työn valvoja TkT Jussi Nikander

Työn ohjaaja DI Sakari Ellonen

Yhteistyötaho Mapita Oy

Päivämäärä 17.11.2025 **Sivumäärä** 69+16 **Kieli** englanti

Tiivistelmä

Tämä diplomityö tarkastelee, miten verkkopohjaisten PPGIS-kyselyiden muotoilu vaikuttaa sekä kyselyiden palautusasteeseen että paikkatiedon laatuun. Tutkimuksessa hyödynnettiin 216 Maptionnaire-alustalla toteutetun karttapohjaisen kyselyn aineistoa. Kyselyiden muotoiluominaisuuksien ja laatumittareiden välisiä yhteyksiä tutkittiin kvantitatiivisesti korrelaatio- sekä regressioanalyysien avulla. Tutkimus tunnistaa kyselyn pituuden, taustakarttojen käytön ja visuaalisen monimutkaisuuden kaltaisia muotoiluominaisuuksia, jotka vaikuttavat palautusasteeseen ja paikkatiedon laatuun. Tulokset viittavat, että pitkät kyselyt, joissa on suhteellisesti paljon ponnahdusikkunoita, laskevat palautusastetta. Paikkatiedon laatua heikentävät useiden erilaisten taustakarttojen käyttö sekä sisällöllisesti monimutkaiset ja raskaat kyselysivut. Vaikka tulokset ovat tilastollisesti merkitseviä, niitä tulisi pitää suuntaa-antavina. Kyselyn muotoilun lisäksi monet muut, tämän tutkimuksen ulkopuolelle jääneet tekijät voivat vaikuttaa merkittävästi kyselyn laatuun. Tutkimus tuottaa empiiristä tietoa vähän tutkitusta aiheesta PPGIS-kentällä ja tarjoaa tieteellisesti perusteltuja käytännön suosituksia kyselyjen muotoiluun.

Avainsanat PPGIS, paikkatiedon laatu, kyselyn muotoilu, eksploratiivinen data-analyysi, kvantitatiivinen data-analyysi, monen selittäjän regressio

Preface

I would like to thank my supervisor Jussi Nikander and my advisor Sakari Ellonen for their insightful guidance. Thank you to the other people of Mapita Oy, as well, by letting me be part of their team and giving me a chance to work there.

I also want to thank my dad and my friend Wouter. Dad for the conversations that helped me better understand my own, often disorderly thoughts, and Wouter for giving me valuable tips all the while being busy conquering the city of dreams. Lastly, I want to thank Luisa for being there for me and reminding me of what is and isn't important.

Espoo, 17 November 2025

Tuukka Korpela

Contents

Symbols and abbreviations

Symbols

r	Pearson's correlation coefficient
ρ	Spearman's correlation coefficient
β_0	y-intercept (the constant term) in the regression model
β_p	slope coefficient for the p-th predictor variable
ε_i	residual (error) for observation i
R^2	coefficient of determination
z	z-score, i.e. standard score
μ	feature mean
σ	feature standard deviation
e	Euler's number

Operators

$\sum_i^n$	sum over index i with an upper limit n

Abbreviations

AI	Artificial Intelligence
AIC	Akaike Information Criterion
API	Application Programming Interface
BIC	Bayesian Information Criterion
CI/CD	Continuous Integration and Continuous Delivery
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
Df	Degrees of Freedom
FGDC	(U.S.) Federal Geographic Data Committee
GIS	Geographic Information Systems
HC3	Heteroskedasticity-Consistent estimator
ISO	International Organization for Standardization
LLM	Large Language Model
OLS	Ordinary Least Squares (Regression)
OSM	OpenStreetMap
Oy	Osakeyhtiö (Limited company)
PGIS	Participatory Geographic Information Systems
PM	Participatory Mapping
PPGIS	Public Participation Geographic Information Systems
Q-Q	Quantile-Quantile
SQL	Structured Query Language
UTM	Urchin Tracking Module
VGI	Volunteered Geographic Information
VIF	Variance Inflation Factor
WMS	Web Map Service

1 Introduction

Since the turn of the century, the spatial planning practice has shifted from top-down approaches towards participatory planning and integrated methods, which have increased the legitimacy, transparency, and quality of planning outcomes [1]. Participatory planning emphasizes the importance of local stakeholder involvement in decision-making and planning action concerning the locals' own living environments [2]. The UN's Sustainable Development Goals highlight the current focus on participation, advocating for "... inclusive and sustainable urbanization and capacity for participatory, integrated and sustainable human settlement planning and management in all countries," [3].

Congruent to the shift in spatial planning practice, the field of geoinformation has shifted its frame, as well. Geographic information systems (GIS) – information systems and technologies dealing with spatial entities and relationships with capabilities for spatial analytics and modeling [4] – have evolved from processes and computer systems controlled and utilized by an elite few onto methods and technologies that are more accessible and used by the people whose lives these systems help to shape [5]. An example of this development is the emergence of public participation geographic information systems (PPGIS), coined as a term in 1996 "to describe how GIS technology could support public participation for a variety of applications with the goal of inclusion and empowerment of marginalized populations," [6, p. 7].

PPGIS is a method where local stakeholders are asked to mark down and contextualize their location-based values, thoughts, behaviors, or preferences on a digital map for the purposes of i.a. policy-making and land development [7]. While non-digital public participation, such as town-hall meetings or hearings, are still a prominent tool in the toolbox of planners and consultants, web-based PPGIS offer a way to visualize data directly on a map and reach parts of the population who might otherwise be excluded from participating in decision-making [8]. In addition, online participation makes it easier to collect, store, aggregate, and analyze data when compared to traditional data collection methods [9].

According to [7], a PPGIS project can be split up into multiple distinct phases: survey/tool design, respondent recruitment, data collection, and data analysis. Each phase has implications for the analysis and data quality, and has been studied in its own right. Survey design, which comprises the PPGIS survey's aesthetic and technical properties, influences the respondent's user-experience and thus the collected data's quality [10]. Similarly, recruitment and data collection methods have been proven to impact data representativeness and quality [11].

Concerns about data quality have been raised due to the different phases within a PPGIS project and the inherently non-expert, non-standardized nature of the data. Multi-phase projects create idiosyncratic methodologies which are difficult to situate into a systematic framework [7] that would allow clear quality assessments. Different PPGIS tools and their unique properties generate differing user-experiences impacting data quality in varying ways [12]. The representativeness and credibility of data producers and thus the accuracy and completeness of the produced data is often difficult to validate [11], [13], [14]. Additionally, a vexing problem for PPGIS in

terms of data quality is its goal to understand subjective values – there are usually no objective criteria with which subjective values can be evaluated [15].

Therefore, while PPGIS has shown great promise as part of the wider context of participatory planning [15], multi-phase projects, crowdsourced and subjective data, and the variety in available tools pose challenges to generate high quality data and outcomes [16].

1.1 Study aim, questions, and scope

The aim of this study is to quantitatively analyze the relationship between web-based mapping survey design and survey quality, using data from the Maptionnaire platform. Building on a recognized gap in existing literature, this research seeks to provide evidence-based guidelines for effective survey design in the field of PPGIS. The main research question guiding this study is:

What effect does the design of a Maptionnaire survey have on survey submission rate and spatial data quality?

The main question is supported by two sub-questions:

- Which Maptionnaire features constitute survey design?
- How can spatial data quality be operationalized for Maptionnaire surveys?

To answer these questions, this research will use correlation and multiple regression analyses to examine the relationship between specific survey design features and two key metrics operationalized as survey quality: the survey submission rate and spatial data quality. Submission rate is a straightforward metric – the number of submitted surveys divided by the total number of survey visitors – and has been used in previous research as a quality metric [17]–[19]. Submission rate is also an intuitive quality metric: a survey creator wants respondents to answer the whole survey, not just a part of it, and designs the survey accordingly. If respondents leave the survey before reaching the end, it may suggest poor survey design.

Spatial data quality, on the other hand, is a more complex metric and will be discussed in more detail, as there are no well-established metrics for it within the PPGIS field [15]. The operationalization of this metric will be elaborated on in subsequent chapters.

This research is delimited to a quantitative analysis of Maptionnaire surveys, a web-based platform for community engagement via cartographic surveys. This restriction is justified by the fact that web-based mapping surveys are one of the most common methods for collecting PPGIS data [20].

Specifically, the study's scope is further delimited to the survey design aspects of Maptionnaire surveys. While other factors – such as respondent recruitment, data representativeness, or the concrete use of project outcomes in decision-making [11], [21], [22] – are important for overall PPGIS project quality, they are beyond the scope of this study. This focus allows for a large-scale quantitative analysis using

Maptionnaire's database, addressing the lack of quantitative research on this topic. Previous PPGIS studies have often relied on highly specific methods applied to single surveys [7], leaving a gap in understanding the impact of design choices across a large number of surveys [22], [23].

Through collaboration with Mapita Oy, the company behind Maptionnaire, a practical objective of this research is to produce scientifically justified guidelines for survey creators. While Maptionnaire provides tutorials, there is a lack of scientific justification for design choices, and a notable scarcity of best practices for PPGIS surveys in the existing literature.

1.2 Thesis structure

After this introduction, a literature review is performed of previous research on the topic in chapter two. Within the literature review, the concepts of participatory planning, participatory mapping, spatial data quality, and survey design will be elaborated on. Following the literature review, an explanation of the used materials and methods is given in chapter three. The fourth chapter presents the results that were retrieved according to chapter three. Finally in chapter five, the results are discussed and reflected upon in a wider context. The limitations of this study and future research directions are also provided in the last chapter.

2 Literature review

This chapter reviews the theoretical foundations of this research. It is structured hierarchically, beginning with the overarching concepts that situate this study in a larger body of research and narrowing to the concepts that are more specific to this study. First, the concept of participatory planning is examined. Then participatory mapping, the direct methodological context for PPGIS, is looked at. Subsequently, the chapter delves into the concept of spatial data quality, a central theme of this thesis, exploring its theoretical underpinnings. Finally, some principles of survey design are examined, highlighting survey design's impact on data quality. Before going into these theoretical foundations, however, a clarification of how the literature review was conducted is given.

2.1 Method

The literature review was prefaced with an initial scoping review. A scoping review – usually only a preliminary step for a more detailed review – assesses the size and scope of the available research literature based on a topic of interest [24].

Google Scholar was used for the scoping review, accompanied by expert consultation. Google Scholar doesn't strictly filter for peer-reviewed articles and has been noted to lack transparency [25], [26], which led to the use of the peer-reviewed and more reliable ScienceDirect database for the literature review [27], [28]. The experts who were consulted are affiliated with Mapita Oy. They are experts in the field of PPGIS and have written multiple scientific articles about the topic.

The search string for the initial scoping study in Google Scholar was "PPGIS OR public participation OR public participation GIS OR Maptionnaire AND web-based survey AND data quality". The search resulted in approximately 40 300 results from the Google Scholar database. The first 50 hits sorted by relevance were scoped for their keywords and the keywords were stored (appendix A) in order to formulate the search string for the literature review.

A search string for the literature review in ScienceDirect was formulated from a subset of the keywords identified in the scoping review: "('public participation geographic information system' OR 'public participation GIS' OR 'PPGIS' OR 'Maptionnaire') AND ('criteria' OR 'spatial data quality') AND ('online survey' OR 'web-based survey')". This resulted in 77 hits in the ScienceDirect database. A selection of articles was used as the basis for the review, leaning on recommendations from the consulted experts. The selection was also extended through a backward snowballing method (see e.g. [29]) by including some suitable articles in the selected articles' bibliographies.

2.2 Participatory planning

Participatory planning is a collaborative process that focuses on the involvement of all relevant stakeholders, especially local communities and end-users, in decision-making for planning and socio-economic goals [2]. It moves beyond expert-driven models to

promote legitimacy, foster ownership, and create plans that reflect community needs, thereby leading to more effective and inclusive outcomes, especially for marginalized groups [1].

The theoretical foundations for participatory planning are drawn from multiple different fields of study. Brown and Kyttä [20] emphasize the advances in (environmental) psychology in the 1970s and 1980s, when theories of human culture and activity acknowledged the importance of individuals and communities in assigning value to their environments (e.g. [30], [31]). Supporting and exemplifying these theories, studies in anthropology asserted how "one must acknowledge that local understandings of external realities are ineluctably fashioned from local cultural materials, and that, knowing little or nothing of the latter, one's ability to make appropriate sense of 'what is' and 'what occurs' in one's environment is bound to be deficient," [32, p. 100]. Similarly, the field of cartography started to critically evaluate the understanding of maps as objective realities in the 90s and reasoned for maps as socially embedded constructions instead [33], [34].

Thus, the theoretical seeds for the importance of local involvement and citizen participation in planning were sown. However, as the seminal text on citizen participation by Arnstein [35] notes, participation in planning is not a one-dimensional concept that is always realized in a specific way, but rather a scale where the locals' power to practically impact the planning process can greatly differ. Informing locals about a finished plan can be marketed as participation, when realistically this kind of "tokenism" doesn't give any power for the locals to alter the plan. Consulting locals in the earliest phases of a planning process, on the other hand, would give them real opportunities to influence the end-result.

In addition to these epistemic discoveries, advancements in information and communication technologies (ICT) have had a significant impact on the implementation of participatory planning. Traditional methods for incorporating the local community in the planning process – such as public meetings, workshops, or hearings – have been proven to often result in unrepresentative and low rates of participation [36].

Digitalization makes it easier to collect, store, aggregate and analyze data [9], and thus, web-based methods, like online surveys, are lauded as a possible solution to reach a wider and more diverse group of people by allowing participation to take place regardless of time and place [8]. Digital methods also enable sophisticated data analysis and visualization through e.g. online maps [37] and are relatively easy to deploy in the earlier stages of a planning process [11].

2.3 Participatory mapping

Within the broader framework of participatory planning, participatory mapping (PM) has emerged as a set of tools for embedding local knowledge into spatial decision-making.

Participatory mapping is a process where a location and the properties related to that location are mapped and contextualized by the local community [38]. When it was realized that local people's contributions are essential for decision-making,

participatory mapping "spread like a pandemic" [39, p. 1] within the wider context of participatory planning.

Nowadays, participatory mapping multiple methods such as softGIS, participatory GIS (PGIS), volunteered geographic information (VGI), and public participation GIS (PPGIS) which all roughly describe the generation, collection, and sharing of spatial data [20]. Albeit related, these terms are often separated and used in their own contexts.

The term softGIS emphasizes the importance of experiential (or "soft") knowledge in understanding people's local environments. Technical expertise and pure statistical approaches ("hard" knowledge) alone are not satisfactory to describe an environment. Additionally, softGIS views participation through a holistic lens, where the tools and methods themselves mediate the experience of participation and have implications on the outcomes of GIS procedures [40].

PGIS is a term highlighting the practice that utilizes the epistemic ideas from softGIS [41]. PGIS was coined to denote the evolution of GIS technologies from systems exclusively aimed at experts to democratic systems also catering to laypeople (Abbot, 1998). The focus of PGIS is in community empowerment and fostering social identity, not on large data nor necessarily on data quality [20].

PPGIS utilizes modern information and communication technologies for applications such as urban development, zoning, or environmental conservation by digitally gathering and sharing local voices, concerns, and aspirations [42]. The difference between PGIS and PPGIS arises mainly from the applied methods: PPGIS leans heavily on technical, web-based methods whereas PGIS encompasses a set of analog and mixed mapping methods for community engagement [43]. In addition, PPGIS has historically been applied in the global north, whereas PGIS emerged in the global south [20].

VGI describes a phenomenon where volunteers create, collect, and share geographic information online [44]. OpenStreetMap (OSM) – "an initiative to create and provide free geographic data, such as street maps, to anyone" [45] – is one of the most well-known examples of VGI with around ten million volunteered contributors worldwide (see https://planet.openstreetmap.org/statistics/data_stats.html). Where PGIS concentrates mainly on participatory methods and outcomes, VGI focuses on application and big data [46].

When compared to the other three terms (softGIS, PGIS, and PPGIS), VGI is unique by focusing on information creation, dissemination, and volunteering rather than decision-making supported by participation [47]. However, VGI – and more precisely OSM – is introduced here because a novel proxy developed for this study utilizes OSM in measuring spatial data quality.

2.4 Spatial data quality

Even though primarily considered a positive development for project legitimacy and acceptance [48], the democratization of GIS systems has introduced issues for data validation and quality assessment:

> [The democratization of GIS] facilitates the use of data for non-intended purposes, as well as the overlaying of heterogeneous data collected at different periods, by different organizations, using various acquisition technologies, standards, and specifications. Such a context increases the risks of geospatial data misuse [16, p. 261].

This concern is not new; the rise of participatory mapping in the 1990s was accompanied by a surge in academic focus on spatial data accuracy, leading to a need for structured quality frameworks [49].

2.4.1 Data quality traditionally

Early frameworks for spatial data quality were developed for expert-driven, authoritative datasets where data could be validated against an objective reality [50]. These frameworks usually treat data quality as a measure of accuracy and consistency against established standards or thresholds. For instance, Fisher [51] proposed a classification of uncertainty, discerning between error (the discrepancy between a mapped value and a true reference value), vagueness (imprecise definitions of features or classes), and ambiguity (features having multiple possible definitions).

Building on such concepts, formal standards were developed. The International Organization for Standardization's (ISO) 19157-1 is a global standard for geographic data quality, outlining six core components for spatial data quality [52], [53]:

- **Completeness**: No necessary data are missing and no excess data is present.
- **Logical consistency**: Consistency of data with similar data.
- **Positional accuracy**: Mapped data is spatially close to its reference data.
- **Temporal quality**: Time-associated data reflects the actual timing of the phenomenon it represents.
- **Thematic quality**: Attributes of non-spatial data describe the data accurately.
- **Usability**: Data is fit for its purpose.

These ISO components, and similar national standards like that of the U.S. Federal Geographic Data Committee (FGDC) (https://www.fgdc.gov/csdgmgraphical/dataq.htm), provide a robust system for assessing the quality of objective spatial data, such as infrastructure maps or topographical surveys. However, their direct application to crowdsourced and participatory data can be problematic.

2.4.2 Data quality in PPGIS

The traditional, technical criteria for data quality are often insufficient for PPGIS data because this data is frequently subjective, experiential, and perceptual [15], [52]. Concepts like natural beauty, sense of safety, or cultural value do not have an objective

"ground truth" against which positional or thematic accuracy can be measured. A participant's perception is, by definition, true for them.

This challenge has led researchers to explore proxies – indirect indicators that suggest data quality in the absence of objective verification [54]. These proxies shift the focus from the data's accuracy to the contributor's behavior, assuming that certain actions correlate with more thoughtful and reliable contributions.

One early proxy was "mapping effort", proposed by Brown et al. [55] and operationalized through two metrics: response time and the number of mapped markers. The hypothesis is that a user who spends more time and maps more features is more engaged and thus provides higher-quality data. However, this proxy has been criticized for its sensitivity to the cognitive load of the survey task; a complex question will naturally take more time, independent of the user's diligence [17].

A more recent proxy is zooming behavior. Zooming behavior in this context is defined as the survey respondent's utilization of a map's zoom feature when answering a place-based question. Rzeszewski and Kotus [23] found a positive correlation between the map's zoom level and the positional accuracy of responses when users were asked to locate known landmarks. This suggests that zooming in may indicate greater care and precision for survey responses. Their findings also highlighted the issues within "mapping effort", as younger participants achieved similar accuracy in less time than older participants, showing that speed does not necessarily negate quality.

Despite their utility, these proxies are imperfect generalizations. It is easy to envision scenarios where they fail: mapping fuzzy emotional concepts may not require zooming, and a user can still be zoomed in on an incorrect location. As Johnson [54] cautions, proxies can be misleading and should be used with a clear understanding of their limitations.

In addition to proxies, Fagerholm et al. [7] have collated some more traditional measures to explore spatial data quality in PPGIS: point density or dispersion estimations can be used to examine the intensity/density of mapped features (e.g. standard distance), clustering (e.g. DBSCAN) of features can be evaluated to understand the behavior of different data categories, and proximity (e.g. buffer analysis and distance thresholds) of features can give an estimate of positional accuracy. However, these measures often fail to give a holistic view of spatial data quality, and are mostly focused on positional accuracy.

Finally, when examining spatial data quality in PPGIS, previous studies have usually examined only one specific project [15], [23], [55], [56]. Syntheses and metareviews about data quality have been conducted [15], [20], [57], but they are mostly qualitative in nature and focus on describing the methodological incoherence within the field, not on evaluating survey design with specified quality metrics. As far as the literature reviewed for this study can suggest, there's a gap in PPGIS research where multiple surveys created with the same tool would be examined against clearly defined criteria.

2.4.3 Survey design

Given the challenges in assessing PPGIS data quality retrospectively, the literature increasingly emphasizes the importance of proactive survey design to ensure quality at the point of collection.

PPGIS survey design extends beyond traditional survey features to include the spatial and technological aspects of mapping. Guidelines for designing effective surveys are well-established [58]–[60], but PPGIS surveys introduce unique challenges that directly impact data quality and user-experience [13], [23].

One key challenge in PPGIS is the wide array of design options available [20], [23]. The spatial features to be mapped vary (point, line, polygon, volume); map location and scale differ between surveys; and as a visual tool, a PPGIS survey is usually highly modifiable aesthetically (colors, images, map themes, etc.). As an example, Maptionnaire has 33 elements (e.g. likert scale, text input, white space, video, budgeting question, etc.) that a survey creator can utilize in designing a survey. What's more, nearly all elements have some aesthetic options in them.

Three survey design features, which are strictly speaking settings in Maptionnaire, were added as Maptionnaire "elements" for the purposes of this study. The three added elements represent the use of custom button colors, use of custom icons, and the use of an element that shows the respondent their progress in answering the survey. These added elements are visible in figures 3 and G3.

Design choices impact the time it takes to answer surveys. General surveys that are easy and relatively quick to answer usually produce the most accurate data [59]. A recent study by Salminen et al. [19] suggests that PPGIS surveys also benefit from relatively short surveys. Surveys that take over 15 minutes to answer have a significant drop-off in participation rates – people are not keen on taking more than 15 minutes for an online survey.

Furthermore, the design of a PPGIS survey should account for the cognitive burden placed on participants. Answering surveys is a cognitively demanding task (Brosnan et al., 2021), and the load is amplified in PPGIS due to mapping [23]. Cognitive load in PPGIS is compounded by several factors [12]:

- **Digital literacy**: A participant's comfort and skill with technology in general.
- **Familiarity**: A participant's personal knowledge of the mapped location.
- **Mapping device**: GPS device automates data collection whereas self-administered surveys require cognitive effort.
- **UX design**: The intuitiveness of the user-interface.
- **Background maps**: The clarity and scale of the visual context provided.
- **Question type**: Whether the question is a mapping, open-ended, or numeric query.
- **Mapped feature**: The geometric type of the feature being mapped.

- **Subjectivity**: The objective or subjective nature of the information being requested (e.g. mapping pedestrian road crossings versus mapping oral heritage).

These factors underscore the importance of a PPGIS survey design. Cognitive load, user-experience, and consequently data quality all connect to the way a survey is designed.

It should be emphasized that "design" in the context of this study does not solely refer to aesthetics, but is an overarching term for features that can impact the user-experience when answering a survey. For instance, the length of a survey is considered to be a design option (indicated by the median time it takes for respondents to finish a survey), as is the theme of the survey (e.g. mobility and transport).

3 Maptionnaire data and research methodology

This chapter outlines the data and methodology used to carry out the quantitative analysis examining how survey design influences Maptionnaire survey quality. First, the data is described in detail. After that, a three-step theoretical framework – explore-explain-predict – created by Fagerholm et al. [7] is introduced as the overarching methodological data analysis pipeline for this research. The applied methods and the reasoning for their selection are clarified after the introduction of the methodological framework.

3.1 Data description, source and selection

All analyses in this study are based on surveys hosted on the Maptionnaire platform. A Maptionnaire survey is similar to a slide show where a respondent answers questions or views information by moving through survey pages. Each page can have a different background. There are four types of backgrounds: (1) maps (system maps, own basemaps, map overlays), (2) background images, (3) interactive images, and (4) color.

In addition to a background, different survey elements can be added to each survey page from an element library, instituting the main content of a survey. The elements in the library are grouped into four categories and are presented in table 1. Most elements also have customization possibilities which allow modifying their visual look. For example, buttons which the respondent would click to initiate the placement of a point marker can be customized by changing their text (e.g. "Click here to place a marker!" or "Place a pin on your favourite spot!"), color, or icon.

Table 1: Maptionnarie element categories with examples.

Category	Description	Examples
Map questions	Elements that prompt an answer on a map.	Points, lines, areas, select elements (map overlays added on top of base maps that the respondent can select)
Questions	Generic question types for contextual information.	Open questions, multiple-choice questions, Likert scale
Content	Tools for visual design.	Images, headings, spacers (line breaks)
Programmatic elements	Invisible elements primarily for the survey creator's use.	Device information, UTM tracker

To ensure data consistency and quality, surveys were included in the analysis only if they met a series of predefined criteria applied before data querying. First, surveys had to have a **published status**, meaning they were available for the public.

Second, only surveys with a **publication date of 2023 or later** were included to avoid deprecated elements resulting from Maptionnaire's continuous integration and deployment (CI/CD) framework. Third, surveys were required to contain **map-based content** (at least one map-based question) to align with the study aim, which focuses on the relationship between survey design and spatial data quality. Fourth, a **minimum participation** of at least ten submissions was necessary to indicate a minimum level of survey validity, as fewer submissions could easily indicate the survey was published for testing purposes. Fifth, surveys had to be available in either **Finnish or English** to ensure accurate interpretation during the explore phase. Finally, the origin of creation was limited to those produced by Finnish or U.S. consultation organizations to maintain a manageable thematic scope, in line with the preferences of Mapita Oy.

With these filters, two datasets were compiled from the Maptionnaire database. The dataset used for the explore phase consisted of 37 surveys, each from a different consultation organization to get a broad understanding of different survey designs. Based on the results drawn from the explore phase, a dataset of 216 surveys from 50 unique organizations was used in the explain phase. The explore and explain phases are described in more detail below.

3.2 Methodological framework

Quantitative analysis uses data that is structured and can be represented numerically [61]. It allows statistical analysis of relationships between survey features and quality by aiming to build accurate and reliable numerical measurements [62]. The main tools used for conducting the quantitative analysis were PostgreSQL, Python, and Google Sheets. PostgreSQL was used to query and transform the data, Python was used to transform, analyse and visualise it, and Google Sheets was used as an easily modifiable database that enables descriptive statistical analysis. In addition, Google's large language models (LLM) Gemini 2.5 Flash and Pro were used to help with querying and scripting. More on the general use of AI in writing this thesis in appendix B.

To place the analysis within established frames, the research follows the data analysis pipeline proposed by [7]. This framework was developed to reduce methodological inconsistency within PPGIS research and offers a gradual process for moving from raw data to explanatory and predictive insights. It consists of three sequential phases:

1. **Explore**: Feature engineering to query and organize data, and to deal with outliers. Descriptive and univariate examination of PPGIS data, including visualization and detection of patterns or anomalies.

2. **Explain**: Application of more advanced analyses to reveal relationships and potential causal mechanisms within the data.

3. **Predict**: Extrapolation of results to new settings or unseen data. This phase was not implemented in this research.

This framework, even though presented as sequential, is often an iterative process in practice [7]. The phases are repeated and overlap with each other. This was also

true for this research and multiple iterations of feature engineering and analysis were carried out before arriving at the final results. The explore phase guided the first iteration of the explain phase, but the explain phase was not a one-off analysis where everything worked out on the first iteration. There were elements of exploration within the explain phase, as well.

3.3 Data analysis process

3.3.1 Explore phase

The explore phase aimed to operationalize two key concepts – **survey design** and **survey quality** – before performing statistical modelling. The process was carried out on a subset of 37 Maptionnaire surveys that met the filtering criteria described earlier. The process is visualized as a flowchart in figure 1.

The purpose of this phase was fourfold: (1) to understand the structure and functionality of Maptionnaire as a survey creation tool, (2) to identify measurable design features that may influence survey quality, (3) to define and test potential proxies for spatial data quality, and (4) to engineer a dataset for performing preliminary correlation and regression analyses.

Exploratory analysis was applied to find and optimize the features constituting **survey design**. Exploratory analysis aims to understand raw data through statistical techniques and graphical visualization [63]. Exploratory analysis helps to detect patterns and outliers within the data, and can be especially useful when there's a large dataset to be analyzed [63], [64]. Descriptive and univariate analyses are often part of exploratory analysis and provide information about the variables under investigation. They can be used to, for instance, calculate mean values or visualize data distributions [65].

The exploratory analysis started with a **visual inspection** of the selected surveys and the goal was to get a high-level understanding of Maptionnaire surveys. There was one practical reason for selecting 37 surveys for this phase: going through surveys manually and taking notes is time-consuming. Even if not comprehensive, this number of surveys gave a good idea about the possibilities of Maptionnaire as a survey creation platform. For each survey, notes were taken on used elements, background maps, design choices, and a thematic classification was made based on a previous classification made by Mapita Oy.

After the visual inspection, a **feature engineering** process was initiated. Feature engineering transforms raw data into relevant information for statistical analysis such as regression modeling [66]. In this research the process included querying the Maptionnaire database for potential survey design features, calculating aggregates (e.g. median time it took for survey participants to finish a survey), tabulating the data into a machine-readable dataset, one-hot encoding, and standardizing features into a uniform format (z-score). One-hot encoding transforms a categorical variable into binary variables by creating new columns for each category where 1 means the category is present and 0 means it is absent [67].

In combination with the feature engineering process, **descriptive and univariate**

analyses were conducted. As can be viewed in figure 1, the feature engineering and descriptive statistics followed an iterative process where the descriptive analysis provided insights for fit and unfit features. If a feature proved to be unfit, the feature was dropped or it went through another feature engineering process. An example of an unfit feature is a one-hot encoded feature with little variance (almost only 1s or 0s).

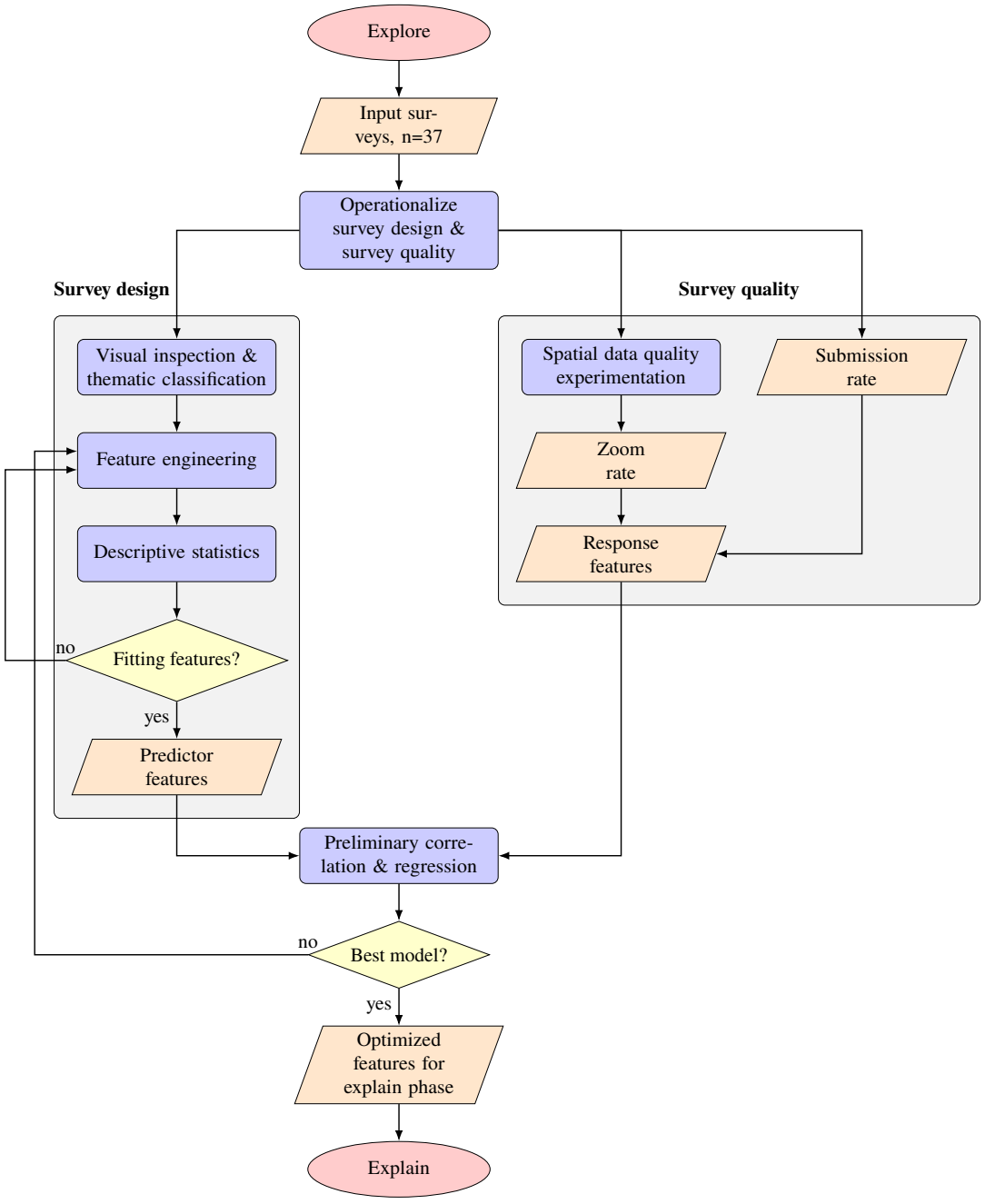


Figure 1: Flowchart of the explore phase leading to the explain phase. Red ellipses are the start and end node, orange rhomboids are data input/output, blue rectangles are processes, and yellow diamonds indicate decisions.

Spatial data quality was operationalized as **zooming behavior**, because zooming behavior has been used in previous research as a proxy [23], [68] and, thus, has some established support in the field of PPGIS. In addition to theoretical support, zooming behavior was chosen for its practicality. The Maptionnaire database stores both default and respondent zoom levels which makes zooming behavior readily measurable.

The zooming behavior (**zoom rate** from here onward) in this study is measured as the number of mapping responses that use a zoom level higher than the default zoom level(s) divided by all mapping responses. This measurement follows an intuitive logic: using a higher zoom level indicates a more accurate and thought-out feature placement when compared to using the default or lower than the default zoom level.

Some experimentation was conducted for spatial data quality before settling on zoom rate. The experimentation was prompted by the novel setting of this research: a single threshold value or an objectively correct dataset denoting positional accuracy or spatial data quality is not applicable for an analysis of multiple PPGIS surveys.

Three alternative proxies were tested with: (1) a proxy based on standard distance, (2) AI assisted definition of a distance threshold for each survey, and (3) an OSM-based proxy focused on thematic quality and inter-rater agreement. These proxies were found to be insufficient as proxies and were excluded from the analysis in the end. However, a short description of each proxy is given here. The reasoning for excluding these proxies is given as a result of the experimentation process in the next chapter.

The first proxy was based on **standard distance**, a measure of dispersion in a set of geographical features [69]. Standard distance measures the dispersion of features around their mean center and is often represented by drawing a circle with a radius equal to the standard distance around the mean center [70]. The proxy would have used standard distance as a binary threshold measure for spatial quality: the percentage of features within the standard distance would have been the spatial data quality metric for a survey.

The second proxy experimented with was based on **a positional accuracy threshold determined by AI**. For each survey, its text content was retrieved from the Maptionnaire database and fed to Google's LLM Gemini Pro 1.5. The LLM, acting as an expert in the field of PPGIS, was asked to come up with a reasonable positional accuracy threshold in meters and an explanation for it based on the mapping task described in the text content. The used prompt and an example of retrieved answers is added as appendix C. After the threshold value was determined by the LLM, similar binary filtering would've been applied as with the standard distance proxy to make up the proxy.

Lastly, an **OSM-based** proxy was developed based on the fact that Maptionnaire uses background maps from MapTiler and Mapbox that are built on top of OSM. The proxy focused on thematic quality and inter-rater agreement of survey responses. Inter-rater agreement is a measure of the consistency and agreement between two or more respondents in their, in this case, feature placement (Gisev et al., 2013). To define the proxy, the OSM database was queried with the survey responses via the Overpass API (`https://dev.overpass-api.de/overpass-doc/en/`) and the five most frequent OSM tag keys (OSM tags are key=value pairs) were computed for each survey. The proxy was then operationalized as the rate of survey responses that

agree with the top OSM tag keys when compared to all responses.

When a preliminary dataset was engineered by defining the survey design features as the predictor features and submission rate and zoom rate as the response features, **preliminary correlation** and **regression analysis** was implemented. Pearson's product moment correlation coefficient (equation 1) and the Spearman's rank correlation coefficient (equation 2) were calculated to measure the strength and direction of the relationship between each feature pair and were treated as a preparatory step for regression analysis. Multiple regression (equation 3) was used to understand the impact of the predictive survey design features on the survey quality features and the ordinary least squares (OLS) estimator was used to find the best fitting regression line.

The Pearson's correlation coefficient quantifies the linear relationship between two variables by dividing their covariance – how they vary together – by the product of their standard deviations, which measure how much they vary individually [71]. Pearson's correlation assumes linear relationships and is parametric by relying on normally distributed data [72].

The Pearson's correlation coefficient,

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

where

r is the correlation coefficient,
x_i, y_i are the values of the x and y variables for observation i,
$\bar{x}, \bar{y}$ are the means of the x and y variables, respectively,
n is the number of observations,

ranges from -1 to +1, where a negative or positive score close to one indicates a strong association between variables, while a score near zero indicates no correlation. A coefficient, r, value of $\pm$ 1 denotes perfect correlation, $\pm$ 0.9 . . . 0.7 denotes strong correlation, $\pm$ 0.69 . . . 0.4 denotes moderate correlation, $\pm$ 0.39 . . . 0.1 denotes weak correlation, and $\pm$ 0.09 . . . 0 denotes no correlation [73].

Spearman's rank correlation is a nonparametric alternative to Pearson's correlation and is used to assess the monotonic relationship between two variables, making no assumptions about the underlying data's linearity or normality [71]. A monotonic relationship is one that is consistently unidirectional but is not necessarily linear (e.g., an exponential relationship). The reason for applying Spearman's correlation in addition to Pearson's correlation was to account for possible non-linearity and non-normality in the data. Spearman's correlation was used to compare the results with the Pearson's correlation results and therefore to not rely solely on the assumption of linearity in finding the best survey design features to estimate survey quality.

For calculating Spearman's correlation coefficient, this method ranks the values for both variables by ordering them from smallest to largest by one variable – preserving the observation pairs – and then encoding each value with its corresponding integer rank [74]. Subsequently, each observation has two paired ranks associated with it.

For instance, observation i has the value/rank 2 for variable X and the value 15 for variable Y. The interpretation of the Spearman's correlation coefficient (ρ) follows the same criteria as the Pearson's coefficient.

Spearman's correlation coefficient is calculated with the following formula:

$$\rho = 1 - \frac{6 \sum d_i^2}{n^3 - n}, \tag{2}$$

where

ρ is the Spearman's rank correlation coefficient,
d_i is the difference between the ranks for observation i,
n is the number of observations.

Multiple regression was the main method to evaluate how survey design features predict survey quality. The primary goal of multiple regression is to find the best-fitting straight line that describes how a dependent variable changes as the independent variables change. Multiple regression describes the change individually for each independent variable while controlling for the others and for the whole model defined as the combination of independent variables [75].

The equation for multiple linear regression model builds on the basic equation for a line and can be formulated as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \tag{3}$$

where for $i = 1, \ldots, n$ observations

y_i is the value of the response variable for observation i,
x_{ip} is the value of the p-th predictor variable for observation i,
β_0 is the y-intercept (the constant term),
β_p is the slope coefficient for the p-th predictor variable,
ε_i is the model's residual (error) for observation i.

Essentially, multiple linear regression seeks to draw a line by estimating the values of β_0 and $\beta_{1\ldots p}$ that minimize the residuals ε_i between observed/true values y_i and the values predicted by the linear model.

To minimize the residuals, the OLS method was used. OLS is a least squares estimator to find the regression line with the minimum sum of the squared residuals between the observed and predicted values and is one of the most common methods for estimating a regression line [73], [76].

Multiple OLS regression comes with five assumptions about the underlying data. It assumes (1) a linear relationship between the independent and dependent variables, (2) no autocorrelation between the residuals, (3) that the residuals are normally distributed, (4) the variance of the residuals to be constant ("homoskedastic"), and (5) no multicollinearity – a phenomenon where two independent variables in a regression

model correlate nearly perfectly between each other possibly leading to unreliable results [77], [78].

To comply with these assumptions, Durbin-Watson statistic, an indicator for autocorrelation across the model residuals, and quantile-quantile plots, a graphical tool to check residuals' normality [77], were assessed in selecting between regression models. Regarding the assumption of homoskedasticity, a heteroskedasticity robust variant of the OLS estimator was used to account for possible violation of the assumption in the data [75]. There are multiple variants of the heteroskedasticity robust estimator and the one chosen for this research is the so called HC3, being the most suitable for samples where N $\leq$ 250 [79].

Multicollinearity was examined through the correlation analysis and the variance inflation factor (VIF). VIF expresses the degree to which multicollinearity among the predictor variables degrades the precision of the regression result [80]. A rule of thumb is that the VIF should not exceed the value of 10 [81].

The reason for applying regression as the main analytical tool was that as an exploratory piece of research, it is sensible to start from examining linear, rather than non-linear, relationships between PPGIS survey design and quality features. Some deep-learning frameworks could capture and explain complex relationships between the features, but would have the danger of ignoring obvious survey attributes that impact survey quality linearly.

Applying **multiple** regression was justified by the fact that a survey contains multiple different content attributes. A survey fits into the concept of multiple regression by treating it as an entity containing multiple predictor features evaluated against one response feature at a time.

3.3.2 Explain phase

Using the optimized data drawn from the explore phase, a larger scale analysis was conducted for the explain phase. Instead of 37 surveys, correlation and regression analyses were applied on a set of 216 surveys. Adding more data makes the results more robust by decreasing the degree of sampling error (Dancey and Reidy, 2020). In addition, it is considered a good methodological practice to develop a model with one dataset and to test a hypothesis with another [82].

The explain phase followed a progression visualized in figure 2. Visual inspection was excluded from this phase, as was the experimentation with spatial data quality. Because the dataset was new, feature engineering was applied to the predictor features if the model from the previous phase wasn't statistically significant to obtain the most powerful regression model. Multiple models were built and variables were either added or removed based on theoretical and/or statistical criteria.

The predictor features were decided based on their statistical significance to the model's fit (p-value for the F-statistic) and effect on the model's adjusted coefficient of determination (adj. R^2). F-statistic is the ratio between the explained and unexplained variance within the regression model, taking into consideration the amount of data and number of predictor values. It is used to indicate the overall significance of the model [83].

The p-value indicates the likelihood that the observed effect could occur by chance under the null hypothesis. The null hypothesis (H_0) in this study assumes that survey design features have no statistically significant effect on survey quality. By convention, p-values smaller than 0.05 or 0.01 denote statistical significance [84]. This research adopts the more lenient 0.05 significance threshold due to its exploratory nature.

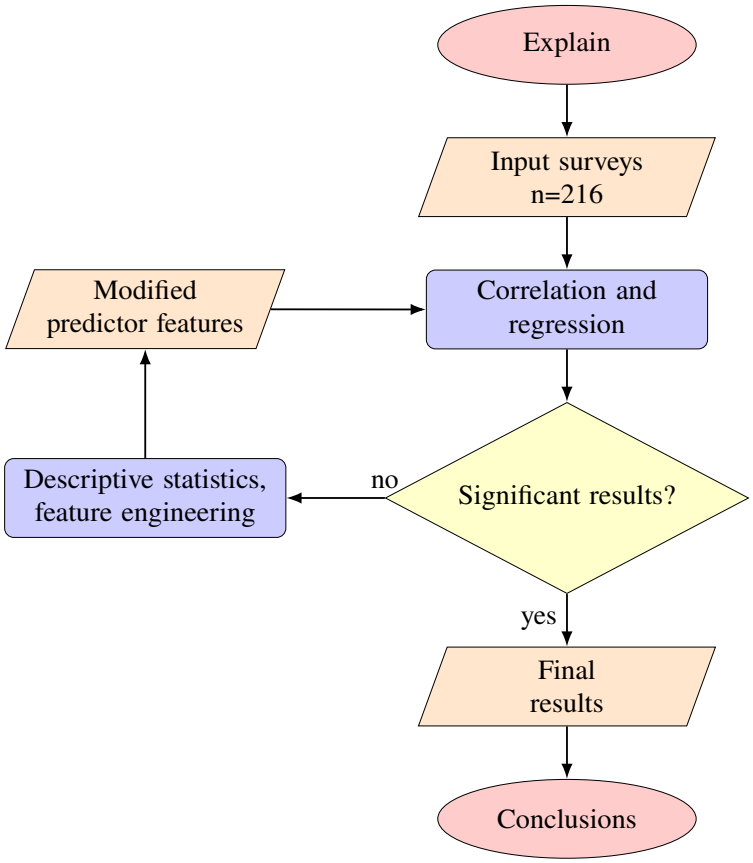


Figure 2: Flowchart of the explain phase.

The coefficient of determination measures how much of the variation in the model's outcome is explained by the variation in the independent variables [85]. The adjusted R^2 is a more robust version that accounts for the number of independent variables in the model [86]. Additionally, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were compared between models. The statistically significant model with the lowest AIC and BIC values combined with the highest adjusted R^2 value was selected. Lower AIC and BIC values indicate a better fit for a model [87].

In addition to selecting the final regression model via statistical metrics, some features were added or removed based on intuition and subjective deliberation. When examining the effect of survey design features on survey quality, the non-significant features can be interpreted as equally valuable as the significant features for understanding survey quality. If a design feature doesn't have a statistically significant effect on survey quality, that feature can be utilized without being concerned about its

impact on the quality. This is valuable information for a survey creator.

As an example of such deliberation, the geographic features (point, line, area) used in a survey and information about the used background maps were included into the analysis as a set of features by default. This decision is purely based on the nature of PPGIS: mapping and the use of geographic features is at the core of PPGIS. Even if found non-significant in terms of survey quality, excluding them from the analysis would leave out critical information regarding one of the main functionalities of PPGIS.

3.3.3 Predict phase

The predict phase was not concretely implemented in this research even though possible with regression analysis. Coming up with guidelines on how to create a high-quality PPGIS survey is, of course, aimed for future survey creation, but predicting the quality of a survey based on the results of this research would constitute a research on its own. The scope of this research was limited to the explanation of PPGIS survey design features' impact on survey quality, and thus the results are not extrapolated into predictive scenarios.

4 Results

The structure of this chapter reflects the two phases of data analysis. It starts from the results of the explore phase, then moves onto the main results acquired from the explain phase.

4.1 Explore phase results

This section presents the findings from the initial examination of 37 Maptionnaire surveys, answering the two research sub-questions about survey design and spatial data quality. This section has three parts. First, the results of operationalizing the survey design features are presented. Second, the experimentation results for spatial data quality are deliberated, and finally, the results of the preliminary correlation and regression analysis are shown.

4.1.1 Survey design features

Following the procedural flow presented in figure 1, the results from the visual inspection are presented first, followed by the combination of feature engineering and descriptive statistics.

The **visual inspection** resulted in a tabular dataset of features and comments about the inspected surveys (appendix D). Some critical observations were made during this phase regarding the final analysis.

Firstly, it was noted that the visual design of a survey has an impact on how engaging the survey is. Well thought out combinations of visual elements, such as colors and icons, and neatly arranged pages make it more pleasing to go through a survey. Therefore, the need to operationalize a survey's visual design was assumed. More detailed descriptions of feature operationalizations are given later on.

Secondly, a hypothesis of the survey theme's impact on the submission rate was formed: surveys dealing with mobility and transportation seemed to evoke a lot of emotions in the respondents, apparently leading to high response rates. Therefore, categorizing surveys into themes seemed significant. An existing classification scheme developed by Mapita Oy was applied for the categorization. Each survey was assigned one of the following themes: buildings & neighborhoods, inclusion & cohesion, mobility & transport, parks & recreation, urban planning & design.

It should be noted that the categorization of a survey into one theme is not straightforward: many surveys could be categorized into multiple different categories. For example, a survey that is collecting data about the accessibility of an urban park could be thematically categorized either into parks & recreation, mobility & transport, or even inclusion & cohesion. The survey theme feature used in the following sections should be understood with this in mind.

Thirdly, it was noted that data from the "programmatic elements" (see table 1 in section 3.1) are mostly invisible for the respondent and aren't stored readily in the Maptionnarie database. Thus, it was decided not to include them in the analysis.

However, it is worth mentioning that the device information element would potentially have a significant impact on a survey's submission rate and spatial data quality.

Lastly, the visual inspection indicated that user experience is an important survey design feature. Lengthy surveys had a high cognitive load, even if they were visually pleasing. Excessive number of pages or elements on one page made the user experience frustrating. Regarding cognitive load, if a survey page took more than a few seconds to load due to large image files or computationally heavy map overlays, one was easily distracted from the main task of the survey. Zooming in onto a map also became frustrating if each change in the zoom level prompted seconds of loading time. Therefore, measures for survey length and page loading times were considered potentially significant features regarding submission rate and zoom rate.

After the visual inspection, potential survey design features were **feature engineered** to get a fitting dataset for correlation and regression analysis. The process followed the methods described in the previous chapter: Potential features were retrieved from the Maptionnaire database with SQL. (Postgre)SQL allows to use aggregate functions like count, sum, average, etc. to combine raw data into sensible features. Features, such as the survey background map, were categorized and one-hot encoded to enable their use in quantitative analysis. Through testing, the features were optimized (e.g. median values were used instead of averages to tackle outliers) as part of the engineering process.

When some potential features were engineered, **descriptive statistics** helped to get a better understanding of the features and to assess their fitness as variables for regression analysis.

First, some general trends about the whole dataset were charted. Figure 3 shows the frequency of surveys that used a specific element from the Maptionnaire element library. This gave an idea of the elements that are used in practice and, thus, an indication of their significance to be included in the analysis. If an element is not used, there's no reason to analyze its impact on the response features.

A few observations can be highlighted from figure 3. Firstly, all surveys in this sample used paragraphs (text content created by the survey creator) as an element, and no survey used a geobudgeting point (enables the respondent to show how they would allocate resources such as funds or CO2 emissions), datepicker (respondent chooses a date from a calendar), audio (an audio file), youtube (an embedded YouTube video), discussion (an element allowing for discussion between participants), nor panel tracker (a way to restrict who has access to the survey) element. Secondly, it can be observed that the geopoint feature is by far the most popular geographic feature when compared to other mapping elements; 34 surveys used geopoints, whereas only 13 used the second most used geographic geoline feature.

The results guided the analysis away from considering every single element as a predictor feature in the research. For example, a decision was made against analysing the effect of audio files or embedded videos on submission rate by hypothetically increasing the respondent's cognitive load. It would be highly likely that not enough surveys which use an audio or video file could be gathered for the data to be representative of such effect.

In addition to examining used elements across surveys, the usage of background

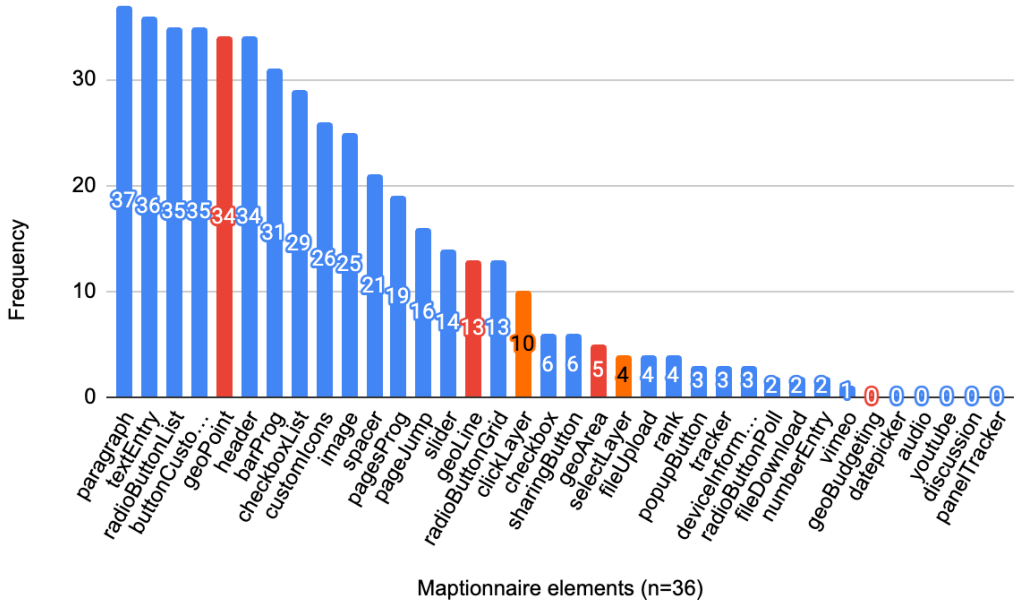


Figure 3: Frequency of surveys in the explore sample (n=37) using a Maptionnaire element. Mapping elements are highlighted in red (point, line, area, geobudgeting) and orange (select elements).

maps was looked into. Figure 4 shows the distribution of background map usage across the surveys. Full encoding of the maps is given in appendix E.

Figure 4 indicates a concentration in map selection across surveys. 24 surveys used map MapTiler Streets (115) as the background map in a survey. The next most used map type with 7 surveys using it was an interactive image. Interactive images are user imported image files that respondents can pin and draw on. Additionally, figure 4 underlines that some surveys use multiple different background maps instead of using a map once, or one map multiple times.

Because of these observations and the large number of possible maps, it was decided to operationalize the use of maps not through individual maps, but to compare whether a survey uses a single map or multiple different maps. A binary feature was created, being true (= 1) if two or more different maps were used as a background map for mapping questions, and false (= 0) if only one map was used.

As well as looking into larger trends in the data, univariate analyses were conducted while operationalizing and optimizing the survey design features. In figure G1, the frequency of survey default page layout is plotted. 32 surveys used the narrow page layout as the default layout and 5 surveys used the wide layout.

Based on the low variance in page layouts and the realization that the choice of a default layout doesn't dictate a consistent use of layouts – surveys with a default narrow layout can have some pages with a wide layout – it was decided that the page layout feature was not an insightful survey design feature to be included for correlation and regression analyses.

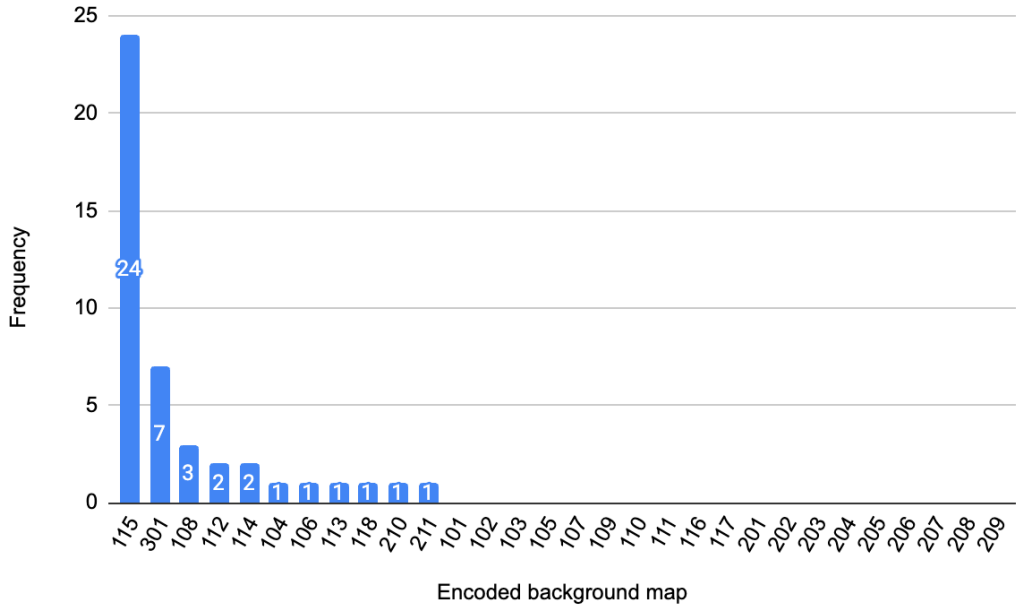


Figure 4: Frequency of surveys in the explore sample using a specific background map. 24 surveys used map 115 (MapTiler Streets), 7 surveys used map 301 (images as maps), etc. Five surveys used two different and one survey used three different background maps which totals to 44 maps used across the 37 surveys. Encoding scheme: 101-118 = Maptionnaire's OSM-based system maps, 201-211 user imported maps (e.g. a wms map), 301 = user imported image as a map (e.g. a ground plan).

This example illustrates the iterative nature of the feature engineering process visualized in figure 1. A considerable amount of time was used to come up with sensible and measurable features.

Finally, histograms were drawn to understand the distributions of variables. For instance, the total popup count for a survey is charted in figure G2 having a right skew. The figure exemplifies that some features were not normally distributed, having a potentially significant effect on Pearson's correlation results (see 3.3.1).

The set of engineered survey design features predicting the survey quality is given in section 4.1.3. Before that, a short deliberation of the experimentation process for spatial data quality is given.

4.1.2 Survey quality features

Survey quality was operationalized as two response features: submission rate and spatial data quality. Selecting submission rate as a quality feature was justified by theory and intuition (see section 1.1). Using zoom rate as the spatial data quality proxy, however, was a result of an experimentation process where three other proxies – standard distance proxy, AI assisted proxy, and OSM proxy – were investigated before deciding on zoom rate. The proxies were introduced in the previous chapter

and the focus here is to show why these proxies made way for zoom rate as the proxy for spatial data quality.

The proxy based on **standard distance** proved to be too simplistic as a measure for spatial data quality due to its sensitivity to dispersion (and potential outliers) of data within a geographical area. As seen in figure 5, a survey concerning Australia as a whole would have a low spatial data quality mainly due to the geographical setting of the survey. Thus, dispersion as a sole metric is not a good indicator of spatial data quality.

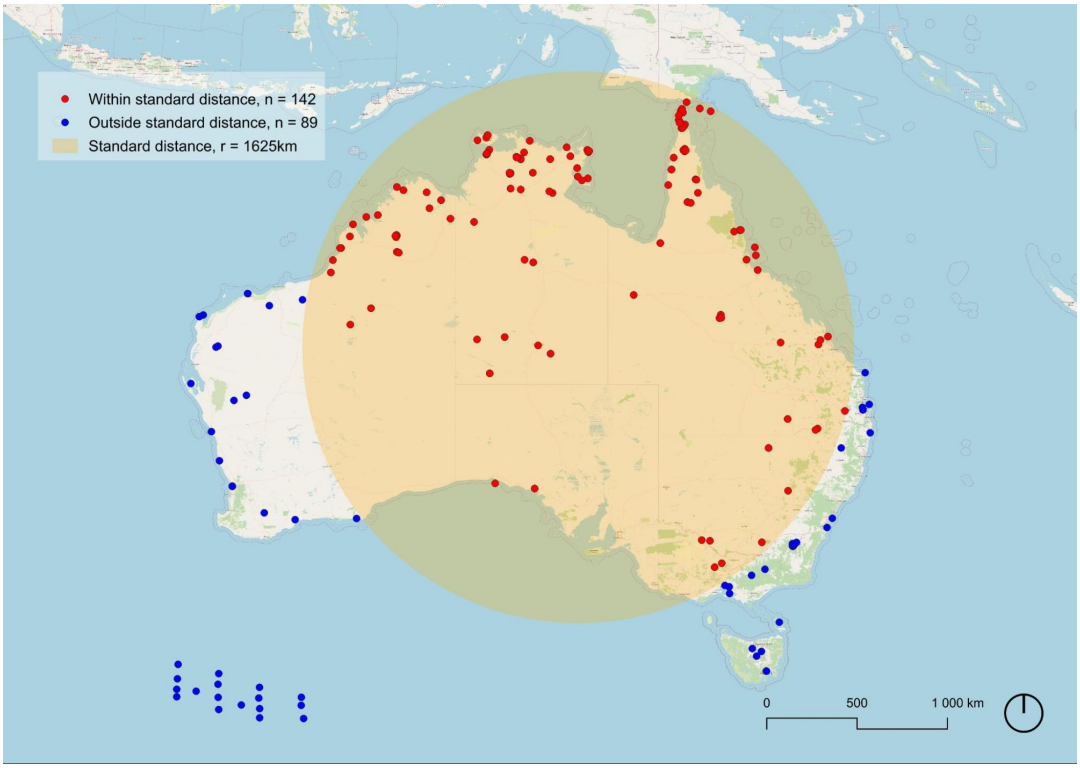


Figure 5: An example of the unfitness of standard distance as a proxy for spatial data quality. The blue and red points are responses given to a Maptionnaire survey. No data cleaning was done for the data and assumably the blue data points in the left bottom corner of the figure are outliers impacting to the calculation of standard distance. The orange circle is drawn with the standard distance as the radius. Because Australia's population and points of interest are mainly concentrated on the coastal areas, data points in cities like Sydney or Melbourne on the south-eastern coast of the continent would be considered as "inaccurate" by default with the proxy.

Just like with standard distance, it was understood fairly quickly that the proxy based on **AI assistance** was flawed for this research. Firstly, LLMs can hallucinate, i.e. provide nonsensical information [88], so relying purely on a threshold value determined by an LLM is problematic. Going through each survey to confirm that a value is reasonable would be time consuming and make the use of AI for automation redundant. Secondly, the answers an LLM gives are difficult to replicate. Retrieving scientifically robust results would be impossible, and so the idea of using an LLM this

way was dropped. Clarifying the attempted use of AI, an example of the prompt used and its example answers are given in appendix C.

The **OSM-based proxy** (see appendix F) showed more potential. Defining spatial quality not through an expert-derived dataset or threshold value, but through agreement between survey respondents fits the foundational idea of PPGIS to let locals be the experts of their living environments. Quality is defined by the survey responses and their associated OSM tag keys, essentially shifting the definition for quality to the survey respondents.

As seen from figure 4, it is a reasonable assumption that most Maptionnaire surveys use an OSM-based background map instead of a custom map, making an OSM proxy practically feasible. Figure 6 visualizes OSM hits queried with one survey response from a survey through the Overpass API. The five most frequent OSM tag keys queried with the line feature were surface (10), footway (6), building (4), lit (3), and highway (2).

Figure 6: A visualization of the OSM hits queried with a line feature. The red dashed line defined by the red coordinate points is one survey response. The blue lines and areas are the OSM hits for the intersecting response line visualized in QGIS.

Even though promising, the proxy was flawed in its current state. One major flaw was that querying the OSM database is different for each type of geographical feature. A line and an area feature are large features that often intersect or enclose multiple OSM data features, whereas a point feature needs a user-defined buffer with which to query data. This difference easily leads to more data being queried with line and area features than with point features, favoring line features in measuring data quality.

Based on these experiments, it was decided that zoom rate is the most robust proxy for spatial data quality. Developing the OSM proxy further is discussed briefly in the fifth and final chapter 5.

4.1.3 Preliminary correlation and regression

After using descriptive statistics to understand and screen for fit and unfit survey design and quality features, a set of features was used for a preliminary correlation and regression analysis. This section presents the correlation results first and then moves onto the regression results. The first engineered set included 11 survey design features and two survey quality features.

The predictor features included the total visitor, page, and word count of the survey contents; the median submission time in minutes; whether multiple maps were used; the popup count; the average element and overlay counts per page; the overlay file size (kB) per map page; and the average number of geographic and select elements mapped per respondent. The response features were submission rate and zoom rate.

The geographic features used in answering mapping questions (point, line, area) and the survey theme, even though mentioned as desirable survey design features in sections 3.3.1 and 4.1.1, were not part of this dataset. The number of binary variables produced by one-hot encoding these features would inflate the number of features and decrease the validity of the results due to lower degrees of freedom in the model (explained in more detail below). As an example, one-hot encoding five survey themes creates four features for each observation (more on this in section 4.2).

Figure 7 presents the Pearson's and Spearman's correlation coefficients between each pair of features. The interpretation for the matrix is explained in detail in the figure caption.

By combining the Pearson's and Spearman's coefficients into a single figure, it visualizes the difference between the two methods for measuring correlation: Pearson's r is reliant on the data linearity and normality, whereas Spearman's ρ can detect non-linear, monotonic associations between non-normally distributed features.

The Pearson's correlation results between the predictor features will be looked at first, followed by a few remarks about the Spearman's correlation results. The purpose is to investigate the relationship between predictors to chart potential dropout variables and to preliminarily check the possible multicollinearity of the predictors (calculated more rigorously in the regression analysis phase with VIF). The second purpose is to check whether Spearman's correlations reveal any non-linear relationships of variables that may inflate regression coefficients based on the Pearson correlations.

Starting with the **Pearson's** matrix, there is only one feature pair with a strong correlation. This suggests that there is no strong multicollinearity between this set of predictor features, justifying the subsequent regression analysis.

The average number of geographic features mapped by a respondent correlates strongly and positively ($r = 0.7$) with the median time it takes to submit a survey. This suggests that the more time respondents use on a survey, the more mapped responses they give, and vice versa. Moderate correlations can be observed (five strongest correlations highlighted here) between total popup count and word count

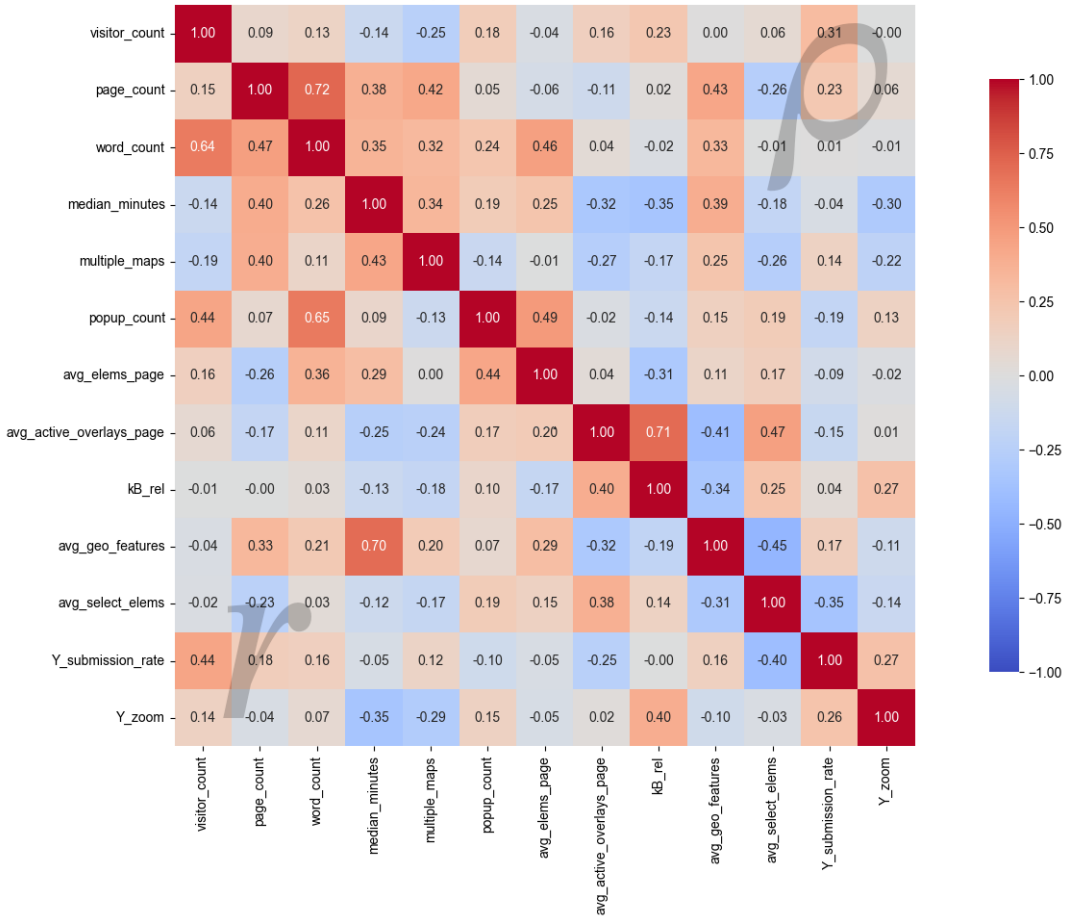


Figure 7: Pearson's and Spearman's correlation matrix calculated with the explore phase features (n=13). The rows and columns list the same variables, in the same order. The crossing point of each pair represents the correlation coefficient between them. The diagonal (from top left to bottom right) always shows 1.00, because each variable is perfectly correlated with itself. The diagonal line in this matrix (and the other correlation matrices in this study) is also the divider between the results from the two methods: the coefficients to the left and underneath of the diagonal line are the coefficients (r) for Pearson's correlation, and the coefficients to the right and above of the diagonal line are the coefficients (ρ) for Spearman's correlation. Darker colors represent stronger correlation between features when compared to lighter shades. Red colors are positive correlations and blue colors are negative correlations.

($r = 0.65$); total word count and visitor count ($r = 0.64$); total word count and page count ($r = 0.47$); total visitor count and submission rate ($r = 0.44$); and median time to submit a survey and total page count ($r = 0.4$). The strongest negative correlation is between submission rate and the average select elements mapped per respondent ($r = -0.4$).

Moving onto the **Spearman's** correlation matrix, there are two feature pairs that have a strong positive correlation ($\rho > 0.7$): total word count and total page count, and

average count of active overlays per page and relative file size of overlays (kB_rel). The next three strongest correlations fall into the moderate correlation strength category and are between average elements per page and the total popup count in a survey ($\rho = 0.49$); average map overlays per page and the average count of select elements ($\rho = 0.47$); and total word count and average elements per page ($\rho = 0.46$).

On a general level, the two correlation methods produce similar results. No sign is flipped from minus to plus, or the other way around, and the depth and variety in the coefficient colors denoting the correlation strength are alike. There are more positive than negative correlations in general and the upper left corner of both matrices is more reddish than the bluer right bottom corner.

Even though similar, a couple clear differences can be pointed out. The correlation coefficient between word count and visitor count for Pearson's $r = 0.65$, but only $\rho = 0.13$ for Spearman's ρ with an absolute difference of 0.52. Another relatively large difference is between word count and popup count, $r = 0.65$ and $\rho = 0.24$, with an absolute difference of 0.41. These results are likely a result of skewed data: some features – like popup count (see figure G2) – are highly skewed and, thus, might have inflated Pearson's correlation results.

No strong correlations between a predictor feature and a response feature are present in either matrix. However, for most of the predictor features the correlations with response features seemed sufficiently large to justify the subsequent regression analysis.

The strongest correlation between a predictor feature and a response feature is between total visitor count and submission rate ($r = 0.44$). Looking into this result, a crucial detail was recognized: even though only moderately correlated, visitor count should not be used as a survey design feature. Submission rate for a survey is defined by the total number of survey visitors (= visitor count) and if the same feature (visitor count) is used both in the response feature and as a predictor feature, it creates a purely technical relationship between them. Therefore, visitor count was eliminated from the predictor feature pool at this point.

A few other predictor features were eliminated, as well. The total page count and word count were eliminated due to their moderately high correlation between each other ($r = 0.47$ and $\rho = 0.72$) and a subsequent reflection: page and word count effectively measure the length of a survey, but are worse metrics for that than the median minutes to submit a survey. A Maptionnaire survey creator can utilize an element that makes the respondent skip certain pages based on their answers. Thus, a survey with 30 pages might, in practice, have only 15 pages that a respondent views while answering it. This makes the total page count, and the highly correlating word count with it, imperfect measures for survey length.

Lastly, it was decided to drop the average count of active overlays and rely on the overlay file size per map page (kB_rel). This was based on their relatively high correlation $r = 0.40$ and $\rho = 0.71$ and on the consideration that if a survey uses many overlays, but those do not affect the loading time of a page negatively, it can be viewed as a positive aspect of distinctive survey design. The file size is a more straightforward measure of how heavily overlays are used on a given survey page and the possible effect on loading times.

After these selections, the feature set included 7 survey design features (predictors) and 2 survey quality features (response features) for the **multiple OLS regression** analysis. The included design features were the median submission time in minutes; whether multiple maps were used; the popup count; the average element count per page; the overlay file size (kB) per map page with overlay(s); and the average number of geographic and select elements mapped per respondent.

Table 2 shows the results and diagnostics of the multiple OLS regression for **submission rate**.

Table 2: Multiple OLS regression summary, diagnostics, and coefficients for **submission rate**.

Model summary			
No. observations	37	Log-Likelihood	14.019
Df residuals	29	Df model	7
R^2	0.221	F-statistic	3.969
Adj. R^2	0.033	F-statistic p-value	0.004*
AIC	-12.04	BIC	0.848

Model diagnostics			
Omnibus	0.752	Durbin-Watson	2.010
Omnibus p-value	0.686	Jarque-Bera (JB)	0.341
Kurtosis	3.064	JB p-value	0.843

Features	**Coefficient (β_p)**	**std. err.**	**p-value**	**[95% conf. int.]**
const	0.349	0.039	0.000*	[0.273, 0.425]
median_minutes	-0.065	0.056	0.240	[-0.174, 0.044]
multiple_maps	0.032	0.031	0.306	[-0.029, 0.093]
popup_count	-0.007	0.088	0.940	[-0.179, 0.166]
avg_elems_page	0.012	0.037	0.746	[-0.060, 0.084]
kB_rel	0.018	0.068	0.795	[-0.115, 0.150]
avg_geo_features	0.048	0.058	0.407	[-0.066, 0.162]
avg_select_elems	-0.066	0.037	0.071	[-0.138, 0.006]

Note: Heteroskedasticity robust standard errors are used (see 3.3.1).

* $p < 0.05$

Starting from the **model summary**, one can see that 37 surveys were included in the analysis from the number of observations. The degrees of freedom (df) for the residuals (no. observations (37) - no. predictors (7) - 1) and model (no. predictors) describe the extent of available data relative to the number of features that are to be estimated. If the degrees of freedom are low, the model estimates are more likely to be imprecise and low in statistical explanatory power [89].

Here, the degrees of freedom suggest that the model is too complex for the amount of data available and may be unreliable or overfitted. This will be addressed by adding

more data into the analysis in the explain phase. There are several suggestions as to how much data there should be to achieve robust results with multiple regression, but one of the most common "rules of thumb" is a ratio of 10:1, i.e., one should have 10 observations per a predictor feature [90], [91].

The model as a whole is statistically significant; F-statistic p-value < 0.05. Therefore, the null hypothesis (H_0) of predictor features having no statistically significant effect on submission rate (see 3.3.2) can be rejected.

However, the adjusted R^2 is only 0.033, indicating that even though statistically significant, only 3.3% of the variance in the submission rate can be explained by the variance in the selected predictor features. Neither AIC or BIC are interpretable on their own and will be used as part of the selection criteria for following models in the next section.

The model **diagnostics** indicate that the model complies with the regression assumptions laid out in section 3.3.1. Being close to two, the Durbin-Watson statistic = 2.010 indicates no significant autocorrelation between the model residuals (Nerlove and Wallis, 1966). Omnibus, Jarque-Bera, and kurtosis are all measures for assessing the residual normality in this model. Both omnibus and Jarque-Bera are non-significant (p-values > 0.05), suggesting that the residuals are normally distributed.

Kurtosis at 3.064 reveal that the residuals have outliers creating long "tails" in the distribution, but looking at the quantile-quantile plot on the left side in figure 8, the residuals seem mostly normally distributed by following the theoretical normal distribution. This is a positive sign for model robustness and the validity of the results drawn with the model.

The **coefficients** table presents the effect of individual predictor features on submission rate. All the predictor features are named on the left, followed by their coefficient β. The coefficient β indicates the change in submission rate (response feature) for a one-unit increase in the corresponding predictor feature, holding all other features constant [83].

For each non-binary **predictor feature**, all the observations were standardized ($z_i = (x_i - \mu)/\sigma$, where x_i is the ith original observation, μ is the average for that feature, and σ is the feature's standard deviation) before the regression analysis due to their non-uniform scales. Standardization transforms data so that it has an average of 0 and a standard deviation of 1. The binary features weren't standardized, because they would lose their categorical meaning (the absence or presence of a feature) if standardized [92]. The response features weren't standardized either in order to make the interpretation of the regression coefficients slightly more intuitive.

A standardized coefficient tells the predicted change in the response feature for a one standard deviation increase in the specific predictor feature [93]. However, making exacting interpretations based on the standardized coefficients, like the one above, has been critiqued [94], [95]. Because standardization is reliant on a feature's standard deviation, differences in standardized coefficients may reflect not only genuine differences in their effect size but also differences in the features' standard deviations.

Therefore, the regression coefficients will be treated as indicative, rather than absolute in this research. Interpretations about the direction of a linear relationship and the relationships' relative strength are made, but no definite nor absolute estimates

should be drawn.

Only the model's constant term ("const") is statistically significant (p-value < 0.05). The constant term coefficient (β = 0.349) represents the y-intercept for the regression line when all the predictor features are zero (see β_0 in equation 3). However, observing all the predictor values to be zero in the context of this research is purely theoretical (e.g., no survey would ever have zero elements on its pages), and thus there is no sensible interpretation of the constant term.

The constant term being the only statistically significant "feature" suggests that no single predictor feature has a significant effect on submission rate when considered as part of this feature set. For now, this non-significance is regarded as an impact of the low number of observations in the dataset. The features do imply significance as a group of features and should therefore be explored further.

The coefficient standard error ("std. err.") measures the precision of the respective β coefficient. It tells how much the estimated β coefficient is likely to vary from the true population coefficient if the sampling and estimation process were repeated many times. The 95% confidence interval provides a range of plausible values for the true population coefficient for each feature based on the sample coefficient [83].

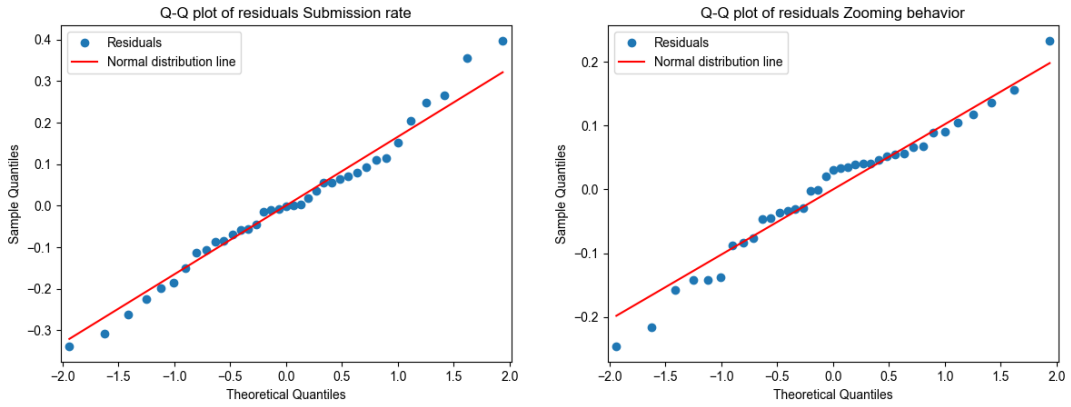


Figure 8: Quantile-quantile (Q-Q) plots of the response feature residuals. On the left the plot for submission rate and on the right for zoom rate. The red line is a theoretical normal distribution, the blue points are the residuals. Strong residual deviations from the normal distribution line would indicate non-normality.

Table 3 presents the regression results for **zoom rate** with the same predictor features. Contradictory to submission rate, the model is not statistically significant for estimating zoom rate (F-statistic p-value = 0.077). However, the adjusted coefficient of determination (adj. R^2) is clearly higher at 18% than for submission rate at 3.3%. With the model being close to statistical significance and expressing stronger explanatory power, one reasonable interpretation would be that the features chosen for this set estimate zoom rate more readily than submission rate.

Again, the model shows robustness in terms of its diagnostics. The omnibus and Jarque-Bera are once more non-significant suggesting residual normality, and the Durbin-Watson statistic is close to two indicating no autocorrelation between the

Table 3: Multiple OLS regression summary, diagnostics, and coefficients for **zoom rate**.

Model summary			
No. observations	37	Log-Likelihood	31.864
Df residuals	29	Df model	7
R^2	0.340	F-statistic	2.088
Adj. R^2	0.180	F-statistic p-value	0.077
AIC	-47.73	BIC	-34.84

Model diagnostics			
Omnibus	1.267	Durbin-Watson	1.520
Omnibus p-value	0.531	Jarque-Bera (JB)	0.808
Kurtosis	3.015	JB p-value	0.668

Features	**Coefficient (β_p)**	**std. err.**	**p-value**	**[95% conf. int.]**
const	0.868	0.024	0.000*	[0.822, 0.915]
median_minutes	-0.065	0.025	0.010*	[-0.115, -0.015]
multiple_maps	-0.007	0.031	0.826	[-0.067, 0.054]
popup_count	0.017	0.026	0.513	[-0.034, 0.067]
avg_elems_page	0.004	0.032	0.893	[-0.058, 0.066]
kB_rel	0.048	0.018	0.006*	[0.013, 0.083]
avg_geo_features	0.038	0.062	0.542	[-0.083, 0.159]
avg_select_elems	-0.012	0.020	0.546	[-0.050, 0.027]

Note: Heteroskedasticity robust standard errors are used.
* $p < 0.05$

residuals. On the right side of figure 8 one can see that the residuals for the zoom rate also roughly follow the normal distribution line.

Because the model as a whole is statistically non-significant, the coefficients are not explored in detail. However, it is good to note that there are two statistically significant features in explaining the variance in zoom rate: median minutes (p-value = 0.01) and relative overlay size per page (kB_rel, p-value = 0.006). This implies that the moderately strong correlations between these two features and zoom rate (figure 7) are strong enough to drag the whole model close to statistical significance. This was confirmed by running the model without the overlay size feature: the F-statistic p-value rose to 0.158, denoting clear non-significance, and the adjusted coefficient of determination (adj. R^2) fell from 18% to 5.2%.

There seems to be a clear linear relationship between the overlay size and zoom rate in this dataset. However, examining this relationship in more detail will be ignored for now by recognizing the low number of observations in this dataset. If similar results would be obtained from the regression analysis in the next phase, the association should be assessed thoroughly.

In addition to looking at the correlation matrices to understand possible multicollinearity between the predictor features, VIF (see 3.3.1) values were calculated as another model diagnostic. The VIF values for the features were the following: median_minutes = 2.49; multiple_maps = 1.36; popup_count = 1.31; avg_elems_page = 1.47; kB_rel = 1.12, avg_geo_features = 2.36; and avg_select_elems = 1.25. All VIF values are under 2.5, suggesting no significant multicollinearity between the predictor features.

The results from the explore phase give a solid foundation to move into the explain phase. Other than having a limited sample size, no clear limitations seem to be present in the regression models.

4.2 Explain phase results

4.2.1 Testing the explore phase model

After the preliminary correlation and regression analysis, a dataset of 216 surveys were retrieved from the Maptionnaire database. This number of observations allows for more complexity in the model based on the previously mentioned 10:1 ratio: the model could have 21 predictor features to still be considered statistically robust.

Like was done for the smaller explore dataset, descriptive statistics were also charted for the full dataset (appendix G). The used survey elements (figure G3) and background maps (figure G4) were plotted and they follow similar trends for the full dataset as for the exploratory dataset. The survey theme distribution was also plotted (figure G5), and the most frequently represented theme in the surveys was mobility and transport (47.2%).

For the survey elements, the trend is similar as it was for the smaller dataset: paragraph and text entry elements are used by almost every survey, and the point element is clearly the most utilized geographic element. Additionally, the four elements that were not used by any survey were also unused in the smaller dataset (geo-budgeting points, date picker, audio, and panel tracker).

The averages for submission and zoom rate were also calculated (figure G6) for the full dataset to get an understanding of the rates that a survey creator might expect when deploying a Maptionnaire survey. The average submission rate was 33.6% and the average zoom rate 66.7%.

As described in the previous chapter (see figure 2), no visual inspection nor experimentation with spatial data quality was done with the new dataset. Correlation and regression analyses were conducted with the same set of features that were used in the final explore phase analysis (7 predictor features and 2 response features).

Running Pearson's and Spearman's correlation analyses with the larger dataset resulted in weaker correlations for each feature pair (figure G7) when compared to the smaller dataset. Only weak correlations ($\pm$ 0.39 . . . 0.1) were observed between the features. The Spearman's ρ = 0.38 between average geographic features mapped by a respondent and the median time to submit a survey was the strongest observed correlation. The two correlation methods also produced more similar results with the bigger dataset; the largest absolute difference in corresponding coefficients was 0.16,

contrasting to the 0.52 absolute difference within the explore dataset.

The larger dataset flipped the regression results when compared to the explore phase: submission rate (table G1) as a function of the seven predictor features wasn't statistically significant (F-statistic p-value = 0.143), but zoom rate (table G2) turned out to be highly significant (p-value = 5.00e-8). Submission rate's adjusted coefficient of determination (adj. R^2) went down to only 2% from 3.3%, but zoom rate's adjusted R^2 rose from 18% to 20%.

The diagnostics for submission rate do not suggest any problems within the model, the predictor features as a group simply do not have enough explanatory power in predicting the variance within submission rate based on the collected data. However, the diagnostics for zoom rate are alarming. The Q-Q plot for zoom rate shows that the model residuals deviate clearly from the normal distribution (see right side of figure G8) and thus suggests residual non-normality. This is confirmed by both omnibus and Jarque-Bera tests being significant with p-values 0.005 and 0.01 respectively. Consequently, the model's p-values for zoom rate cannot be fully trusted.

Based on these results with the larger dataset, the defined seven predictor features do not explain submission rate's variance and the results for zoom rate are non-robust. Thus, a handful of decisions were made in order to improve the explanatory power of the model.

4.2.2 Creating an enriched model

First, allowed by the increased degrees of freedom in the larger dataset, **nine predictor features were added** to address the model's non-significance in explaining submission rate. Like mentioned in section 3.3.2, evaluating how a geographic feature impacts survey quality is considered as critical in understanding PPGIS survey design, and thus, geographic features (point, line, area) were now added into the predictor pool. In addition, features defining the survey theme, element customization, and survey progression were included at this point (see table 4 for full explanations).

The geographic features were one-hot encoded and added into the predictor feature pool as binary variables, which added three features. The survey themes were one-hot encoded, as well, resulting in five binary features. However, only four theme features were added into the predictor pool by excluding the buildings and neighborhoods theme. If all five features would have been added, there would have been perfect linearity between them; when a survey has one of the theme features as true (= 1), all the other ones are false (= 0). This would have violated the regression assumption of no multicollinearity. There is no such artificial multicollinearity between the geographic features, because a survey can utilize multiple different geographic features.

The element customization and survey progress features were also added as binary features indicating whether or not there are element customization (utilization of custom icons) and whether or not a progress indicator has been used (pages or a progress bar are shown to the respondent).

Second, because the total popup count was one of the worst predictors for both submission rate and zoom rate in the first regression model (see tables G1 and G2), it was decided to be modified: **instead of the survey's total popup count, an average**

popup count per page was engineered as the feature to express the use of popups. This was thought to be a reasonable modification, because having evenly distributed popups, even if many, in a survey might not have any effect in answering, but having multiple popups on one page might increase the cognitive load for the respondent.

Third, **two features were excluded completely**: average geographic features mapped per respondent and average select features mapped per respondent. Just like total popup count, they turned out to be poor predictors for submission rate and zoom rate. Additionally, and perhaps more importantly, it was realized that these two features aren't exactly design features, but result features of a survey. A survey creator doesn't have direct control over the number of answers the respondents give – the number of given answers is more correctly a potential indicator for survey quality, rather than a design feature. Thus, they were dropped from the predictor feature pool.

Fourth, **mirroring and a natural logarithmic transformation** were applied for zoom rate to address the non-normality of residuals. The transformation was only done after the correlation analyses were calculated, targeting the regression analysis.

Different data transformations can be applied to make data applicable for parametric statistical analysis. Logarithmic and power (square, cubic) transformations are examples of common data transformations – logarithmic transformations being most suitable for right skewed (long tail to the right) and power transformations for left skewed data [96]. On the left of figure 9 is the original zoom rate histogram with a left skew, indicating that the most suitable transformation would be a power transformation.

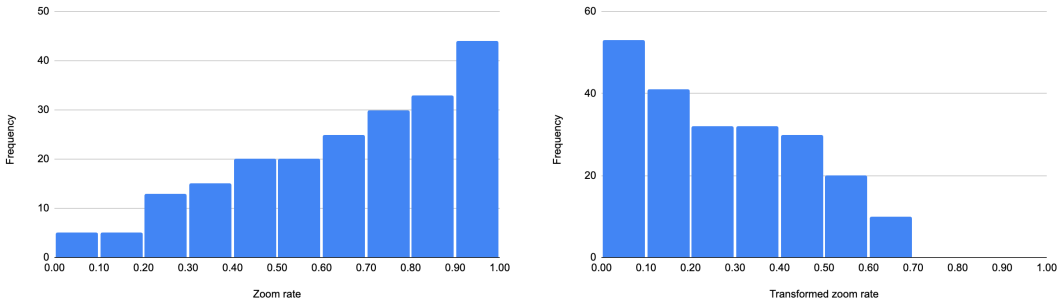


Figure 9: Histograms for zoom rate data. On the left is the original data, on the right is the data after mirroring and log transformation. Bucket size is 0.1.

After testing with power transformations and a natural log transformation, the log transformation was selected. The AIC and BIC values were compared between the models (with power transformation AIC = 47.10 and BIC = 97.72, with log-transformation AIC = -175.0 and BIC = -124.3), and the omnibus and Jarque-Bera tests were evaluated. Both test results were higher for the power transformed model, signifying larger deviations from normality.

Before taking the natural logarithm, the zoom rate data was mirrored to create a right skew: $0 - y_i + \text{const.}$, where y_i is one observation and $\in [0, 1]$, and const. = 2. Adding a constant value ensured that the logarithm base was positive. Then, the natural logarithm was taken to give a distribution closer to normal distribution (right

side of 9) when compared to the distribution with raw zoom rate values or the values obtained with power transformations.

Mirroring the response feature has a critical effect on the interpretation of the regression coefficients: a negative coefficient for a predictor feature estimating zoom rate is, actually, a positive coefficient and vice versa. In addition, the log-transformation impacts the interpretation of the coefficients. Following Benoit [97], assuming small coefficients ($e^{\beta_p} \approx 1 + \beta_p$), the following interpretation for the coefficients can be used: $100 * \beta_p$ is the expected **percent point change** in the response feature for a unit (with standardized coefficients unit is one standard deviation) increase in the predictor feature, holding the other predictors constant.

As can be seen from figure 9, the transformed zoom rate data is approximately within the same range as the original data was (original $[0, 1]$; transformed $[0, 0.67]$). Therefore, the log-transformation – in this specific case – doesn't have a critical impact on the interpretation of the coefficient magnitude: because zoom rate (and submission rate for that matter) is a rate between 0 and 1, the interpretation of an expected percentage change would've also applied for the non-transformed data.

4.2.3 Final results

With the mentioned changes to the model and predictor features, the final model consisted of 14 predictor features. Sorted to match the following results, the features and their data types are detailed in table 4.

The correlation results for the final set of features are presented in figure 10. As in the previous correlation matrices, the Pearson's r coefficients are presented in the lower left half of the matrix, and the Spearman's ρ are in the upper right half. Yet again, the two methods produce similar results for correlation. The largest absolute difference of coefficient values for a feature pair is between the feature expressing whether the respondent sees their progress in a survey ("progression") and the relative size of map overlays in the survey ("kB_rel"): Pearson's r = -0.08 and Spearman's ρ = 0.13, the difference being $|-0.08 - 0.13| = 0.21$.

The strongest positive correlation is between the average number of elements per page and the average number of popups per page (r = 0.53) and the strongest negative correlation is between two survey themes: mobility and transport ("m_t"), and inclusion and cohesion ("i_c") (r = -0.46). These are moderate correlations and do not suggest critical multicollinearity. This was also confirmed by the VIF values for the predictors in this dataset: the maximum VIF value was 3.75, suggesting only weak multicollinearity between the predictor features.

After the correlation analysis, the final regression analysis was run. Table 5 shows the regression results for submission rate and table 6 shows the results for the log-transformed zoom rate (the mirroring effect has been corrected in the table).

With the mentioned modifications, the model for **submission rate** is now highly significant (p-value = 1.00e-06). Even though highly significant, the model doesn't reach a high explanatory power: the variance in the predictor features explains 11.3% of the variance in submission rate. The diagnostics signify robustness with no significant non-normality, also confirmed by the Q-Q plot on the left side of figure 11.

Table 4: Explanations for the features used in the final analyses. Interval notation is used for the feature ranges/data types.

Abbreviation	Explanation	Range
median_minutes	Median minutes to submit a survey.	$[0..\infty)$
avg_popup_page	Average number of popup windows on a survey page. Total number of popups divided by page count. Popups are used to give and gather extra info for a mapping response.	$[0, \infty)$
avg_elems_page	Average number of visible elements on a survey page.	$[0, \infty)$
kB_rel	The map overlay file size (in kilobytes) per page with map overlay(s). Calculated as the total sum of map overlay sizes divided by number of pages with active map overlays.	$[0, \infty)$
multiple_maps	Are multiple different background maps used in the survey?	$\{0, 1\}$
customization	Are non-default icons used?	$\{0, 1\}$
progression	Do respondents see how much of the survey is completed?	$\{0, 1\}$
geo_point	Is a point feature utilized in the survey?	$\{0, 1\}$
geo_line	Is a line feature utilized in the survey?	$\{0, 1\}$
geo_area	Is an area feature utilized in the survey?	$\{0, 1\}$
i_c	Is the theme inclusion and cohesion?	$\{0, 1\}$
m_t	Is the theme mobility and transportation?	$\{0, 1\}$
p_r	Is the theme parks and recreation?	$\{0, 1\}$
up_d	Is the theme urban planning and design?	$\{0, 1\}$
Y_submission_rate	The number of submitted surveys divided by the total number of survey visitors	$[0, 1]$
Y_zoom	The number of mapping responses that use a zoom level higher than the default zoom level(s) divided by all mapping responses	$[0, 1]$

Each predictor feature coefficient is rather small (max. $\beta = |-0.079|$). This is explained by the scale of submission rate itself ($[0, 1]$, where a theoretical maximum for a coefficient would be 1) and the model's overall low explanatory power: this set of features explains only a tenth of the variance in submission rate. One can expect small individual coefficients when the predictor features' combined explanatory power is weak.

There are three features that are statistically significant in explaining the variance in submission rate: median minutes to submit a survey (β = -0.029, p-value = 0.005), average number of popups per survey page (β = -0.036, p-value = 0.001), and the theme inclusion and cohesion (β = -0.079, p-value = 0.033).

Each of the three features have a negative coefficient. For the two first predictor

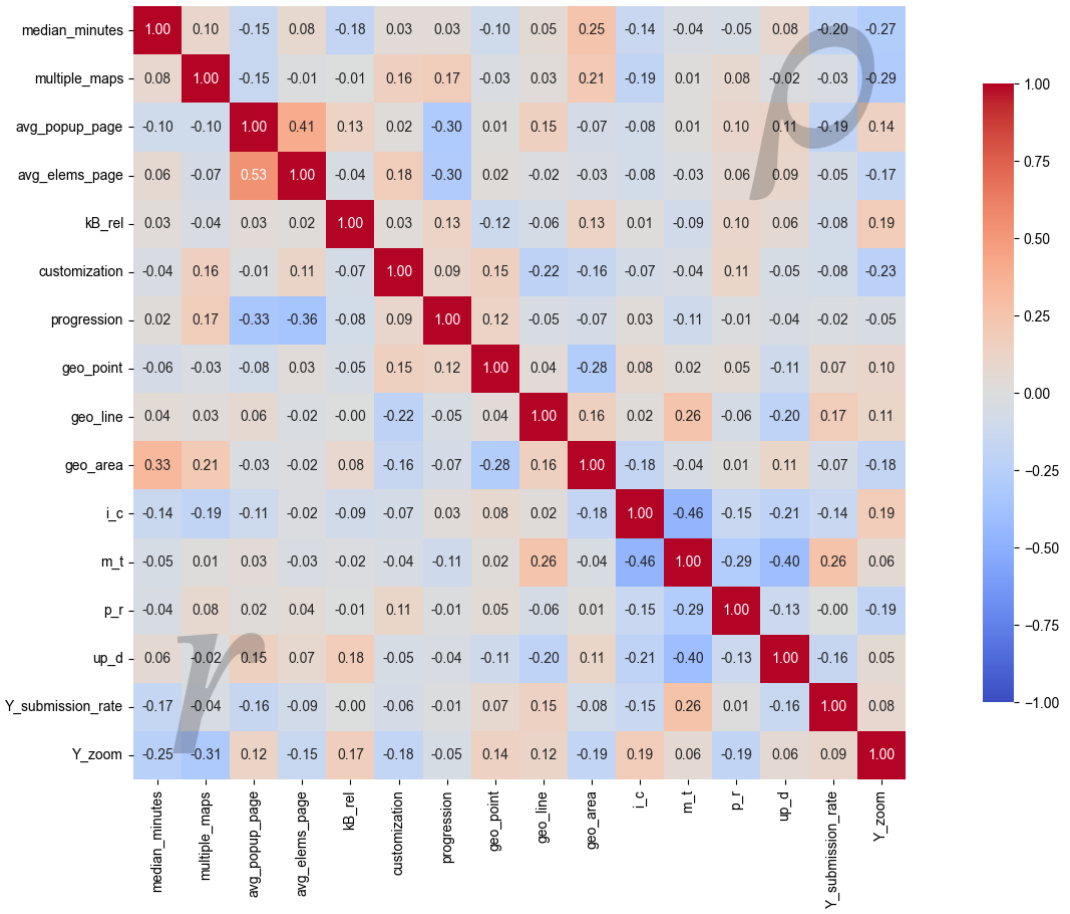


Figure 10: Pearson's and Spearman's correlation matrix calculated with the final feature set (n=16). The matrix shows the Pearson's correlation coefficient (r) value for each feature pair in the lower left half and the Spearman's coefficient (ρ) in the upper right half.

features, when there's a one standard deviation increase, the submission rate decreases for the (absolute) value of the coefficient, assuming all the other features being constant. For median minutes the standard deviation = 7.63 and for the average number of popups standard deviation = 2.66. Consequently, the interpretation for median submission time would be that if it increases about 7.5 minutes, the submission rate decreases by 0.029 (2.9 percentage points). Yet as mentioned earlier, the exact interpretation of the effect's magnitude can be misleading with standardized coefficients.

For the non-standardized binary theme feature, the interpretation is slightly different. If the survey theme is inclusion and cohesion, it is estimated that the survey rate decreases by 0.079 (7.9 percentage points), when all the other features are constant. It is good to note that the standard error and the 95% confidence interval for the theme feature are larger than for the two other features, signifying relatively large variance in this theme's effect magnitude on submission rate. Additionally, as stated in section 4.1.1, the survey theme feature is somewhat sensitive to subjective

Table 5: Final multiple OLS regression summary, diagnostics, and coefficients for **submission rate**.

Model summary			
No. observations	216	Log-Likelihood	132.07
Df residuals	201	Df model	14
R^2	0.170	F-statistic	4.340
Adj. R^2	0.113	F-statistic p-value	1.00e-06*
AIC	-234.1	BIC	-183.5

Model diagnostics			
Omnibus	4.225	Durbin-Watson	1.973
Omnibus p-value	0.121	Jarque-Bera (JB)	4.194
Kurtosis	2.943	JB p-value	0.123

Features	**Coefficient (β_p)**	**std. err.**	**p-value**	**[95% conf. int.]**
const	0.336	0.079	0.000*	[0.181, 0.491]
median_minutes	-0.029	0.010	0.005*	[-0.049, -0.008]
avg_popup_page	-0.036	0.010	0.001*	[-0.056, -0.015]
avg_elems_page	0.007	0.011	0.529	[-0.015, 0.028]
kB_rel	0.001	0.014	0.927	[-0.025, 0.028]
multiple_maps	-0.026	0.025	0.309	[-0.075, 0.024]
customization	-0.022	0.023	0.357	[-0.067, 0.024]
progression	-0.015	0.027	0.564	[-0.067, 0.037]
geo_point	0.033	0.078	0.675	[-0.120, 0.186]
geo_line	0.041	0.023	0.079	[-0.005, 0.086]
geo_area	-0.026	0.029	0.375	[-0.083, 0.031]
i_c	-0.079	0.037	0.033*	[-0.152, -0.006]
m_t	0.022	0.030	0.477	[-0.038, 0.081]
p_r	-0.001	0.048	0.977	[-0.096, 0.094]
up_d	-0.039	0.033	0.235	[-0.103, 0.025]

Note: Heteroskedasticity robust standard errors are used.
* $p < 0.05$

interpretation in the original categorization. Therefore, this result shouldn't deter one from creating surveys about inclusion and cohesion.

As was noted in section 3.3.2, one shouldn't completely ignore the non-significant coefficients. The regression results indicate that the geographic feature type (point, line, or area) chosen by the survey creator has no significant effect on the submission rate. Based on the frequency of used elements (figure G3), survey creators are cautious in using geographic elements other than the point element. However, at least when submission rate is concerned, there's no reason for such cautiousness. Additionally, the overlay file size, customization of a survey and showing a respondent the progress

they've made do not seem to affect submission rate in any significant way.

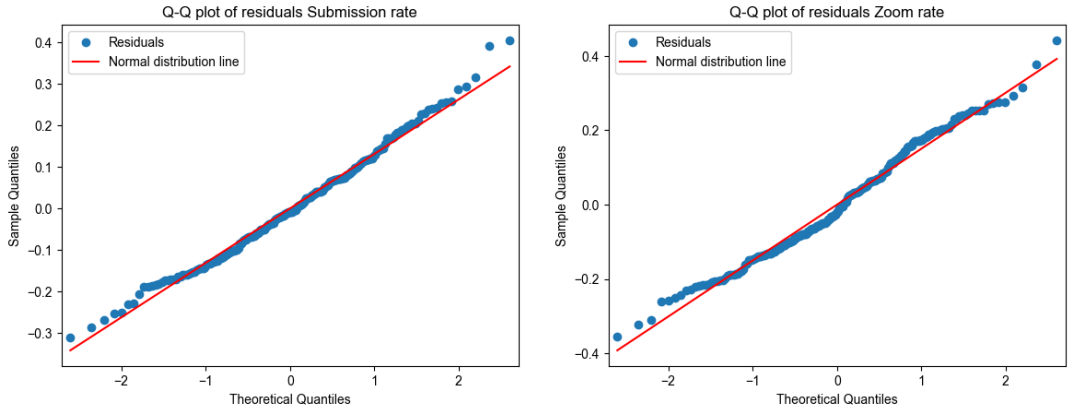


Figure 11: Quantile-quantile (Q-Q) plots of the response features' residuals. On the left the plot for submission rate and on the right for zoom rate.

For the **zoom rate**, the coefficients seem small enough ([97] doesn't define a clear threshold for the "smallness" of the coefficients) to utilize the interpretation mentioned in the previous section: $100 * \beta_p$ is the expected percentage point change in the zoom rate for one standard deviation increase in the predictor, given that the other predictor features stay constant. Take, for example, the largest absolute predictor feature coefficient (geo_point $\beta = 0.173$): $e^{0.173} = 1.18 \approx 1 + 0.173$.

Just like for submission rate, the whole model is highly significant (p-value = 4.33e-11). The explanatory power, however, is clearly stronger for zoom rate than for submission rate: the variance in the predictor features explain 28.1% of the variance in the zoom rate. The omnibus test and its significant p-value (0.032) suggest some skewness in the residuals, but the other test for normality, Jarque-Bera, is non-significant (p-value = 0.084). Looking at the right side of figure 11, one sees that there are some deviations from the normal distribution, but the overall trend looks acceptable. Therefore, the results for zoom rate are considered robust.

There are five significant predictors for zoom rate: average popup number per page ($\beta = 0.043$, p-value = 0.003), average elements per page ($\beta = -0.053$, p-value = 0.006), use of multiple different background maps ($\beta = -0.105$, p-value = 0.001), usage of the geographic point feature ($\beta = 0.173$, p-value = 0.049), and the theme inclusion and cohesion ($\beta = 0.108$, p-value = 0.021). There are three positive predictor feature coefficients and two negative coefficients.

The geographic point feature has the largest positive coefficient value, but also the largest standard error (0.088) and the widest 95% confidence interval ([-0.346, -0.000]). Therefore, the interpretation that zoom rate would increase by 0.173 (17.3 percentage points) when a geo-point is used would probably be too exacting. However, this seems to give some justification as to why survey creators prefer the point feature: on average, respondents use higher than default zoom levels when placing point features, which can signify higher spatial data quality.

The survey theme inclusion and cohesion has a relatively strong positive effect ($\beta = 0.108$), as well. Instinctively, there is no clear connection between a survey

Table 6: Final multiple OLS regression summary, diagnostics, and coefficients for the log-transformed **zoom rate**. The coefficients (β_p) and the confidence intervals in this table are mirrored back to their realistic interpretation.

Model summary			
No. observations	216	Log-Likelihood	102.48
Df residuals	201	Df model	14
R^2	0.328	F-statistic	6.681
Adj. R^2	0.281	F-statistic p-value	4.33e-11*
AIC	-175.0	BIC	-124.3

Model diagnostics			
Omnibus	6.892	Durbin-Watson	2.147
Omnibus p-value	0.032*	Jarque-Bera (JB)	4.957
Kurtosis	2.438	JB p-value	0.084

Features	**Coefficient (β_p)**	**std. err.**	**p-value**	**[95% conf. int.]**
const	0.454	0.096	0.000*	[0.265, 0.642]
median_minutes	-0.025	0.016	0.114	[-0.056, 0.006]
avg_popup_page	0.043	0.014	0.003*	[0.014, 0.071]
avg_elems_page	-0.053	0.019	0.006*	[-0.090, -0.015]
kB_rel	0.031	0.029	0.284	[-0.026, 0.088]
multiple_maps	-0.105	0.032	0.001*	[-0.168, -0.042]
customization	-0.041	0.026	0.115	[-0.093, 0.010]
progression	-0.007	0.030	0.811	[-0.066, 0.052]
geo_point	0.173	0.088	0.049*	[0.000, 0.346]
geo_line	0.026	0.027	0.333	[-0.027, 0.079]
geo_area	-0.035	0.040	0.374	[-0.113, 0.043]
i_c	0.108	0.047	0.021*	[0.016, 0.200]
m_t	0.060	0.041	0.141	[-0.020, 0.139]
p_r	-0.043	0.053	0.420	[-0.146, 0.061]
up_d	0.076	0.044	0.084	[-0.010, 0.163]

Note: Heteroskedasticity robust standard errors are used.
* $p < 0.05$

theme and zooming behavior, but a suggestion of the connection can be crafted. When examining the surveys with this theme, one gets a feeling that many such surveys are about a small geographical area (e.g., a neighborhood) that the respondents likely know well. Giving zoomed-in answers in an environment one knows by heart would make sense.

Smallest of the positive coefficients, the effect of average popups per page gives an interesting insight about zooming behavior. With a popup, a mapping question effectively wants the respondent to confirm the feature placement asking the respondent

whether or not the pinned location is the intended location for the marker. This is not the case with a default mapping question. This additional confirmation step, with the potential added information about the marker placement, might encourage respondents to consider their feature placements more carefully.

The stronger of the two negative coefficients is the use of multiple different background maps ($\beta = -0.105$). The standard error is relatively small when compared to the geographic point feature. One interpretation for the negative effect of using multiple different background maps could be related to cognitive load: it is demanding to read one map correctly, not to mention two. Mental effort needs to be put into correctly identifying same elements between different background maps, and this effort might lower respondents' willingness to focus on zooming in into the maps and placing accurate responses.

Average elements per survey page also has a negative coefficient ($\beta = -0.053$), but clearly smaller than for the use of multiple background maps. It might suggest that a busy survey page with many elements on it can create a feeling of confusion and a lack of care, and thus zooming, for the respondent in answering a mapping question. For example, if a respondent is asked multiple mapping questions on one page, one could hypothesize that the feeling of importance per one response is low.

When comparing the results between submission rate and zoom rate, a few things should be highlighted. Firstly, the clearest difference between the models is the explanatory power. The adjusted coefficient of determination (adj. R^2) for the submission rate was 11.3%, whereas for zoom rate it was 28.1%. The predictor features are, thus, clearly more insightful when estimating the zoom rate than submission rate. This is not surprising considering that zooming is a mainly technical feature, whereas submission rate is a feature that is impacted by a multitude of non-technical influences difficult to measure through survey design. For example, the larger context in which the survey is done might not necessarily affect the zooming level of one question, but it might affect the willingness of a respondent to submit the whole survey. If a respondent has a feeling that the survey results won't be used in a concrete way after the surveying, answering the survey in full might feel irrelevant.

Secondly, there are two predictor features that are significant for both response features: average popups per page, and the theme inclusion and cohesion. For submission rate, these features have a negative impact, but a positive one for zoom rate. A clear-sighted interpretation about this connection is difficult, but one reasonable action could be to reflect back on the explanatory power of the models: when the whole model can only explain about a tenth of the variance in submission rate, one shouldn't get too stuck up in optimizing the survey design based on those results. A reasonable option would be to instead choose the clearer positive impact of those features on zoom rate and boost one's spatial data quality.

Thirdly, there are eight predictor features having no significant effect on either response feature: kB_rel, customization, progression, geo_line and geo_area, and three themes; mobility and transport, parks and recreation, and urban planning and design. These predictor features could be excluded from the regression model based on these results to end up with a more powerful model. It was a conscious decision, however, to leave these features in the analysis, rooted in the thinking presented in

section 3.3.2. Excluding all features having no effect would misrepresent the true effect of survey design on quality: there are design features that do not have a measurable effect on survey quality. Trying to inflate statistical effects could give a false sense of significance when considering design features and survey quality.

Moreover, the inclusion of even more features to the model (with even larger datasets) in future studies may also change the statistical contribution (i.e. the strength of the coefficients) of each individual feature, even of these current non-significant features.

5 Conclusions and discussion

The overarching purpose of the study was to examine the effect of Maptionnaire survey design on survey quality, operationalized as survey submission rate and spatial data quality. Supporting this aim, the Maptionnaire platform and database were inspected and queried to engineer a feature set that would represent survey design. Additionally, testing for potential spatial data quality proxies was done before selecting zoom rate as the main proxy.

In the literature review, the foundations of PPGIS as a method for public participation were first laid out before looking into the challenges of data quality in PPGIS. Tools like Maptionnaire are built on top of epistemological and technological advancements in the fields of participatory planning and mapping, encouraging and enabling planners to engage the public in democratic decision making. Involving the public in planning processes through mapping creates challenges in determining high quality data inputs: standardized quality metrics do not always catch the fuzzy nature of subjective PPGIS surveys.

A research gap was also recognized in the literature review. Previous studies in the field of PPGIS have not assessed the survey design's impact on data quality. Especially quantitative data analysis is lacking and, thus, this study aimed to fill that gap in the field.

After the literature review, the study's methodological framework was introduced. An established framework by Fagerholm et al. [7] was used in order to address the methodological inconsistency within PPGIS research. The framework has three distinct phases: explore, explain, and predict. Used data was also described along with the methods and specific decisions made within this overarching framework. The main method used to understand the impact of survey design on survey quality was multiple regression.

Finally, the study results were presented. The results were split into two: the results for the explore phase and the explain phase. The explore phase included 37 surveys that were studied in detail. In addition, experimentation for spatial data quality proxies was conducted and presented in that part. The explore phase was conducted to answer the two research sub-questions: which Maptionnaire features constitute survey design, and how can spatial data quality be operationalized for Maptionnaire surveys? Survey design features were discovered through visual inspection and feature engineering, while the spatial data quality proxy's operationalization was a result of a separate experimentation process.

The explain phase included 216 surveys. In the final regression model, 14 features were used to explain survey submission rate and zoom rate (the proxy for spatial data quality), the two features operationalized as survey quality. Findings from the regression analysis were used to answer the main research question, **what effect does the design of a Maptionnaire survey have on survey submission rate and spatial data quality**?

Based on the results, the impact of Maptionnaire survey design features on survey quality is relatively weak. Significant statistical effects were observed, but the explanatory power of the chosen design features on submission rate and zoom

rate were small. Only 11.3% of the submission rate's variance was explained by the variance of the 14 design features. For zoom rate, 28.1% of its variance could be explained by the combined variance of the design features.

Drawing conclusions about the survey design's effect on submission rate is especially difficult. The average rate to submit a survey in the final dataset was only 33.6% (figure G6), indicating a low submission rate for PPGIS surveys overall. Combined with low participation rates for PPGIS surveys in general [19], this suggests that the main issue for submission rate is not survey design. Other aspects in a PPGIS process, such as respondent recruitment and concrete actualization of the survey results [11], [21], can have a more significant impact on submission rate. Subsequently, a sensible course of action when designing PPGIS surveys would be to optimize survey design for spatial data quality (in this case zoom rate), instead of submission rate.

Even though more tangible, clear and straightforward conclusions about design's impact on zoom rate are not easy to draw either. It has been shown that respondent characteristics, such as age, impact zooming behavior [23]. Consequently, it is tough to isolate the effects of survey design features from other characteristics and attributes outside of survey design, that can strongly affect zoom rate. Additionally, the assumption made in this research that spatial data quality has a strong and positive association with zoom rate, can be misleading. It is not always obvious that a zoomed-in response is more accurate than an answer with a lower level of zoom. This is why the spatial data quality proxy was experimented with in this study, and why more research would be welcomed in operationalizing spatial data quality in PPGIS.

In any case, these results can provide a basis for experimentation and guidelines on how to design a Maptionnaire survey effectively. No positive effects were observed for submission rate, but to limit the respondent dropout rate, a survey creator should pay attention to the length of the survey and to the number of popup windows per page. Chronologically long surveys that have multiple popup windows per page seem to result in lower submission rates. For zoom rate, the use of multiple different background maps is discouraged, due to a potentially elevated cognitive load of interpreting multiple maps. The use of popup windows has a positive effect on respondents zooming in when placing map markers. That can be because of the confirmatory effect of a popup window; the popup window forces the respondent to double check their feature placement.

Based on the frequency of used elements (figure G3), survey creators are cautious in using geographic elements other than the point element. When submission rate is concerned, there's no reason for such cautiousness. However, respondents use higher than default zoom levels when placing point features, but not when placing line or area features. This gives some justification to the popularity of the simple point feature.

The results also suggested that a survey's theme has an impact on submission and zoom rate. The theme "inclusion and cohesion" had a significant negative effect on submission rate, but a positive effect on zoom rate. The effect of a survey theme is challenging to interpret, like was cautioned in section 4.1.1, because categorizing a survey into a theme or another is often fraught with subjectivity.

However, a hypothesis for the inclusion and cohesion theme's negative effect on submission rate could be that those surveys are often slightly esoteric in nature; there

might be no evident outcome that the survey seeks to achieve. For instance, a survey themed "mobility and transport" might be about a concrete infrastructure project that has taken or is going to take place, while an "inclusion and cohesion" survey could aim for abstract goals of trust, solidarity, or democracy. Survey fatigue (see e.g. [98]) can affect respondents' willingness to submit surveys if they have no obvious real-life impacts.

The positive impact of the theme inclusion and cohesion on zoom rate was hypothesized in the previous chapter: if a respondent knows the geographical area in question well, it might be easier to pinpoint zoomed-in and accurate map features.

A survey's overlay file size, aesthetic customization, or showing a respondent the progress they've made do not seem to affect submission nor zoom rate in any significant way.

This study was a novel exploration into possible linear effects of design features on survey quality. Addressing the gap in current PPGIS research, these purely quantitative results are best to be taken as the first steps into studying the connection between survey design and quality. Like written above, no effect of the wider PPGIS process was taken into consideration to control the study's scope. A negative side effect of this restriction is that it hampers the overall generalizability of the results.

Thematically speaking, the field of application for PPGIS is extremely diverse. PPGIS surveys can be utilized, among others, to examine how children's living environments impact physical activity and transportation modes [99], to guide conservation management [100], to understand the benefits of urban parks [101], or to assist transportation planning [102]. PPGIS is not a unanimously defined set of methods and approaches where universal rules about the quality and effectiveness of the tools used in the field apply.

Therefore, presenting generalized results about impactful PPGIS survey features must be done with care. Precaution should be taken in interpretation and application of results that attempt to summarize an often idiosyncratic process into a few measurable metrics.

This research was conducted in collaboration with Mapita Oy and all the data used in it are from their database. Other PPGIS platforms have different features and applications, so one should be mindful in translating the results of this study to other platforms. Additionally, filters were used to restrict the studied surveys; only surveys created by consultation companies in Finland and the U.S. were analyzed. Maptionnaire is also used by universities, cities, non-governmental organizations, and so forth, which might use surveys in a different way and for different purposes than consultation companies.

To address the described study restrictions, more research would be needed to combine the obtained results into the effects of respondent recruitment and the concrete application of survey results in decision making. Increasing the amount of data could also increase the explanatory power of this study. Even more design features could be added into the regression model in order to predict survey quality, possibly revealing new points of interest for survey designers.

In addition, different study designs would be useful to understand the effects of survey design features on survey quality. For instance, one could gain more concrete

answers about the effects of individual features through a comparative experimental study that creates multiple, otherwise similar surveys and systematically varies the design features. In such study, the respondent recruitment should also be controlled to solely focus on the design features.

References

[1] M. K. McCall and C. E. Dunn, "Geo-information tools for participatory spatial planning: Fulfilling the criteria for 'good' governance?", *Geoforum*, vol. 43, no. 1, pp. 81–94, Jan. 2012, ISSN: 00167185. DOI: 10.1016/j.geoforum.2011.07.007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0016718511001369 (visited on 05/13/2025).

[2] J. Sebunya and A. Gichuki, "The Impact of Participatory Planning on Sustainable Development: A Literature Review", *Journal of Strategic Management*, vol. 4, no. 4, pp. 1–9, Nov. 25, 2024, ISSN: 2789-4851. DOI: 10.70619/vol4iss4pp1-9. [Online]. Available: https://edinburgjournals.org/journals/index.php/journal-of-strategic-management/article/view/404 (visited on 08/13/2025).

[3] UN. "Goal 11 | Department of Economic and Social Affairs". (), [Online]. Available: https://sdgs.un.org/goals/goal11#targets_and_indicators (visited on 07/03/2025).

[4] D. Maguire, "An overview and definition of GIS", *Geographical information systems. Vol. 1: principles*, Maquire D.J., Ed., pp. 9–20, 1991. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0026304549&partnerID=40&md5=df756d650ea411e438005103102f608a.

[5] T. Denwood, J. J. Huck, and S. Lindley, "Effective PPGIS in spatial decision-making: Reflecting participant priorities by illustrating the implications of their choices", *Transactions in GIS*, vol. 26, no. 2, pp. 867–886, Apr. 2022, ISSN: 1361-1682, 1467-9671. DOI: 10.1111/tgis.12888. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/tgis.12888 (visited on 07/03/2025).

[6] G. Brown, "Public Participation GIS (PPGIS) for Regional and Environmental Planning: Reflections on a Decade of Empirical Research", *URISA Journal*, vol. 25, no. 2, 2012.

[7] N. Fagerholm, C. M. Raymond, A. S. Olafsson, *et al.*, "A methodological framework for analysis of participatory mapping data in research, planning, and management", *International Journal of Geographical Information Science*, vol. 35, no. 9, pp. 1848–1875, Sep. 2, 2021, ISSN: 1365-8816, 1362-3087. DOI: 10.1080/13658816.2020.1869747. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/13658816.2020.1869747 (visited on 04/10/2025).

[8] M. Kahila-Tani, A. Broberg, M. Kyttä, and T. Tyger, "Let the Citizens Map—Public Participation GIS as a Planning Support System in the Helsinki Master Plan Process", *Planning Practice & Research*, vol. 31, no. 2, pp. 195–214, Mar. 14, 2016, ISSN: 0269-7459, 1360-0583. DOI: 10.1080/02697459.2015.1104203. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/02697459.2015.1104203 (visited on 08/17/2025).

[9] P. Parviainen, M. Tihinen, J. Kääriäinen, and S. Teppola, "Tackling the digitalization challenge: How to benefit from digitalization in practice", *International Journal of Information Systems and Project Management*, vol. 5, no. 1, pp. 63–77, Feb. 1, 2022, ISSN: 2182-7788. DOI: 10.12821/ijispm050104. [Online]. Available: https://revistas.uminho.pt/index.php/ijispm/article/view/3856 (visited on 08/17/2025).

[10] S. Gottwald, T. E. Laatikainen, and M. Kyttä, "Exploring the usability of PPGIS among older adults: Challenges and opportunities", *International Journal of Geographical Information Science*, vol. 30, no. 12, pp. 2321–2338, Dec. 1, 2016, ISSN: 1365-8816, 1362-3087. DOI: 10.1080/13658816.2016.1170837. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/13658816.2016.1170837 (visited on 09/05/2025).

[11] M. Kahila-Tani, M. Kytta, and S. Geertman, "Does mapping improve public participation? Exploring the pros and cons of using public participation GIS in urban planning practices", *Landscape and Urban Planning*, vol. 186, pp. 45–55, Jun. 2019, ISSN: 01692046. DOI: 10.1016/j.landurbplan.2019.02.019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169204618301932 (visited on 08/13/2025).

[12] C. M. Burnett, Ed., *Evaluating Participatory Mapping Software*. Cham: Springer International Publishing, 2023, ISBN: 978-3-031-19593-8 978-3-031-19594-5. DOI: 10.1007/978-3-031-19594-5. [Online]. Available: https://link.springer.com/10.1007/978-3-031-19594-5 (visited on 07/09/2025).

[13] M. Kyttä, N. Fagerholm, V. H. Hausner, and A. Broberg, "Maptionnaire", in *Evaluating Participatory Mapping Software*, C. M. Burnett, Ed., Cham: Springer International Publishing, 2023, pp. 71–91, ISBN: 978-3-031-19593-8 978-3-031-19594-5. DOI: 10.1007/978-3-031-19594-5_4. [Online]. Available: https://link.springer.com/10.1007/978-3-031-19594-5_4 (visited on 09/02/2025).

[14] S. E. Spielman, "Spatial collective intelligence? Credibility, accuracy, and volunteered geographic information", *Cartography and Geographic Information Science*, vol. 41, no. 2, pp. 115–124, Mar. 15, 2014, ISSN: 1523-0406, 1545-0465. DOI: 10.1080/15230406.2013.874200. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/15230406.2013.874200 (visited on 09/02/2025).

[15] G. Brown and N. Fagerholm, "Empirical PPGIS/PGIS mapping of ecosystem services: A review and evaluation", *Ecosystem Services*, vol. 13, pp. 119–133, Jun. 2015, ISSN: 22120416. DOI: 10.1016/j.ecoser.2014.10.007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2212041614001235 (visited on 09/11/2025).

[16] R. Devillers, Y. Bédard, R. Jeansoulin, and B. Moulin, "Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data", *International Journal of Geographical Information Science*, vol. 21, no. 3, pp. 261–282, Mar. 2007, ISSN: 1365-8816, 1362-3087. DOI: 10.1080/13658810600911879. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/13658810600911879 (visited on 09/08/2025).

[17] G. Brown, "A Review of Sampling Effects and Response Bias in Internet Participatory Mapping (PPGIS/PGIS/VGI)", *Transactions in GIS*, vol. 21, no. 1, pp. 39–56, Feb. 2017, ISSN: 1361-1682, 1467-9671. DOI: 10.1111/tgis.12207. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/tgis.12207 (visited on 09/05/2025).

[18] K. M. Hansen, "The Effects of Incentives, Interview Length, and Interviewer Characteristics on Response Rates in a CATI-Study", *International Journal of Public Opinion Research*, vol. 19, no. 1, pp. 112–121, Mar. 13, 2006, ISSN: 0954-2892, 1471-6909. DOI: 10.1093/ijpor/edl022. [Online]. Available: https://academic.oup.com/ijpor/article-lookup/doi/10.1093/ijpor/edl022 (visited on 09/05/2025).

[19] E. A. Salminen, V. H. Hausner, F. J. Ancin Murguzur, and S. Engen, "Optimizing recruitment in an online environmental PPGIS—is it worth the time and costs?", *International Journal of Geographical Information Science*, vol. 39, no. 3, pp. 652–674, Mar. 4, 2025, ISSN: 1365-8816, 1362-3087. DOI: 10.1080/13658816.2024.2427267. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/13658816.2024.2427267 (visited on 04/11/2025).

[20] G. Brown and M. Kyttä, "Key issues and research priorities for public participation GIS (PPGIS): A synthesis based on empirical research", *Applied Geography*, vol. 46, pp. 122–136, Jan. 2014, ISSN: 01436228. DOI: 10.1016/j.apgeog.2013.11.004. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0143622813002531 (visited on 07/03/2025).

[21] S. Kantola, N. Fagerholm, and A. Nikula, "Utilization and implementation of PPGIS in land use planning and decision-making from the perspective of organizations", *Land Use Policy*, vol. 127, p. 106 528, Apr. 2023, ISSN: 02648377. DOI: 10.1016/j.landusepol.2022.106528. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0264837722005555 (visited on 05/09/2025).

[22] V. Nurminen, S. Rossi, T. Rinne, and M. Kyttä, "How has digital participatory mapping influenced urban planning: Views from nine planning cases from Finland", *Computers, Environment and Urban Systems*, vol. 112, p. 102 152, Sep. 2024, ISSN: 01989715. DOI: 10.1016/j.compenvurbsys.2024.102152. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0198971524000814 (visited on 08/13/2025).

[23] M. Rzeszewski and J. Kotus, "Usability and usefulness of internet mapping platforms in participatory spatial planning", *Applied Geography*, vol. 103, pp. 56–69, Feb. 2019, ISSN: 01436228. DOI: 10.1016/j.apgeog.2019.01.001. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0143622818305460 (visited on 08/13/2025).

[24] M. J. Grant and A. Booth, "A typology of reviews: An analysis of 14 review types and associated methodologies", *Health Information & Libraries Journal*, vol. 26, no. 2, pp. 91–108, Jun. 2009, ISSN: 1471-1834, 1471-1842. DOI: 10.1111/j.1471-1842.2009.00848.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.1471-1842.2009.00848.x (visited on 09/01/2025).

[25] F. R. Jensenius, M. Htun, D. J. Samuels, D. A. Singer, A. Lawrence, and M. Chwe, "The Benefits and Pitfalls of Google Scholar", *PS: Political Science & Politics*, vol. 51, no. 4, pp. 820–824, Oct. 2018, ISSN: 1049-0965, 1537-5935. DOI: 10.1017/S104909651800094X. [Online]. Available: https://www.cambridge.org/core/product/identifier/S104909651800094X/type/journal_article (visited on 09/01/2025).

[26] K. Jordan and S. P. Tsai, "Keywords, citations and 'algorithm magic': Exploring assumptions about ranking in academic literature searches online", *Learning, Media and Technology*, pp. 1–15, Aug. 23, 2024, ISSN: 1743-9884, 1743-9892. DOI: 10.1080/17439884.2024.2392108. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/17439884.2024.2392108 (visited on 09/01/2025).

[27] R. Pranckutė, "Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World", *Publications*, vol. 9, no. 1, p. 12, Mar. 12, 2021, ISSN: 2304-6775. DOI: 10.3390/publications9010012. [Online]. Available: https://www.mdpi.com/2304-6775/9/1/12 (visited on 09/01/2025).

[28] M. Tober, "PubMed, ScienceDirect, Scopus or Google Scholar – Which is the best search engine for an effective literature research in laser medicine?", *Medical Laser Application*, vol. 26, no. 3, pp. 139–144, Aug. 2011, ISSN: 16151615. DOI: 10.1016/j.mla.2011.05.006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1615161511000329 (visited on 09/01/2025).

[29] S. Jalali and C. Wohlin, "Systematic literature studies: Database searches vs. backward snowballing", in *Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, Lund Sweden: ACM, Sep. 19, 2012, pp. 29–38, ISBN: 978-1-4503-1056-7. DOI: 10.1145/2372251.2372257. [Online]. Available: https://dl.acm.org/doi/10.1145/2372251.2372257 (visited on 10/14/2025).

[30] E. H. Zube, “Perceived land use patterns and landscape values”, *Landscape Ecology*, vol. 1, no. 1, pp. 37–45, Jul. 1987, ISSN: 0921-2973, 1572-9761. DOI: 10.1007/BF02275264. [Online]. Available: http://link.springer.com/10.1007/BF02275264 (visited on 08/13/2025).

[31] J. J. Gibson, *The Ecological Approach to Visual Perception: Classic Edition*, 1st ed. Psychology Press, Nov. 20, 2014, ISBN: 978-1-315-74021-8. DOI: 10.4324/9781315740218. [Online]. Available: https://www.taylorfrancis.com/books/9781315740218 (visited on 08/13/2025).

[32] K. H. Basso, “"Speaking with Names": Language and Landscape among the Western Apache”, *Cultural Anthropology*, vol. 3, no. 2, pp. 99–130, 1988, ISSN: 08867356, 15481360. JSTOR: 656347. [Online]. Available: http://www.jstor.org/stable/656347 (visited on 08/13/2025).

[33] J. W. Crampton, “Maps as social constructions: Power, communication and visualization”, *Progress in Human Geography*, vol. 25, no. 2, pp. 235–252, Jun. 2001, ISSN: 0309-1325, 1477-0288. DOI: 10.1191/030913201678580494. [Online]. Available: https://journals.sagepub.com/doi/10.1191/030913201678580494 (visited on 08/13/2025).

[34] N. L. Peluso, “WHOSE WOODS ARE THESE? COUNTER-MAPPING FOREST TERRITORIES IN KALIMANTAN, INDONESIA”, *Antipode*, vol. 27, no. 4, pp. 383–406, Oct. 1995, ISSN: 0066-4812, 1467-8330. DOI: 10.1111/j.1467-8330.1995.tb00286.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-8330.1995.tb00286.x (visited on 08/13/2025).

[35] S. R. Arnstein, “A Ladder Of Citizen Participation”, *Journal of the American Institute of Planners*, vol. 35, no. 4, pp. 216–224, Jul. 1969, ISSN: 0002-8991. DOI: 10.1080/01944366908977225. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01944366908977225 (visited on 09/19/2025).

[36] L. Laurian, “Public Participation in Environmental Decision Making: Findings from Communities Facing Toxic Waste Cleanup”, *Journal of the American Planning Association*, vol. 70, no. 1, pp. 53–65, Mar. 31, 2004, ISSN: 0194-4363, 1939-0130. DOI: 10.1080/01944360408976338. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01944360408976338 (visited on 08/17/2025).

[37] P. Jankowski, M. Czepkiewicz, M. Młodkowski, and Z. Zwoliński, “Geo-questionnaire: A Method and Tool for Public Preference Elicitation in Land Use Planning”, *Transactions in GIS*, vol. 20, no. 6, pp. 903–924, Dec. 2016, ISSN: 1361-1682, 1467-9671. DOI: 10.1111/tgis.12191. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/tgis.12191 (visited on 08/17/2025).

[38] L. Cochrane and J. Corbett, "Participatory Mapping", in *Handbook of Communication for Development and Social Change*, J. Servaes, Ed., Singapore: Springer Singapore, 2018, pp. 1–9, ISBN: 978-981-10-7035-8. DOI: 10.1007/978-981-10-7035-8_6-1. [Online]. Available: http://link.springer.com/10.1007/978-981-10-7035-8_6-1 (visited on 09/07/2025).

[39] R. Chambers, "Participatory Mapping and Geographic Information Systems: Whose Map? Who is Empowered and Who Disempowered? Who Gains and Who Loses?", *THE ELECTRONIC JOURNAL OF INFORMATION SYSTEMS IN DEVELOPING COUNTRIES*, vol. 25, no. 1, pp. 1–11, Jun. 2006, ISSN: 1681-4835, 1681-4835. DOI: 10.1002/j.1681-4835.2006.tb00163.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/j.1681-4835.2006.tb00163.x (visited on 09/08/2025).

[40] H. Rantanen and M. Kahila, "The SoftGIS approach to local knowledge", *Journal of Environmental Management*, vol. 90, no. 6, pp. 1981–1990, May 2009, ISSN: 03014797. DOI: 10.1016/j.jenvman.2007.08.025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0301479708001199 (visited on 09/08/2025).

[41] G. Rambaldi, P. A. K. Kyem, M. McCall, and D. Weiner, "Participatory Spatial Information Management and Communication in Developing Countries", *THE ELECTRONIC JOURNAL OF INFORMATION SYSTEMS IN DEVELOPING COUNTRIES*, vol. 25, no. 1, pp. 1–9, Jun. 2006, ISSN: 1681-4835, 1681-4835. DOI: 10.1002/j.1681-4835.2006.tb00162.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/j.1681-4835.2006.tb00162.x (visited on 09/08/2025).

[42] N. Fagerholm, M. García-Martín, M. Torralba, C. Bieling, and T. Plieninger, "Public participation geographical information systems (PPGIS): Participatory research methods for sustainability - toolkit #1", *GAIA - Ecological Perspectives for Science and Society*, vol. 31, no. 1, pp. 46–48, Mar. 31, 2022, ISSN: 0940-5550. DOI: 10.14512/gaia.31.1.10. [Online]. Available: https://www.ingentaconnect.com/content/10.14512/gaia.31.1.10 (visited on 07/03/2025).

[43] A. Zolkafli, Y. Liu, and G. Brown, "Bridging the knowledge divide between public and experts using PGIS for land use planning in Malaysia", *Applied Geography*, vol. 83, pp. 107–117, Jun. 2017, ISSN: 01436228. DOI: 10.1016/j.apgeog.2017.03.013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0143622816304787 (visited on 09/08/2025).

[44] M. F. Goodchild, "Citizens as sensors: The world of volunteered geography", *GeoJournal*, vol. 69, no. 4, pp. 211–221, Nov. 30, 2007, ISSN: 0343-2521, 1572-9893. DOI: 10.1007/s10708-007-9111-y. [Online]. Available: http://link.springer.com/10.1007/s10708-007-9111-y (visited on 09/08/2025).

[45] "OpenStreetMap Foundation". (), [Online]. Available: https://osmfoundation.org/ (visited on 10/27/2025).

[46] J. Verplanke, M. K. McCall, C. Uberhuaga, G. Rambaldi, and M. Haklay, "A Shared Perspective for PGIS and VGI", *The Cartographic Journal*, vol. 53, no. 4, pp. 308–317, Oct. 2016, ISSN: 0008-7041, 1743-2774. DOI: 10.1080/00087041.2016.1227552. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00087041.2016.1227552 (visited on 09/08/2025).

[47] G. Brown, "An empirical evaluation of the spatial accuracy of public participation GIS (PPGIS) data", *Applied Geography*, vol. 34, pp. 289–294, May 1, 2012, ISSN: 0143-6228. DOI: 10.1016/j.apgeog.2011.12.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0143622811002402 (visited on 09/10/2025).

[48] A. D. Mekonnen and P. V. Gorsevski, "A web-based participatory GIS (PGIS) for offshore wind farm suitability within Lake Erie, Ohio", *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 162–177, Jan. 2015, ISSN: 13640321. DOI: 10.1016/j.rser.2014.08.030. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364032114007060 (visited on 09/08/2025).

[49] P. van Oort, "Spatial Data Quality: From Description to Application", Ph.D. dissertation, Wageningen University and Research, Netherlands, 2006, 141 pp., ISBN: 9798515235710. [Online]. Available: https://www.proquest.com/dissertations-theses/spatial-data-quality-description-application/docview/2564139910/se-2?accountid=98641.

[50] M. F. Goodchild, *Fundamentals of Spatial Data Quality*. John Wiley & Sons, 2010, vol. 662.

[51] P. F. Fisher, "Models of uncertainty in spatial data", *Geographical information systems*, vol. 1, pp. 191–205, 1999.

[52] V. Antoniou and A. Skopeliti, "MEASURES AND INDICATORS OF VGI QUALITY: AN OVERVIEW", *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3/W5, pp. 345–351, Aug. 20, 2015, ISSN: 2194-9050. DOI: 10.5194/isprsannals-II-3-W5-345-2015. [Online]. Available: https://isprs-annals.copernicus.org/articles/II-3-W5/345/2015/ (visited on 07/01/2025).

[53] F. J. Ariza-López, P. Barreira González, J. Masó Pau, *et al.*, "Geospatial data quality (ISO 19157-1): Evolve or perish", *Revista Cartográfica*, no. 100, pp. 129–154, May 1, 2020, ISSN: 2663-3981, 0080-2085. DOI: 10.35424/rcarto.i100.692. [Online]. Available: https://revistasipgh.org/index.php/rcar/article/view/692 (visited on 10/27/2025).

[54] G. M. Johnson, "The hard proxy problem: Proxies aren't intentional; they're intentional", *Philosophical Studies*, vol. 182, no. 5–6, pp. 1383–1411, Jun. 2025, ISSN: 0031-8116, 1573-0883. DOI: 10.1007/s11098-025-02333-9. [Online]. Available: https://link.springer.com/10.1007/s11098-025-02333-9 (visited on 09/16/2025).

[55] G. Brown, D. Weber, D. Zanon, and K. De Bie, "Evaluation of an online (opt-in) panel for public participation geographic information systems surveys", *International Journal of Public Opinion Research*, vol. 24, no. 4, pp. 534–545, Dec. 1, 2012, ISSN: 0954-2892, 1471-6909. DOI: 10.1093/ijpor/eds001. [Online]. Available: https://academic.oup.com/ijpor/article-lookup/doi/10.1093/ijpor/eds001 (visited on 09/11/2025).

[56] K. Hasanzadeh, A. Kajosaari, D. Häggman, and M. Kyttä, "A context sensitive approach to anonymizing public participation GIS data: From development to the assessment of anonymization effects on data quality", *Computers, Environment and Urban Systems*, vol. 83, p. 101 513, Sep. 2020, ISSN: 01989715. DOI: 10.1016/j.compenvurbsys.2020.101513. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0198971520302465 (visited on 09/16/2025).

[57] P. Jankowski, K. Forss, M. Czepkiewicz, H. Saarikoski, and M. Kahila, "Assessing impacts of PPGIS on urban land use planning: Evidence from Finland and Poland", *European Planning Studies*, vol. 30, no. 8, pp. 1529–1548, Aug. 3, 2022, ISSN: 0965-4313, 1469-5944. DOI: 10.1080/09654313.2021.1882393. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/09654313.2021.1882393 (visited on 09/16/2025).

[58] H. R. Bernard, *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. Bloomsbury Publishing PLC, 2017.

[59] J. A. Krosnick, "Questionnaire Design", in *The Palgrave Handbook of Survey Research*, D. L. Vannette and J. A. Krosnick, Eds., Cham: Springer International Publishing, 2018, pp. 439–455, ISBN: 978-3-319-54394-9 978-3-319-54395-6. DOI: 10.1007/978-3-319-54395-6_53. [Online]. Available: http://link.springer.com/10.1007/978-3-319-54395-6_53 (visited on 09/12/2025).

[60] P. Lietz, "Research into Questionnaire Design: A Summary of the Literature", *International Journal of Market Research*, vol. 52, no. 2, pp. 249–272, Mar. 2010, ISSN: 1470-7853, 2515-2173. DOI: 10.2501/S147078530920120X. [Online]. Available: https://journals.sagepub.com/doi/10.2501/S147078530920120X (visited on 09/05/2025).

[61] W. Tan, *Research Methods: A Practical Guide for Students and Researchers*. World Scientific, 2022.

[62] M. J. Goertzen, "Introduction to quantitative research and data", *Library technology reports*, vol. 53, no. 4, pp. 12–18, 2017. [Online]. Available: https://journals.ala.org/index.php/ltr/article/view/6325/8275 (visited on 09/23/2025).

[63] A. Páez and G. Boisjoly, "Exploratory Data Analysis", in *Discrete Choice Analysis with R*. Cham: Springer International Publishing, 2022, pp. 25–64, ISBN: 978-3-031-20718-1 978-3-031-20719-8. DOI: 10.1007/978-3-031-20719-8_2. [Online]. Available: https://link.springer.com/10.1007/978-3-031-20719-8_2 (visited on 09/23/2025).

[64] M. Komorowski, D. C. Marshall, J. D. Salciccioli, and Y. Crutain, "Exploratory Data Analysis", in *Secondary Analysis of Electronic Health Records*. Cham: Springer International Publishing, 2016, pp. 185–203, ISBN: 978-3-319-43740-8 978-3-319-43742-2. DOI: 10.1007/978-3-319-43742-2_15. [Online]. Available: http://link.springer.com/10.1007/978-3-319-43742-2_15 (visited on 09/23/2025).

[65] H. M. Park, "Univariate analysis and normality test using SAS, Stata, and SPSS", 2015.

[66] IBM. "What is a feature engineering?" (), [Online]. Available: https://www.ibm.com/think/topics/feature-engineering (visited on 08/18/2025).

[67] S. Okada, M. Ohzeki, and S. Taguchi, "Efficient partition of integer optimization problems with one-hot encoding", *Scientific Reports*, vol. 9, no. 1, p. 13 036, Sep. 10, 2019, ISSN: 2045-2322. DOI: 10.1038/s41598-019-49539-6. [Online]. Available: https://www.nature.com/articles/s41598-019-49539-6 (visited on 10/28/2025).

[68] A. V. Palacio Buendía, M. Y. Pérez Albert, and D. Serrano Giné, "Online Public Participation Geographic Information System (PPGIS) as a landscape and public use management tool: A case study from the Ebro Delta Natural Park (Spain)", *Landscape Online*, vol. 93, pp. 1–18, Dec. 21, 2021, ISSN: 1865-1542. DOI: 10.3097/LO.202193. [Online]. Available: https://landscape-online.org/index.php/lo/article/view/LO.202193 (visited on 09/17/2025).

[69] R. Bachi, "STANDARD DISTANCE MEASURES AND RELATED METHODS FOR SPATIAL ANALYSIS", *Papers in Regional Science*, vol. 10, no. 1, pp. 83–132, Jan. 1962, ISSN: 10568190. DOI: 10.1111/j.1435-5597.1962.tb00872.x. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1056819023008205 (visited on 09/17/2025).

[70] Esri. "How Standard Distance works—ArcGIS Pro | Documentation". (), [Online]. Available: https://pro.arcgis.com/en/pro-app/3.3/tool-reference/spatial-statistics/h-how-standard-distance-spatial-statistic-works.htm (visited on 10/09/2025).

[71] P. Schober, C. Boer, and L. A. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation", *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763–1768, May 2018, ISSN: 0003-2999. DOI: 10.1213/ANE.0000000000002864. [Online]. Available: https://journals.lww.com/00000539-201805000-00050 (visited on 08/29/2025).

[72] J. J. Berman, "Understanding Your Data", in *Data Simplification*, Elsevier, 2016, pp. 135–187, ISBN: 978-0-12-803781-2. DOI: 10.1016/B978-0-12-803781-2.00004-7. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780128037812000047 (visited on 09/30/2025).

[73] C. P. Dancey and J. Reidy, *Statistics Without Maths for Psychology*, 8th ed. Harlow, United Kingdom: Pearson Education Canada, 2020, 1 p., ISBN: 978-1-292-27644-1.

[74] J. H. Zar, "Significance Testing of the Spearman Rank Correlation Coefficient", *Journal of the American Statistical Association*, vol. 67, no. 339, pp. 578–580, Sep. 1972, ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1972.10481251. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01621459.1972.10481251 (visited on 10/14/2025).

[75] A. F. Hayes and L. Cai, "Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation", *Behavior research methods*, vol. 39, no. 4, pp. 709–722, 2007.

[76] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. John Wiley & Sons, 2021.

[77] A. L. Burton, "OLS (Linear) Regression", in *The Encyclopedia of Research Methods in Criminology and Criminal Justice*, J. Barnes and D. R. Forde, Eds., 1st ed., Wiley, Aug. 24, 2021, pp. 509–514, ISBN: 978-1-119-11072-9 978-1-119-11193-1. DOI: 10.1002/9781119111931.ch104. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/9781119111931.ch104 (visited on 09/29/2025).

[78] J. H. Kim, "Multicollinearity and misleading statistical results", *Korean Journal of Anesthesiology*, vol. 72, no. 6, pp. 558–569, Dec. 1, 2019, ISSN: 2005-6419, 2005-7563. DOI: 10.4097/kja.19087. [Online]. Available: http://ekja.org/journal/view.php?doi=10.4097/kja.19087 (visited on 08/26/2025).

[79] J. S. Long and L. H. Ervin, "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model", *The American Statistician*, vol. 54, no. 3, pp. 217–224, Aug. 1, 2000, ISSN: 0003-1305. DOI: 10.1080/00031305.2000.10474549. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00031305.2000.10474549 (visited on 10/17/2025).

[80] C. Robinson and R. E. Schumacker, "Interaction effects: Centering, variance inflation factor, and interpretation issues", *General Linear Model Journal*, vol. 35, no. 1, pp. 6–11, 2009.

[81] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* (Wiley Series in Probability and Statistics), 1st ed. Wiley, Jun. 24, 1980, ISBN: 978-0-471-05856-4 978-0-471-72515-2. DOI: 10.1002/0471725153. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/0471725153 (visited on 10/14/2025).

[82] F. Maleki, K. Ovens, R. Gupta, C. Reinhold, A. Spatz, and R. Forghani. "Generalizability of Machine Learning Models: Quantitative Evaluation of Three Methodological Pitfalls". arXiv: 2202.01337 [cs]. (Sep. 7, 2022), [Online]. Available: http://arxiv.org/abs/2202.01337 (visited on 10/20/2025), pre-published.

[83] D. N. Gujarati and D. C. Porter, *Basic Econometrics* (The McGraw-Hill Series Economics), 5. ed. Boston, Mass.: McGraw-Hill Irwin, 2009, 922 pp., ISBN: 978-0-07-337577-9.

[84] J. Shreffler and M. R. Huecker, "Hypothesis Testing, P Values, Confidence Intervals, and Significance", in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2023. PMID: 32491353. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK557421/ (visited on 10/14/2025).

[85] N. J. Nagelkerke, "A note on a general definition of the coefficient of determination", *biometrika*, vol. 78, no. 3, pp. 691–692, 1991.

[86] A. K. Srivastava, V. K. Srivastava, and A. Ullah, "The coefficient of determination and its adjusted version in linear regression models", *Econometric Reviews*, vol. 14, no. 2, pp. 229–240, Jan. 1995, ISSN: 0747-4938, 1532-4168. DOI: 10.1080/07474939508800317. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07474939508800317 (visited on 10/09/2025).

[87] E. A. Mohammed, C. Naugler, and B. H. Far, "Emerging business intelligence framework for a clinical laboratory through big data analytics", *Emerging trends in computational biology, bioinformatics, and systems biology*, vol. 2015, pp. 577–602, 2015.

[88] Z. Ji, T. Yu, Y. Xu, N. Lee, E. Ishii, and P. Fung, "Towards Mitigating LLM Hallucination via Self Reflection", in *Findings of the Association for Computational Linguistics: EMNLP 2023*, Singapore: Association for Computational Linguistics, 2023, pp. 1827–1843. DOI: 10.18653/v1/2023.findings-emnlp.123. [Online]. Available: https://aclanthology.org/2023.findings-emnlp.123 (visited on 09/17/2025).

[89] S. Pandey and C. L. Bright, "What Are Degrees of Freedom?", 2008.

[90] S. Innes, R. Granger, and J. Théroux, "Creating and testing a questionnaire to predict immediate and strong positive responders to spinal manipulative therapy for non-specific low back pain. A pilot study", *Chiropractic & Manual Therapies*, vol. 31, no. 1, p. 40, Sep. 26, 2023, ISSN: 2045-709X. DOI: 10.1186/s12998-023-00510-3. [Online]. Available: https:

//chiromt.biomedcentral.com/articles/10.1186/s12998-023-00510-3 (visited on 10/18/2025).

[91] S. E. Maxwell, "Sample size and multiple regression analysis.", *Psychological Methods*, vol. 5, no. 4, pp. 434–458, 2000, ISSN: 1939-1463, 1082-989X. DOI: 10.1037/1082-989X.5.4.434. [Online]. Available: https://doi.apa.org/doi/10.1037/1082-989X.5.4.434 (visited on 10/18/2025).

[92] T. B. Newman and W. S. Browner, "In Defense of Standardized Regression Coefficients:" *Epidemiology*, vol. 2, no. 5, pp. 383–386, Sep. 1991, ISSN: 1044-3983. DOI: 10.1097/00001648-199109000-00014. [Online]. Available: http://journals.lww.com/00001648-199109000-00014 (visited on 10/24/2025).

[93] R. J. Freund and W. J. Wilson, *Statistical Methods*. Elsevier, 2003.

[94] S. Greenland, M. Maclure, J. J. Schlesselman, C. Poole, and H. Morgenstern, "Standardized Regression Coefficients: A Further Critique and Review of Some Alternatives", *Epidemiology*, vol. 2, no. 5, p. 387, Sep. 1991, ISSN: 1044-3983. [Online]. Available: https://journals.lww.com/epidem/citation/1991/09000/Standardized_Regression_Coefficients__A_Further.15.aspx (visited on 10/20/2025).

[95] G. King, "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science", *American Journal of Political Science*, vol. 30, no. 3, p. 666, Aug. 1986, ISSN: 00925853. DOI: 10.2307/2111095. JSTOR: 2111095. [Online]. Available: https://www.jstor.org/stable/2111095?origin=crossref (visited on 10/20/2025).

[96] D. K. Lee, "Data transformation: A focus on the interpretation", *Korean Journal of Anesthesiology*, vol. 73, no. 6, pp. 503–508, Dec. 2020, ISSN: 2005-6419. DOI: 10.4097/kja.20137. PMID: 33271009. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7714623/ (visited on 10/20/2025).

[97] K. Benoit, "Linear Regression Models with Logarithmic Transformations", 2011.

[98] S. R. Porter, M. E. Whitcomb, and W. H. Weitzer, "Multiple surveys of students and survey fatigue", *New Directions for Institutional Research*, vol. 2004, no. 121, pp. 63–73, Jan. 2004, ISSN: 0271-0579, 1536-075X. DOI: 10.1002/ir.101. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/ir.101 (visited on 10/29/2025).

[99] A. Broberg, S. Salminen, and M. Kyttä, "Physical environmental characteristics promoting independent and active transport to children's meaningful places", *Applied Geography*, vol. 38, pp. 43–52, Mar. 2013, ISSN: 01436228. DOI: 10.1016/j.apgeog.2012.11.014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0143622812001579 (visited on 10/27/2025).

[100] G. Brown and D. Weber, "A place-based approach to conservation management using public participation GIS (PPGIS)", *Journal of Environmental Planning and Management*, vol. 56, no. 4, pp. 455–473, May 2013, ISSN: 0964-0568, 1360-0559. DOI: 10.1080/09640568.2012.685628. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/09640568.2012.685628 (visited on 08/15/2025).

[101] G. Brown, M. F. Schebella, and D. Weber, "Using participatory GIS to measure physical activity and urban park benefits", *Landscape and Urban Planning*, vol. 121, pp. 34–44, Jan. 2014, ISSN: 01692046. DOI: 10.1016/j.landurbplan.2013.09.006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169204613001850 (visited on 08/15/2025).

[102] B. Waleghwa and T. Heldt, "Exploring the use of public participation GIS in transportation planning for tourism at a Nordic destination", *Scandinavian Journal of Hospitality and Tourism*, vol. 22, no. 3, pp. 210–234, May 27, 2022, ISSN: 1502-2250, 1502-2269. DOI: 10.1080/15022250.2022.2070541. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/15022250.2022.2070541 (visited on 08/15/2025).

[103] D. Gujarati, "Use of dummy variables in testing for equality between sets of coefficients in linear regressions: A generalization", *The American Statistician*, vol. 24, no. 5, pp. 18–22, 1970.

A Scoping review keywords

Keywords from the scoping review used in formulating the search string for the literature review proper.

Table A1: Google Scholar keywords obtained with the scoping review.

Google Scholar keywords	
Active travel	Participant recruitment
Anonymization	Participatory design
Citizen design science	Participatory mapping
Citizen knowledge	Place values
Citizen science	PPGIS
Conservation planning	Public lands
Crowd-sourcing	Public participation
Crowdsourced	Public participation GIS
Data quality	Public participatory geo-information systems
Decision-making	Social fitness network
Digital divide	Social metrics
Environmental management	Spatial
Eye tracking	Spatial accuracy
FOSS	Spatial analysis
GDPR	Spatial data quality
Geographic information systems	Spatial dispersion index
Geoparticipation	Spatial planning
GeoWeb	Spatial point pattern
Geoweb methods	Spatial representativeness
GIS	Spatial scale
GIScience	Strava
Governance	SWOT analysis
Human computer interaction	Urban planning
Internet-based public participation GIS	Voluntary geographic information
Mobile	Volunteered geographic information
Native vegetation	Web Mapping
Open data	

B Usage of AI tools in writing GIS master's thesis

Category	Details
Author	Tuukka Korpela
Date	17.11.2025
Declaration	I used AI tools in my thesis, the details are explained below.
AI Tool(s) Used	Google Gemini 2.5 Flash + Pro
Purpose of Tool Usage	• Writing and debugging LaTeX, Python, SQL, and Overpass QL code. • Understanding a large database and its contents. • Finding and utilizing tools in QGIS and Overpass API (OSM). • Creating initial ideas for analysis methods. • Explain correlation and regression diagnostics. • Analyzing textual data and categorizing it. • Finding literature sources. • Helping to structure the thesis. • Reformulating a sentence or a paragraph.

Category	Details
How AI Output Was Used / Changed	• **Code**: Some generated code was used "as is" by copying and pasting it onto an IDE, but I always went through the code and checked that I understand it in its main points. I also used it to help me format the LaTeX document. • **Technical/Conceptual**: I used the LLM as a sparring partner and a tutor with complicated topics or technical challenges. • **Writing**: I used the model as an "outsider" reader to point out logical errors and text clunkiness. First I wrote everything on my own, then inputted parts of the text for the LLM. I used the outputs to edit the text to make it more concise and fluent.

C Positional accuracy threshold by AI

An AI assisted accuracy threshold was considered as a proxy for spatial data quality before selecting zoom rate. The whole textual content of a survey was first retrieved from the Maptionnaire database and then plucked into the "survey_text" variable in the prompt below.

Under the prompt you'll find an example of the LLM output for a survey. The LLM used was Gemini 2.5 Flash.

```
prompt = f"""
You are an expert in analyzing Finnish and English urban planning
    PPGIS surveys created with Maptionnaire's mapping platform.
    Analyze the following survey text and extract three specific
    pieces of information:
1.  **mapping_task**: In two sentences, describe what the survey
    asks respondents to draw, place, or mark on the map.
2.  **accuracy_metric**: Based on the survey's mapping_task,
    suggest a threshold value in meters for positional accuracy. If
    there are multiple mapping tasks, suggest a value for each task.
3.  **explanations**: A two sentence description for task number 2
    on how you ended up with the suggestions.

Provide the output in a strict JSON format. Do not include any
    extra text or formatting outside of the JSON object.


EXAMPLE OUTPUT FORMAT:
{{
    "mapping_task": "(one sentence about mapping task)",
    "accuracy_metric": "(threshold value)",
    "explanations": "(two sentence explanation for the mapping task
     and accuracy metric)"
}}


SURVEY TEXT:
---
{survey_text}
---
"""
```

An example of an answer to the mapping task:

- The survey asks respondents to identify locations on a map where new active transportation infrastructure and services are needed, such as additional sidewalk ramps, bike share stations, bike parking, car share locations, and bus stops. It also prompts users to mark preferred future routes for biking, walking, or rolling, and to pinpoint areas of safety concern.

Accuracy metric:

- 3 meters

Explanation for the accuracy metric:

- The suggested accuracy threshold of 3 meters is chosen because many of the mapping tasks require precise identification of specific point features such as bus stops, bike racks, or sidewalk ramps, where a few meters can significantly impact utility and planning.

D Explore phase dataset after visual inspection

The table is behind this link: https://docs.google.com/spreadsheets/d/e/2PACX-1vQ9ZAdBZ4pRB_-IFeBUWZ4UTQfzTIKAowh4PA0EU0IVw2JsbfU3Z8r1hQPv5jJ0y8nxqm7XzVNikmVy/pubhtml?gid=869751183&single=true

This was the first feature set derived from the Maptionnaire database. The set was iterated multiple times and grew significantly for the final dataset. This dataset is given as an example of the data that was included in the analyses.

E Map encoding scheme

The map encoding scheme used to examine the use of background maps.

Table E1: Maptionnaire background map options.

System maps	Base maps	Interactive image
101_mapboxBlue	201_bingmap	301_iaImage
102_mapboxDark	202_geojson	
103_mapboxLight	203_geotiff	
104_mapboxOutdoor	204_mapbox	
105_mapboxRed	205_maptiler	
106_mapboxSatellite	206_shapefile	
107_mapboxSattellite	207_vectortileservice	
108_mapboxStreets	208_wfsmap	
109_mapboxYellow	209_wmsmap	
110_maptilerBright	210_wmtsmap	
111_maptilerDark	211_xyzmap	
112_maptilerLight		
113_maptilerOutdoor		
114_maptilerSatellite		
115_maptilerStreets		

F OSM-proxy notebook

The notebook is behind this link: https://github.com/Stubberson/project-collection/blob/4fbb7122338406f9a99f4120df029e06a7c10349/aalto-thesis/OSM_tagging.ipynb

The OSM-proxy was an attempt to match Maptionnaire survey responses with OSM through the Overpass API. It is a novel exploration into new ways to measure spatial data quality in PPGIS. With limited resources, its development for this study had to be cut short and it couldn't be applied as a robust spatial data quality metric. However, I believe it has potential as a proxy for spatial data quality, if developed further.

G Additional results

Additional and intermediate results referenced in the results chapter are presented here. They are presented in the order they are mentioned in the text. Figures G1 and G2 are from the explore phase, figures after them are from the explain phase.

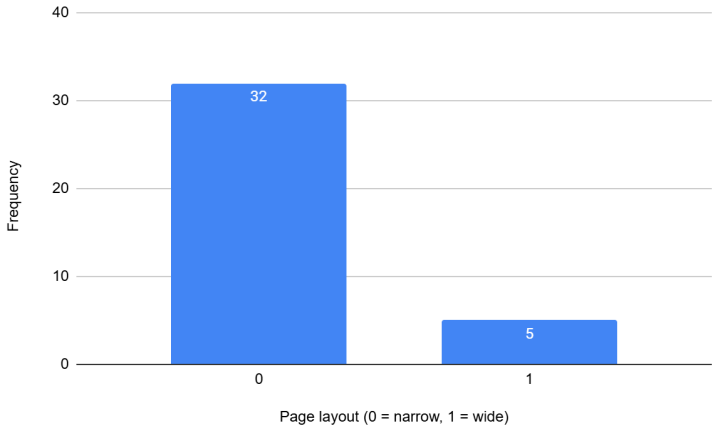


Figure G1: Default survey page layout frequency. 32 surveys used a narrow page layout as default and 5 surveys used the wide page layout.

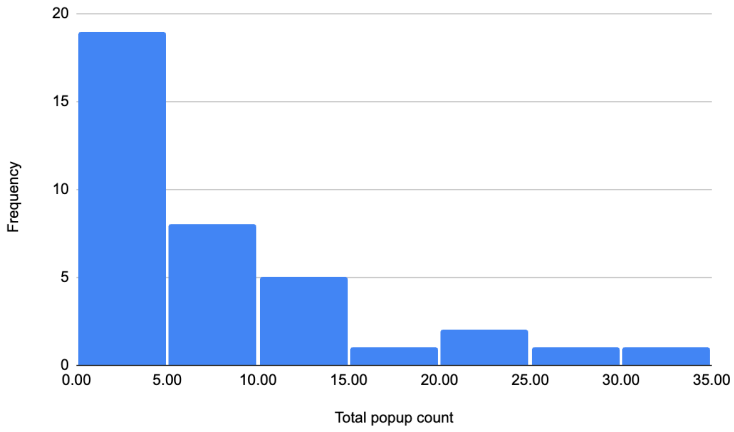


Figure G2: A distribution of surveys by total popup count with a right skew. The histogram has a bucket size of five. For instance, there are 19 surveys that have 0-4 popups, 8 surveys that have 5-9 popups, etc.

The element usage in the full dataset follow a similar trend as for the smaller explore dataset in figure 3.

Just like in the smaller dataset, the MapTiler Streets (115) was the most popular background map with 99 surveys choosing it as a background map in the full dataset (figure G4). However, the interactive image is not the second most popular background in this dataset, but the Mapbox Streets (108) with 40 surveys using it. These two

background maps are very similar in their visual representation with light, but high-contrast color schemes emphasizing the transportation networks. After the streets-style background maps, satellite maps were the most popular system maps (106 = Mapbox Satellite, 114 = MapTiler Satellite).

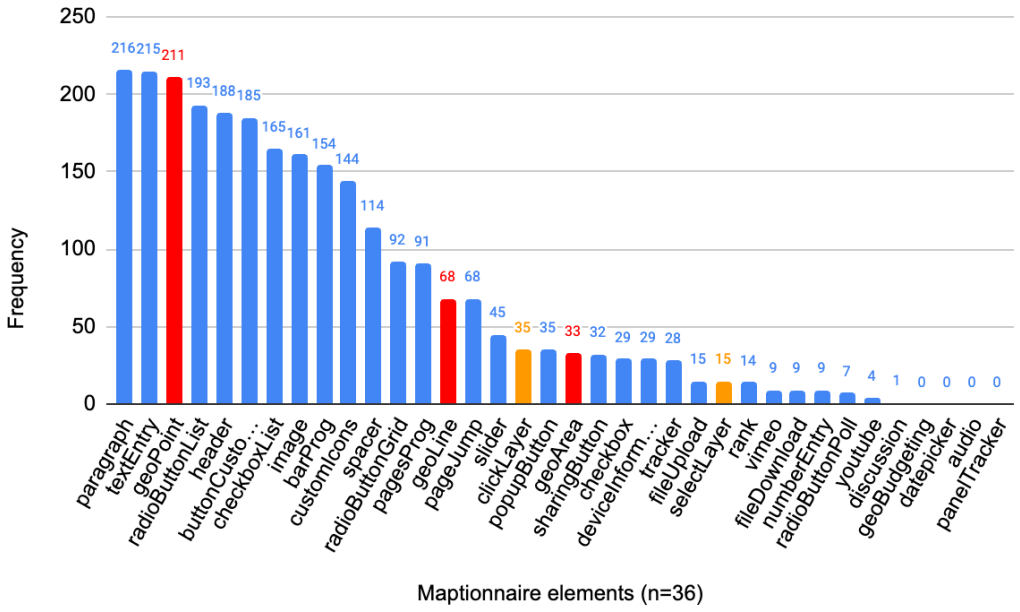


Figure G3: Frequency of surveys in the explain sample (n=216) using a Maptionnaire element. Mapping elements are highlighted in red (point, line, area, geobudgeting) and orange (select elements).

The frequent use of the streets-style maps seems to be tied to the survey themes in this dataset (figure G5: almost half (102, 47.2%) of the surveys were about mobility and transport. It is no big surprise that surveys concentrated on this theme use a background map designed for visualizing the road network. The second most frequent theme was inclusion and cohesion with 42 surveys (19.4%). The theme that was excluded from the analysis due to the so called "dummy variable trap" (e.g. [103]) was buildings & neighborhoods.

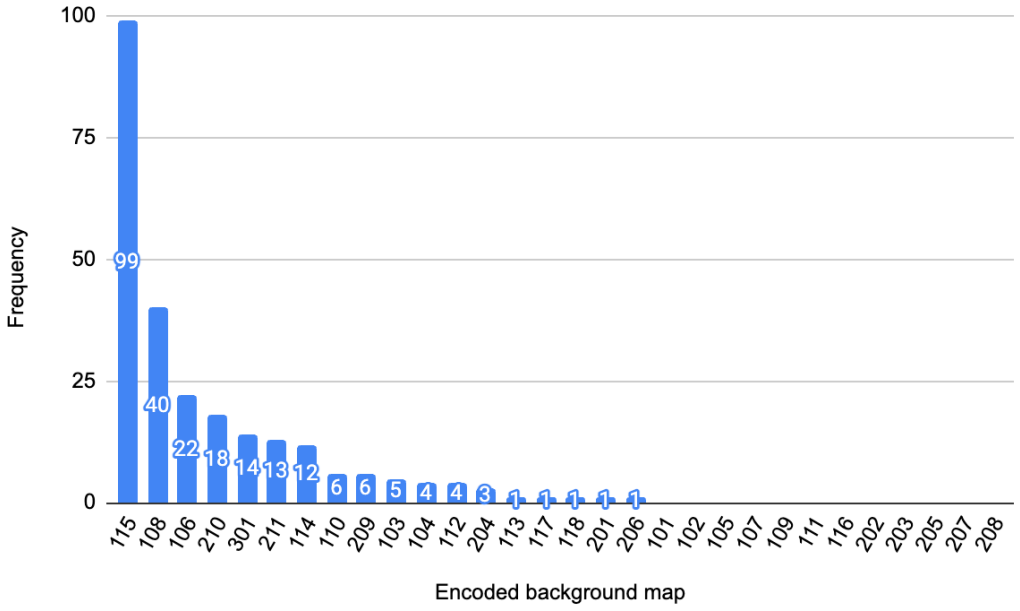


Figure G4: Frequency of surveys in the explain sample using a specific background map. In total, 251 background maps were used across the 216 surveys. Encoding scheme: 101-118 = Maptionnaire's OSM-based system maps, 201-211 user imported maps (e.g., wmtsmap), 301 = user imported image as a map.

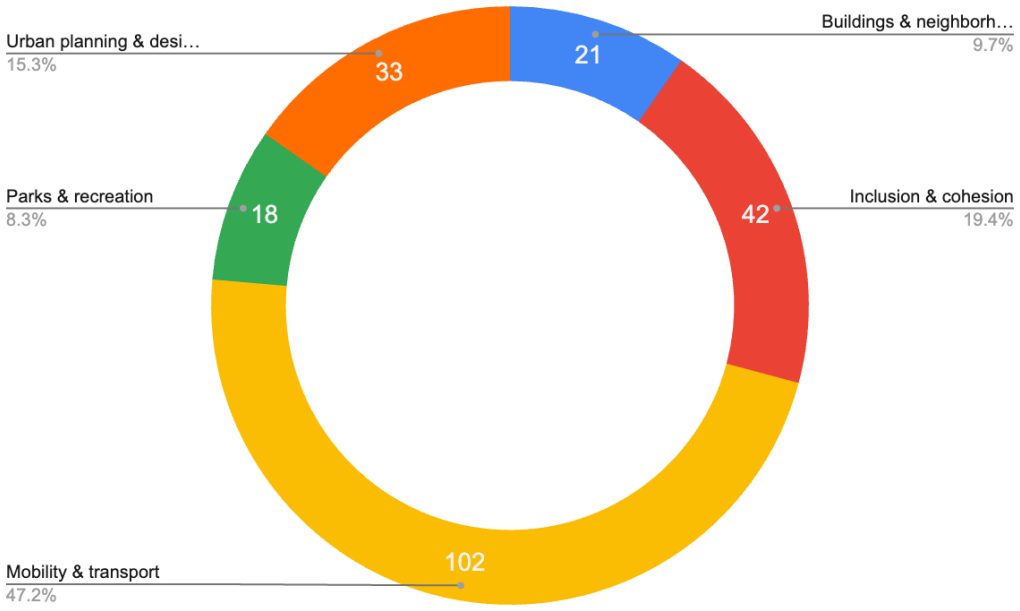


Figure G5: Survey theme distribution in the full dataset.

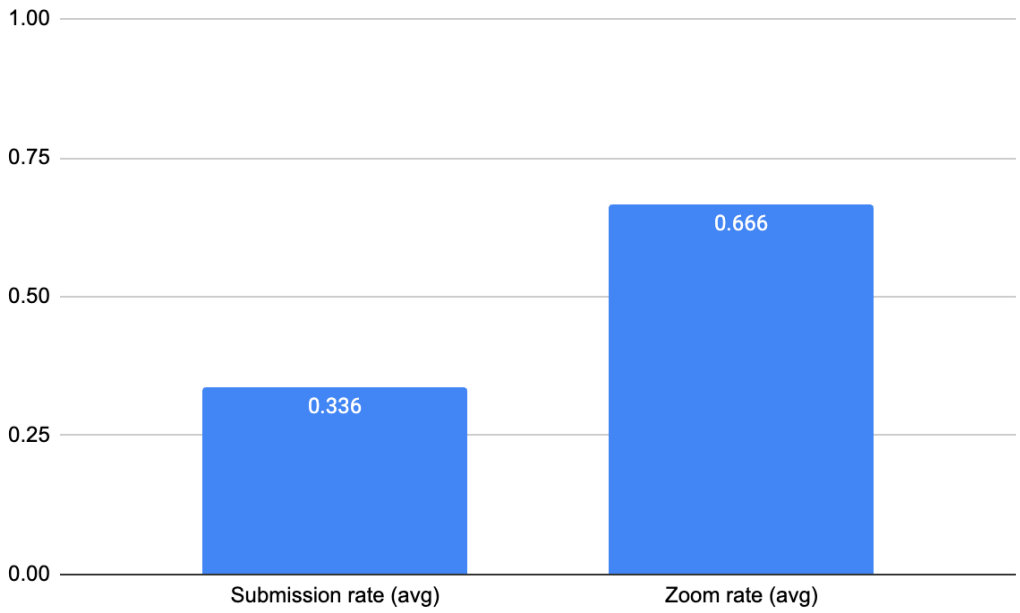


Figure G6: Average survey submission rate and zoom rate for the full dataset.

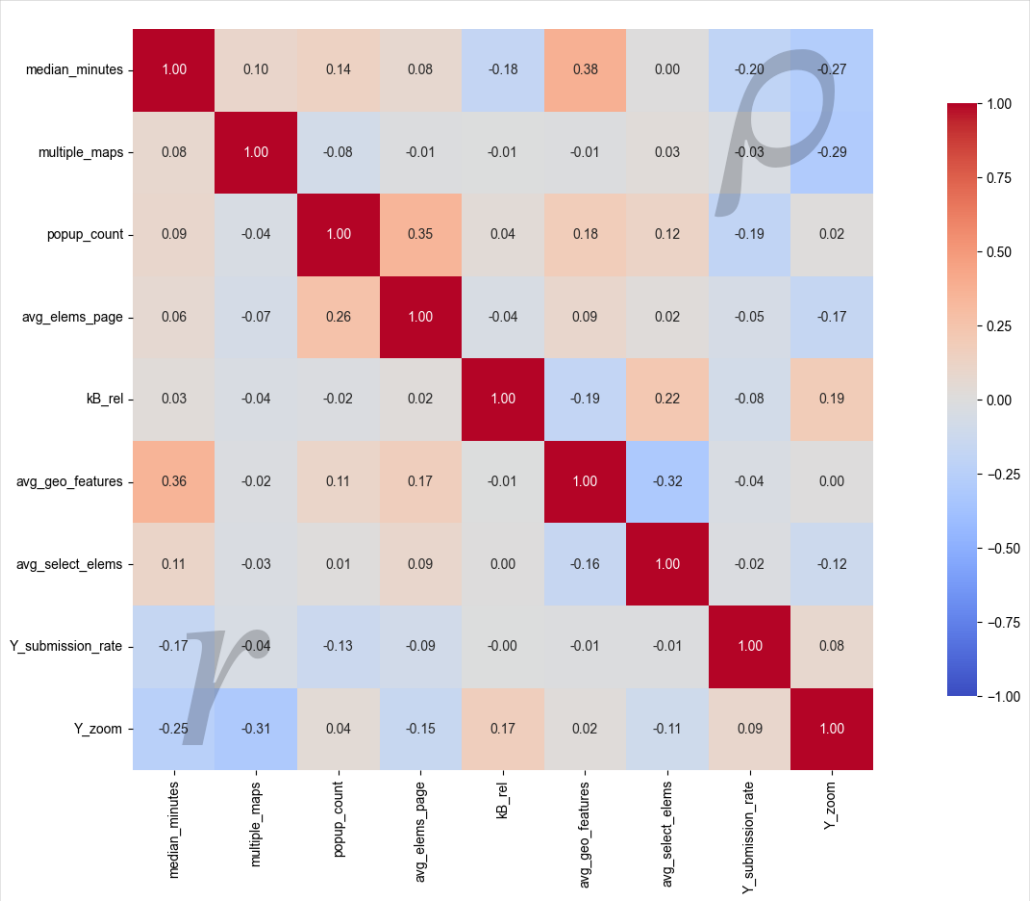


Figure G7: Pearson's and Spearman's correlation matrix for the explain phase with the same predictors as ended up in the explore phase. The matrix shows the Pearson's r and Spearman's ρ value for each feature pair. Pearson's coefficients are in the lower left half and Spearman's in the upper right half.

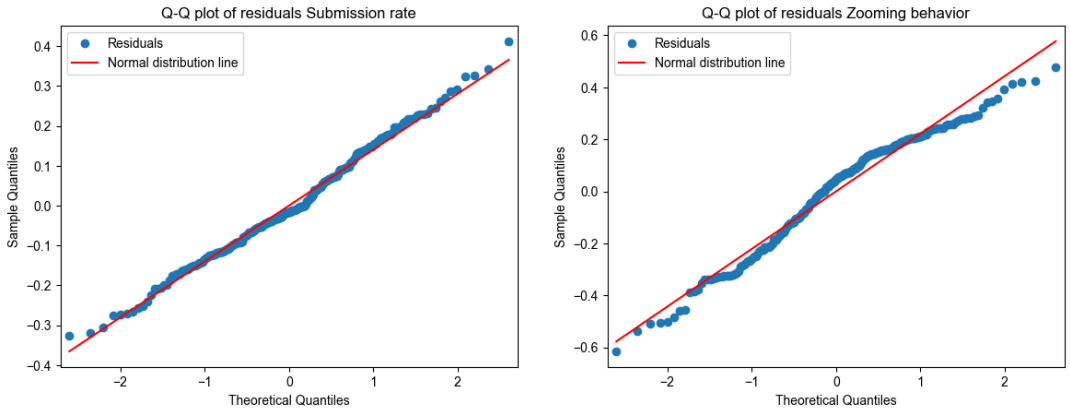


Figure G8: Quantile-quantile (Q-Q) plots of the response feature residuals. On the left the plot for submission rate and on the right for zoom rate. The red line is a theoretical normal distribution, the blue points are the residuals.

Table G1: Multiple OLS regression summary, diagnostics, and coefficients for **submission rate** in the first explain phase model.

Model summary			
No. observations	216	Log-Likelihood	117.72
Df residuals	208	Df model	7
R^2	0.052	F-statistic	1.580
Adj. R^2	0.020	F-statistic p-value	0.143
AIC	-219.4	BIC	-192.4

Model diagnostics			
Omnibus	2.300	Durbin-Watson	2.030
Omnibus p-value	0.317	Jarque-Bera (JB)	2.363
Kurtosis	2.752	JB p-value	0.307

Features	**Coefficient (β_p)**	**std. err.**	**p-value**	**[95% conf. int.]**
const	0.336	0.011	0.000*	[0.315, 0.357]
median_minutes	-0.027	0.009	0.004*	[-0.045, -0.008]
multiple_maps	-0.005	0.008	0.555	[-0.021, 0.011]
popup_count	-0.016	0.051	0.760	[-0.115, 0.084]
avg_elems_page	-0.010	0.014	0.470	[-0.037, 0.017]
kB_rel	0.001	0.013	0.971	[-0.025, 0.026]
avg_geo_features	0.012	0.010	0.225	[-0.007, 0.031]
avg_select_elems	0.004	0.022	0.852	[-0.039, 0.047]

Note: Heteroskedasticity robust standard errors are used.

* $p < 0.05$

Table G2: Multiple OLS regression summary, diagnostics, and coefficients for **zoom rate** in the first explain phase model.

Model summary			
No. observations	216	Log-Likelihood	18.933
Df residuals	208	Df model	7
R^2	0.226	F-statistic	7.479
Adj. R^2	0.200	F-statistic p-value	5.00e-08*
AIC	-21.87	BIC	5.137

Model diagnostics			
Omnibus	10.604	Durbin-Watson	2.144
Omnibus p-value	0.005*	Jarque-Bera (JB)	9.216
Kurtosis	2.464	JB p-value	0.010*

Features	**Coefficient (β_p)**	**std. err.**	**p-value**	**[95% conf. int.]**
const	0.666	0.016	0.000*	[0.635, 0.697]
median_minutes	-0.067	0.023	0.004*	[-0.113, -0.022]
multiple_maps	-0.072	0.016	0.000*	[-0.104, -0.040]
popup_count	0.025	0.038	0.514	[-0.049, 0.099]
avg_elems_page	-0.051	0.018	0.005*	[-0.086, -0.015]
kB_rel	0.043	0.027	0.108	[-0.009, 0.096]
avg_geo_features	0.033	0.021	0.114	[-0.008, 0.074]
avg_select_elems	-0.012	0.015	0.422	[-0.041, 0.017]

Note: Heteroskedasticity robust standard errors are used.

* $p < 0.05$