MSc Programme in Urban Management and Development

Rotterdam, the Netherlands

August 2022

Analysis of Limiting Factors to Biodiversity Data Collection on the Citizen Science Platform iNaturalist

Name: Alicia DeMaio

Supervisor: Dr. Alexander Los

Specialization: Urban Environment, Sustainability, and Climate Change

Country: United States of America

Report number: 1674

UMD 18

IHS

Making cities work

Institute for Housing and
Urban Development Studies of
Erasmus University Rotterdam

Summary

Biodiversity and ecosystem services are in a state of global decline. This is alarming not only for species and ecosystems facing extinction, but also for human populations that are dependent on the services and resources that they provide. Impacts on biodiversity and ecosystems are driven in part from land-use change due to urbanization. Nevertheless, cities provide diverse habitats and are just as important as rural areas in protecting biodiversity. To restore these resources, well-informed policies and actions that safeguard global biodiversity must be implemented. To do so, stakeholders need to understand current baselines and changes to biodiversity over time. This is most effectively done by using primary species occurrence data, such as that collected by citizen science platforms.

iNaturalist, a citizen science platform specializing in biodiversity, provides species population and distribution information by enabling users to record species occurrence data. This helps professional scientists by providing data that would be difficult, time consuming, or expensive to attain. Citizen scientists around the globe represent a huge advantage to biodiversity monitoring by providing the human capital needed for such tasks, yet the data collected exhibit taxonomic, spatial, and temporal gaps. These gaps pose a threat to comprehensive and effective biodiversity management and conservation.

One main objective of this study is to explore various socio-economic, socio-cultural, and platform characteristic barriers to global biodiversity data collection on iNaturalist, that may contribute to spatial, temporal, and taxonomic data gaps. To do this, a survey questionnaire was distributed on the iNaturalist forum. 149 survey responses from 24 different countries around the world were collected and analysed using descriptive and inferential statistics to summarize and interpret the data.

Results of a factor analysis found that income, access to technology, experience, accessibility to transport, and education were significant socio-economic factors related to data collection. Motivation and free time were found to be socio-cultural factors that influence data collection. Platform characteristics were not found to have significance in the context of the survey distributed. The study also found that respondents claim to have little bias towards aesthetically pleasing species. This supports that societal preference, and not species charisma, may be a prominent cause of taxonomic bias and data gaps.

Keywords

Biodiversity, citizen science, ecosystem services

Acknowledgements

This thesis is dedicated to my husband, Nicolas, whose encouragement, love, and endless support has been a source of strength throughout this challenging year. You believe in me more than anyone else ever has. Thank you, from the bottom of my heart, for all the sacrifices you've made to help me achieve my goals. I am a better person with you in my life.

Secondly, I am grateful for the thoughtful guidance and support of my advisor, Dr. Alexander Los. Thank you for sharing your wisdom and enthusiasm on this topic, and for making this an enjoyable process. Thank you, Dr. Los and IHS staff, for all the inspiration, knowledge, and laughs you've shared with me.

Lastly, I thank my UMD18 cohort and the City of Rotterdam for making this an unforgettable year. You have opened my eyes, showed me new ways of thinking, and changed me for the better.

Foreword

“In the center of the American Museum of Natural History’s Hall of Biodiversity, there’s an exhibit embedded in the floor. The exhibit is arranged around a central plaque that notes that there have been five major extinction events since complex animals evolved, over five hundred million years ago. According to the plaque, “Global climate change and other causes, probably including collisions between earth and extraterrestrial objects,” were responsible for these events. It goes on to observe: “Right now we are in the midst of the Sixth Extinction, this time caused solely by humanity’s transformation of the ecological landscape” (Kolbert, 2014 p.267).

Abbreviations

AI	Artificial Intelligence
ES	Ecosystem Services
GBIF	Global Biodiversity Information Facility
IHS	Institute for Housing and Urban Development Studies
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
LGBTQ+	Lesbian, Gay, Bisexual, Transgender, Queer
MEA	Millennium Ecosystem Assessment
PCA	Principal Component Analysis
USA	United States of America

Table of Contents

List of Charts

List of Figures

List of Tables

1. Introduction

1.1 Background

By 2050, cities will be home to up to 75% of the world's population (United Nations, 2012). Globally, land-use change driven primarily by agriculture, forestry and urbanization, is the driver with the largest relative impact on terrestrial and freshwater ecosystems (IPBES, 2019). Urbanization and associated land use change is a major threat to global biodiversity and ecosystem productivity through loss of habitat, biomass, and carbon storage (Seto, Güneralp, & Hutyra, 2012). "Land cover change could lead to the loss of up to 40% of the species in some of the most biologically diverse areas around the world" (Seto et al., 2012 p.16083). More broadly, anthropogenically modified biotic (habitat connectivity, human population etc.) and abiotic drivers (carbon emissions, noise, light etc.), through ecological and evolutionary processes, influence biodiversity (Uchida et al., 2021). Despite these threats, cities can be vital for native biodiversity conservation by providing natural, semi-natural, and artificial ecological systems with a range of habitat types that support urban biodiversity (Aronson et al., 2017). In fact, Kowarik (2011) asserts that cities are often more rich in plant species than rural areas and can thus play an important role in biodiversity conservation. This illustrates that communities all over the planet, from dense megacities of Tokyo and Delhi to rural and agricultural communities in Ghana have equal stake in addressing the biodiversity losses seen today.

According to the IPBES (2019) around one million animal and plant species are currently threatened with extinction, many within decades. The extraordinary pressures and threats to endangered species today prompt thoughtful and expedient actions informed by real-world models and data. Data such as species distribution is currently being collected, in part, by citizens around the world. Citizen science, the mobilization of the general public in scientific research, is expanding in global capacity with the flourishing of on-line platforms. At the same time, global interest in citizen science is growing, fueled by the potential for positive environmental impacts with expectations that collecting environmental data can achieve various goals (van Noordwijk et al., 2021). Indeed, modern citizen science has proved itself to be a vital source of biodiversity data through wildlife monitoring and image classification (van Noordwijk et al., 2021). This data is crucial to monitoring how biodiversity responds to threats and interventions and understanding how to preserve and manage our shared biodiversity (Callaghan, Ozeroff, Hitchcock, & Chandler, 2020).

The accumulation and distribution of biodiversity data, however, is subject to a range of factors and constraints, resulting in spatial, temporal, and taxonomic gaps. These gaps are explained in more detail in Section 4.2. One of the most emblematic disparities in data is summarized by Amano, Lamming, & Sutherland (2016) in which Global Biodiversity Information Facility (GBIF) bird records accelerated dramatically over the past three decades, even in some data-poor regions, while the rate of increase of nonbird records remained low or is slowing down. This illustrates the enduring disparities in species data over space and time. It is important to know the causes of data gaps and why they persist in order to improve the quality and efficacy of data collection and scientific studies.

Efforts to monitor and protect biodiversity are not merely an act of charity for positive ecological change. "Major losses of populations and species clearly impede the provision of ecosystem goods and services, with consequent impacts on human well-being" (Ceballos, Ehrlich, & Raven, 2020 p.13600). Ecosystem services are "the direct and indirect contributions of ecosystems to human well-being" (Braat & de Groot, 2012 p.5). Although there is need for further empirical research, biodiversity is widely described as having a positive relationship with ecosystem services by providing resources like food, medicine, clean air, space for recreation, and tourism, thus producing positive effects on human health and well-being. By protecting ecosystems and global biodiversity, therefore, humans ultimately protect their own best interests. Ecological characteristics and metrics, collected through citizen science, are used to evaluate and quantify ecosystem services. Monitoring ecosystem services over time allows scientists and government officials to assess whether they are improving or degrading. This allows for adjustments and scaling of different policies that relate to ecosystem services and/or biodiversity.

Citizen science organizes and enables human capital at a global scale while contributing to scientific research. Promoting and empowering this resource has enormous potential for the well-being of millions of species, including our own. Humans need diverse species and ecosystems around the globe for survival. Collecting more biodiversity data, through citizen science, allows researchers and scientists to understand current issues through more thorough analyses of species population changes, ecosystem assessments, policy efficacy, and countless other applications, which in turn benefit the entire biosphere.

1.2 Problem Statement

Liu, Dörler, Heigl, & Grossberndt (2021) define citizen science platforms as web-based infrastructures with a single entrance point that contain one or more of the following features: includes active citizen science projects and activities, displays citizen science data and information, provides tools that support citizen science projects and activities, presents good practices for users, and offers relevant scientific outcomes for people involved or interested in citizen science. Platforms are used by numerous stakeholders including citizens, scientific institutions, public administrations, the media, and policymakers (Liu et al., 2021).

iNaturalist (www.inaturalist.org) is an opportunistic citizen science platform hosted by the California Academy of Sciences and National Geographic Society. Users can contribute observations, which include date, time, and coordinates, for any living organism. These observations are then classified to the most specific taxonomic level possible by other community members. Observations with adequate community agreement on taxon identity meet the "research grade" threshold and are regularly uploaded to the Global Biodiversity Information Facility (GBIF). The GBIF is an international network and data infrastructure that collects and shares biodiversity information. GBIF data can be used for taxonomic revisions, environmental modelling, studying threatened species, and biodiversity assessments (Yesson et al., 2007). Since 2012, iNaturalist users have contributed over 44 million occurrences to GBIF (iNaturalist contributors, iNaturalist, 2022). Furthermore, data from iNaturalist is the second most downloaded source of data from the GBIF (Callaghan et al., 2020).

The amount of biodiversity data collected through iNaturalist and GBIF is impressive, yet it is well known that there are stark spatial, temporal, and taxonomic gaps therein. Faith et al. (2013) explain that GBIF information is used in many ways such as modelling species diversity status and tracking biodiversity conservation. In addition, one main advantage of species occurrence data is the ability to provide multiple snap shots of biodiversity status and distribution (Faith et al., 2013). For these reasons, "strategies to create a balanced spread of data (geographically, taxonomically, and temporally) are essential to facilitate meaningful analysis and interpretation of biodiversity as a whole" (Faith et al., 2013 p.45).

iNaturalist is a popular global citizen science platform which has made significant contributions to biodiversity data. It is important to understand all possible dynamics of data collection, in order to optimize collection. Insights on data limitations and data quality issues of citizen science to date have focused on platform design, funding, or other factors, without exploring the user perspective or user characteristics that may contribute to data gaps. This presents an important knowledge gap and an opportunity to learn more about data limitations. This paper aims to explore the global user perspective and experience of iNaturalist, in order to uncover new or confirm existing limitations of biodiversity data collection. This study, without locational or culturally specific context, has potential to be replicated to other local, national, or global citizen science platforms. Results from this study can be used to alleviate, remove, or spark innovative solutions to the barriers of data collection.

1.3 Research Objective

There are two objectives of this study. The first objective is to explore various socio-economic, socio-cultural, and technological barriers to global biodiversity data collection on the platform, iNaturalist, that may contribute to spatial, temporal and taxonomic data gaps.

The second objective is to propose solutions or further research in relation to the barriers to biodiversity data collection. Overcoming these barriers ultimately leads to benefits that include more accurate biodiversity monitoring and ecosystem service assessments, and enhanced recreation to iNaturalist users.

1.4 Research Question

The main research question of this thesis is:

- What factors limit citizen science biodiversity data collection?

The following sub-questions support the main research question:

- What socio-economic factors limit citizen science biodiversity data collection?
- What socio-cultural factors limit citizen science biodiversity data collection?
- What platform characteristics limit citizen science biodiversity data collection?

1.5 Significance of the Study

The overall aim of this study is to help reduce global taxonomic, spatial, and temporal biodiversity data gaps, by identifying various barriers and challenges to citizen science biodiversity data collection. Biodiversity data is used not only for conservation, but also assessing ecosystem services and informing policy. Answering the research questions above would facilitate better biodiversity data collection, thus allowing more robust and thorough assessments and scientific studies. Identifying and removing barriers to data collection would also help to better understand and evaluate ecosystem services that benefit humans.

Furthermore, citizen science provides co-benefits such as recreation and education. It also increases scientific literacy, understanding of the natural sciences, and awareness of environmental issues by providing discussion fora and features such as automatic species identification. Facilitating data collection by identifying and alleviating challenges experienced by users would allow for enhanced enjoyment and satisfaction of participation in citizen science.

1.6 Scope and Limitations

The focus of this study is on factors contributing to biodiversity data gaps for users of iNaturalist. This study does not address user perspectives or experience of any other citizen science platforms. The study chose iNaturalist because it had capability to distribute a survey directly to users, on the forum. Other platforms considered did not share this feature, which exhibits a limitation of citizen science research in general: it is difficult to reach citizen scientists. The unit of analysis is the individual user, and thus does not include the perspective or experience of other users such as policy makers or scientific institutions.

The scope of this study also excludes acute analyses of taxonomic, temporal and spatial data gaps resulting from iNaturalist. The time and resources needed for such an analysis are beyond that of this study.

1.7 Organization of the Paper

This paper is organized into five chapters. The first chapter has provided background information justifying the scope and goals of the study. Chapter two provides a summary of the most relevant concepts to biodiversity citizen science. Chapter three outlines the methodology followed during the design and implementation of the study. This includes the operationalization of research concepts and variables, the sample selection, and data analysis methods used. Chapter four includes a profile of the survey respondents, main survey results, and statistical analysis. Lastly, chapter five provides the main conclusions of the study and recommendations for further research.

2. Literature Review

2.1 Introduction

This chapter provides a review of the relevant concepts to the study including citizen science, biodiversity, and ecosystem services. Citizen science in the context of biodiversity conservation is also included to provide a more detailed analysis of citizen science that is more relevant to this paper. Scientific literature identifies key variables of the associated data gaps. These variables and concepts are ultimately organized into a conceptual research framework, based on the research questions provided herein.

2.2 Citizen Science

Citizen science is the participation of the general public in various stages of scientific research such as collecting, categorising, transcribing, or examining scientific data (Njue et al., 2019). Citizen science is not a new concept and has evolved remarkably over centuries. The timing of cherry blossoms in Kyoto, Japan, for example, have been recorded for 1200 years by local citizens, so long that it has been helpful in climate reconstructions (Kobori et al., 2016). Likewise, biological data collections, cultivated commonly through museums over time, have contributed to spatial models for biodiversity patterns, to species richness estimation, and to ecological/environmental studies on species attributes (Faith et al., 2013). This illuminates the rich history of diverse non-professional volunteers contributing to science.

Today, citizen scientists are involved with a variety of projects spanning scientific disciplines such as astronomy, air quality, deforestation, and weather monitoring (Njue et al., 2019). In addition to the natural sciences, involvement of non-professionals has also contributed to fields of molecular engineering, quantum science, and neuroscience (Troudet, Grandcolas, Blin, Vignes-Lebbe, & Legendre, 2017). Technological developments such as internet provision, sensing equipment, and smartphones with mapping and global positioning systems have increased the feasibility of large-scale citizen science projects in natural sciences (Njue et al., 2019). Smartphone applications and mobile web access are commonly used to carry out nature observations which can then be compiled into datasets and accessed by professional scientists using digital tools. These contributions have made it possible for millions of users to contribute to scientific publications globally. One platform, eBird, collects millions of observations every month and has contributed to at least 90 peer-reviewed articles or book chapters (Kobori et al., 2016).

These advances in technology, together with vast human capital, have also allowed for data collection over extensive geographic regions that would otherwise be impossible to collect due to time and resource limitations (Kobori et al., 2016). There are also monetary advantages to utilizing citizens in research. Theobald et al. (2015) estimated that the contribution of people contributing to their global study of 388 citizen science projects was between $667 million to $2.5 billion annually. These advantages of human capital make a compelling argument to pursue and include citizen scientists in scientific research.

Frigerio, Richter, Per, Pruse & Vohland (2021) argue that the future of citizen science has potential beyond crowdsourcing and data collection. Contributing to other phases of scientific research, such as the formulation of hypothesis and research questions, enables co-creation of knowledge and a better understanding of evidence-based decision making. Engaging citizens in all phases the scientific process will likely increase scientific literacy and usher in a breadth of perspectives that help in describing and addressing issues (Frigerio et al., 2021). Increasing literacy creates a positive feedback cycle resulting in better quality data (Callaghan et al., 2020).

Despite its numerous advantages, citizen science has its share of shortcomings. “Data quality and funding (sustainability) of citizen science projects are still the most critical concerns of citizen science” (Balázs, Mooney, Nováková, Bastin & Arsanjani, 2021 p.141). The meaning of data quality can vary depending on the context and stakeholders involved, but terms like completeness, availability, standards-based, validity, consistency, timeliness, accuracy and bias commonly define and describe issues of data quality (Balázs et al., 2021). For such a wide-reaching problem, there is, understandably, not a single solution or approach to reconcile data quality. Balázs et al. (2021, p.153) however finds that overall, approaches such as “adaptable project aims and survey protocols; volunteer training; the use of experts; automated and statistical analysis; and finding an appropriate project structure” can help to improve data quality.

Data quality can also be improved by artificial intelligence (AI). Automated reasoning and machine learning is currently used to confirm the accuracy and consistency of citizen science contributions (Ceccaroni et al., 2019). There are also a number of types and applications of AI in development for use in the near future. Automated reasoning, for example, can filter out irrelevant data by validating outputs though automatic procedures, and can also alert people about what might occur around them (Ceccaroni et al., 2019). This could be helpful in improving consistency and targeting species that are less represented in citizen science data.

Citizen science has clear challenges yet represents a global network of diverse human capital and potential across a range of scientific disciplines. Citizen science is a valuable resource to understanding global challenges, to increasing public interest, to informing organizations, and to influencing management strategies and policies.

2.3 Biodiversity

Biodiversity is “the variability among living organisms from all sources, including terrestrial, marine, and other aquatic ecosystems and the ecological complexes of which they are part”

(MEA, 2005 p.1). Biodiversity has many complex components including the diversity of all organisms, the diversity within and among species and populations, and the diversity of ecosystems (MEA, 2005). In other words, biodiversity includes genetic, organismal, and ecological diversity (Gaston & Spicer, 2004). Studying biodiversity is no simple matter, considering about 10 million species live on Earth (Troudet et al., 2017).

Biodiversity is declining faster now than any other time in human history (IPBES, 2019). "An average of around 25 percent of species in assessed animal and plant groups are threatened suggesting that around 1 million species already face extinction, many within decades, unless action is taken to reduce the intensity of drivers of biodiversity loss" (IPBES, 2019 p.4). Indeed, urgent global action is needed. Drivers of changes in nature with the largest impact include: changes in sea and land use; exploitation of organisms; climate change; pollution; and invasive species (IPBES, 2019). These rapid changes to habitats, ecosystems, and species populations present an urgent need to monitor global biodiversity (Frigerio et al., 2021). Monitoring biodiversity is crucial for understanding drivers and evaluating solutions by providing critical information to develop conservation strategies (Kelling et al., 2019). International treaties such as the Convention on Biological Diversity express the need to examine change and trends of global biodiversity (Chandler et al., 2016). Also, the main objective of the IPBES is to provide current assessments of available knowledge for better policy decisions and action (IPBES, 2019). It is clear that institutions recognize the need and value of monitoring biodiversity. Methods of doing so are provided in the next section.

2.3.1 Indicators & Indices

Given its complexity and multidimensionality, there is no single, perfect measure of biodiversity, even for small areas. Species diversity indices, however, are used as indicators of biodiversity. Still, "there is absolutely no universal index suitable for all theoretical or real cases in ecological research" (Fedor & Zvaríková, 2019 p.337). Existing indices vary in complexity and performance (what it measures) and should be chosen carefully. Single species richness, the number of species in a given area, is a relatively simple and common method of expressing biodiversity. "Ecologists prefer to express species richness in its more sophisticated form as species richness indices" (Fedor & Zvaríková, 2019 p.338). This is done by "correlating species richness with distribution of all elements in a sample or community within their relative abundance or dominance" (Fedor & Zvaríková, 2019 p.339). The Shannon Index is an example of a well-known biodiversity measure. This index is calculated by "taking the number of each species, the proportion each species is of the total number of individuals, and sums the proportion times the natural log of the negative of this sum" (Nolan and Callahan, 2006 p.334).

The City Biodiversity Index (or Singapore Index) is another tool in which cities can evaluate and monitor progress of biodiversity conservation efforts in relation to their baseline. This index uses 23 indicators that measure native biodiversity, ecosystem services, and governance and management of biodiversity, which are all assigned a value of 0-4 points. Ten of the indicators in the City Biodiversity Index address native biodiversity in the city and include the changes to the number of certain species like birds and butterflies.

"Indicators allow researchers to analyse, monitor and efficiently measure the conditions, characteristics, trends, and rates of change of UES's [urban ecosystem services]" (Haase et al., 2014 p.419). Since biodiversity is so complex, a combination of indicators is needed to cover each aspect of biodiversity. The three most important motivations to improve biodiversity are species conservation, ecological resilience, and biological control of pest organisms (Duelli &

Obrist, 2003). An index for conservation, for example, might contain the conservation values (i.e., red list status) of all species present in that area. The inputs for this indicator illustrate the value for citizen science data collection of species. Species data helps scientists to evaluate and quantify biodiversity in an area, based on the index used. Understanding these metrics would allow for better policy and planning in urban areas, and for comparisons against baseline scenarios.

These methods of biodiversity evaluation serve to illustrate how biodiversity is understood but also show the complicated nature of estimating the biodiversity in both rural and urban areas. Nevertheless, species data is important for not just biodiversity studies but also for assessing the state of ecosystem services, discussed in Section 2.5.

2.4 Citizen Science and Biodiversity Conservation

The current utilization and future potential for citizen science for biodiversity conservation is striking, as biodiversity monitoring projects are the most common citizen science contributions among life science (Frigerio et al., 2021). Over 80% of biodiversity data in Europe and over 50% of GBIF data is recorded by citizen scientists (Frigerio et al., 2021). The majority of citizen science programs are located in North America and Europe, with relatively few in Africa, Asia, Central and South America (Chandler et al., 2016). “Increasingly, such data have helped to assess the impacts of threats to species, including pollution, disease, and climate change” (Faith et al., 2013 p.42). It is clear that citizen science has made clear and plentiful contributions to biodiversity research and conservation.

2.4.1 Data Gaps

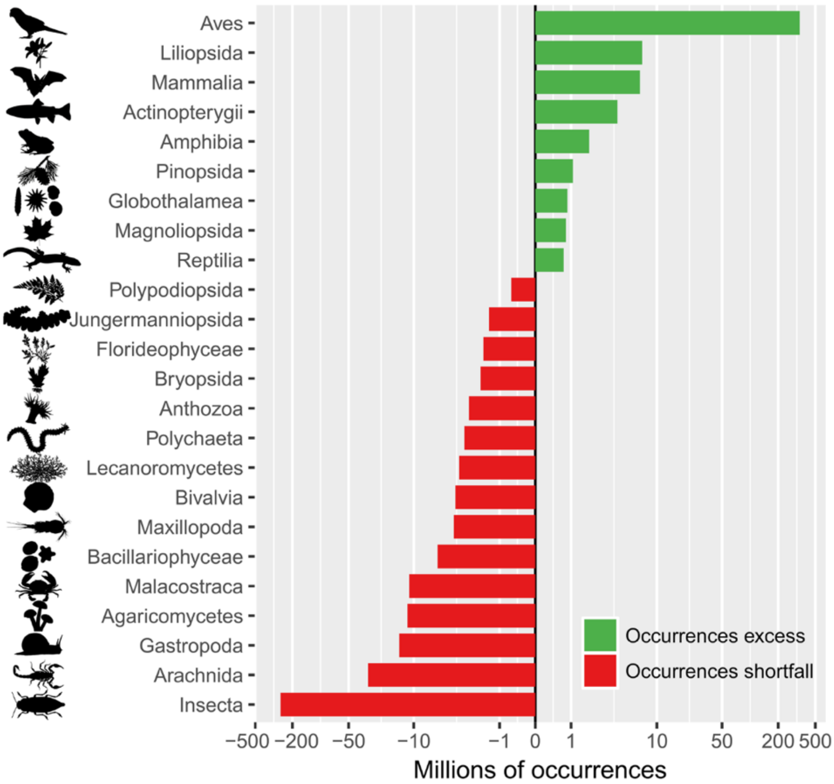
Despite the vast contributions of citizen scientists, accumulation of data is affected by a range of factors, resulting in marked taxonomic, spatial, and temporal gaps and inconsistencies in biodiversity data gathering. “Biodiversity data gaps must be addressed in order to properly find synergies and trade-offs among (these) different aspects of human well-being" (Faith et al., 2013 p.50).

2.4.1.1 Taxonomic Data Gaps

Taxonomic bias, the preference and effort to study some species over others, is evident in opportunistic observation records and the choices of taxa[1] studied professionally. Troudet et al. (2017 p.3) found that in a study of 626 million occurrences in GBIF data, more than half of the records were bird occurrences, “even though birds represent only 1% of the total number of species catalogued in GBIF”. Figure 1 shows the taxonomic bias in biodiversity occurrence data among all 24 species classes studied by Troudet (2017) and shows the variability in species data collection. “The vertical line at x=0 depicts the ‘ideal’ number of occurrences per class, where each class is sampled proportionally to its number of known species. Green and red bars show the classes that are over-and under-represented in the GBIF mediated database compared to this ‘ideal’ sampling, respectively” (Troudet, 2017, p.3). A strong taxonomic bias is shown by the fact that the number of observed species per class is not proportional to their respective species richness. The study shows for example, that in contrast to birds, Arachnida (spiders, ticks, mites etc.) had one of the lowest median numbers of occurrences per species despite being 3 times more species rich.

[1] For the purpose of this paper, “taxa” and “species” may be used interchangeably

Figure 1: "Taxonomic bias in biodiversity occurrence data".



Source: Troudet et al., (2017, p.3)

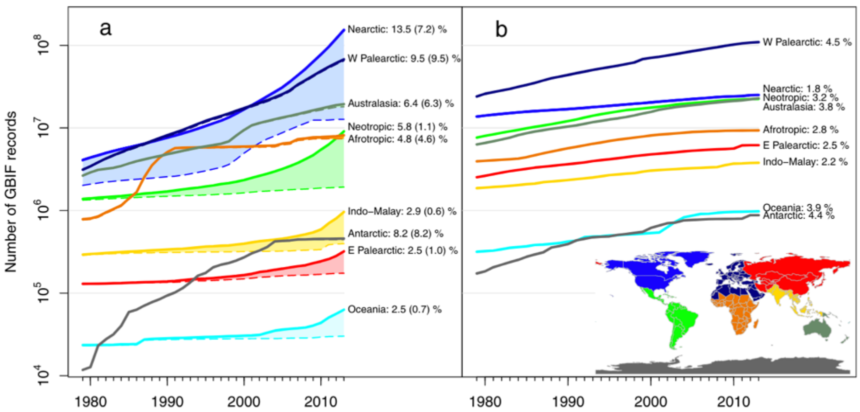
Taxonomic bias can be caused by a number of factors. Dos Santos et al. (2020) explain that among terrestrial mammals, species charisma, or aesthetic appeal, may influence observation research. Human interest can also be driven by anthropomorphism (degree of similarity to humans) (dos Santos et al., 2020). While these traits reflect human interest, their impact on research is less clear (dos Santos et al., 2020). Troudet et al. (2017, p.8) instead suggest that societal preferences (as measured by internet searches) are a factor determining what data is recorded, after their study suggested a "positive and significant correlation between public interest and the number of occurrences in GBIF".

Troudet et al. (2017, p.1) explain that taxonomic bias "prevents reaching global conclusions and developing efficient conservation plans". At the same time, uncharismatic species can play significant functions in ecosystems (Troudet et al., 2017). Reducing taxonomic gaps would help to achieve equitable conservation goals across species and ecosystems.

2.4.1.2 Spatial & Temporal Data Gaps

In addition to taxonomic gaps, there are also considerable differences in the spatial and temporal distribution of data. "The unequal distribution of biodiversity data across the globe, particularly the lack of information in biodiversity-rich regions, has repeatedly been reported since the 1980's" (Amano et al., 2016 p.393). For example, a study of the accumulation of bird and non-bird GBIF data since the 1980's shows a 9% increase of bird records in the two data-richest regions (Nearctic and Western Palearctic realms) and the slowest increase of 3% in the Eastern Palearctic, Oceania, and Indo-Malay realms (Amano et al., 2016). In contrast, non-bird records increased by only 1.8% in the Nearctic realm, with a similar or even higher rate of increase as birds in the Eastern Palearctic and Oceania realms (Amano et al., 2016). The study also showed that the number of non-bird records declined dramatically in two biodiversity rich regions (Afrotropic and Neotropic) compared with the preceding two decades, suggesting that "scientific efforts to collect and share species occurrence data have at best not improved-and even declined-in some data-poor regions despite spatial information gaps being recognized as a challenge since the 1980's" (Amano et al., 2016 p.394). These changes of bird vs non-bird occurrence records by biogeographic realm in the GBIF between 1979 and 2013 are visualized in Figure 2. "Possible factors that have been suggested to cause spatial information gaps include wealth, insufficient experience, infrastructure and communication, and inaccessibility due to geographical location and/or security level" (Amano et al., 2016 p.399).

Figure 2: Cumulative occurrence records of birds and non-birds in GBIF



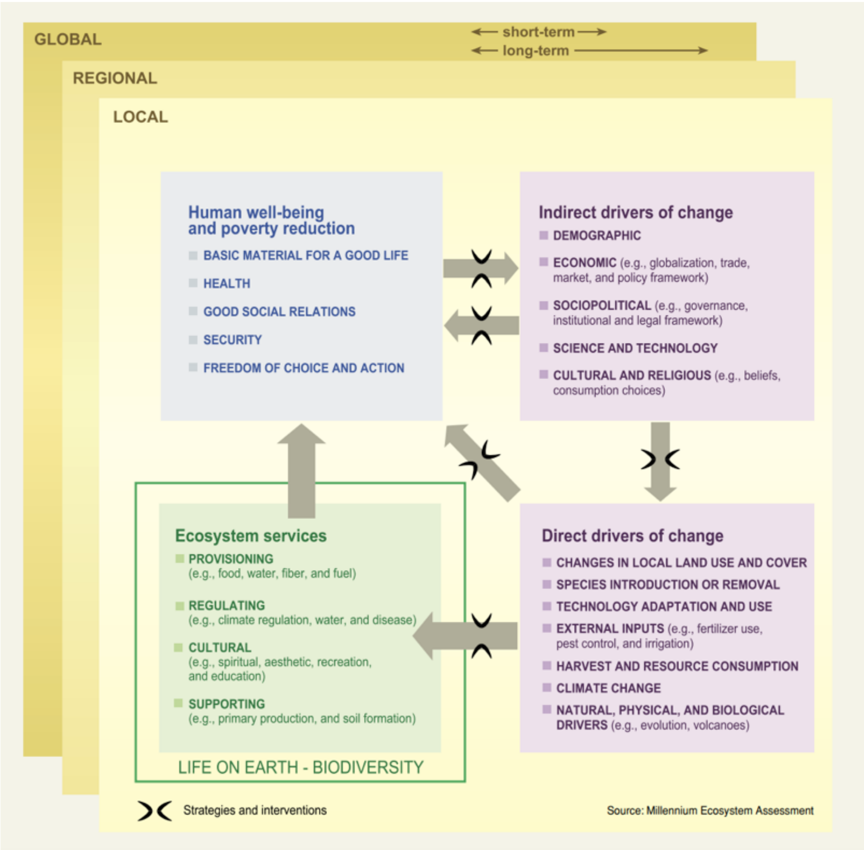
Source: Amano et al. (2016, p.395)

2.5 Ecosystem Services

Ecosystem services are "the direct and indirect contributions of ecosystems to human well-being" (Braat & de Groot, 2012 p.5) . Ecosystem services became a bridging concept of the natural and social sciences after a half-century of growing awareness and policy regarding environmental pollution and resource scarcity, followed by inclinations toward sustainable economic development (Braat & de Groot, 2012). Daily (1997) supplied the first comprehensive account of the services supplied by nature and the ways people depend on them. Costanza et al. (1997) in the same year estimated the economic value of the biosphere, which spurred great interest in the topic.

The Millennium Ecosystem Assessment (MEA) (Millennium Ecosystem Assessment, 2005) was a major international assessment of changes to ecological systems as they pertain to human well-being. This report organized ecosystem services into four categories: supporting, provisioning, regulating, and cultural. "These [services] include provisioning services such as food, water, timber, and fibre; regulating services such as the regulation of climate, floods, disease, wastes, and water quality; cultural services such as recreation, aesthetic enjoyment, and spiritual fulfilment; and supporting services such as soil formation, photosynthesis, and nutrient cycling" (MEA, 2005 p.1-2). The MEA grouped all types of ecosystem services into the realm of biodiversity (Figure 3).

Figure 3: Conceptual framework of interactions between biodiversity, ES, human well-being, and drivers of change.



Source: MEA (2005, p. iii)

The MEA highlighted that ecosystem services were degrading because of the continuous loss of biodiversity (Gan, 2021). “Changes in biodiversity alter the processes of ecosystems as well as their resistance and resilience in the face of environmental changes, exerting a profound influence on ecosystem services and further affects human health and well-being through a feedback mechanism” (Gan, 2021 p.2). This illustrates the inextricable link between biodiversity and ecosystem services, from which we can reason that biodiversity losses ultimately result in service losses to humans.

Figure 4 shows that 15 out of 18 categories of nature’s contribution to people globally, from 1970 to 2019, are in decline. “The observed rapid degradation of the ability of ecosystems to generate services not only necessitates a better understanding of how to maintain important ecosystem functions but also requires that this knowledge is put into a broad institutional and governance context” (Haase et al., 2014 p.414). Thus, understanding the mechanisms that influence ecosystem services is in the best interest of humankind in order to sustain the crucial services we depend on.

Furthermore, Haase et al. (2014 p.414) explain that “ES [ecosystem services] and their contribution to quality of life, human health, and well-being are dependent upon the level of biodiversity at the ecosystem and landscape level”. Few studies, however, discuss the relationship between urban biodiversity and ecosystem services (Haase et al., 2014).

Figure 4: Trends of global ecosystem services from 1970 to 2019

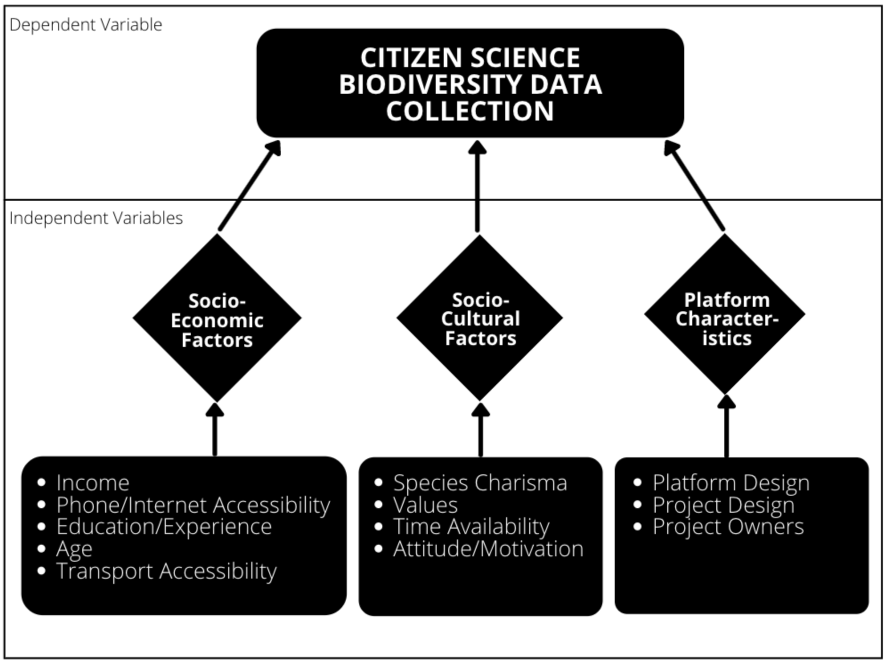
Source: IPBES (2019, p.12)

Comparing ecosystem services around the world can be difficult. Similar to biodiversity, indicators (such as those listed in Figure 4) are used to quantify, understand, and communicate the plethora of services provided by ecosystems to humans, and ultimately aid in performing ecosystem assessments. Indicators have been classified in many different ways by different institutions such as the MEA, The Economics of Ecosystems and Biodiversity, and the Common International Classification of Ecosystem Services (Czúcz et al., 2018). These international classification systems aim to be universal, yet they inherit certain contexts and goals of their origins, which results in stakeholders having to choose relevant services and customize underlying definitions for their own national, regional, and local use (Czúcz et al., 2018). The amount of ecosystem services classifications helps, given the complexity and range of contexts and applications of ecosystem services, yet the lack of a singular classification makes comparisons difficult (Czúcz et al., 2018).

2.6 Conceptual Framework

Figure 5 provides a conceptual framework identifying three factors that are expected to influence citizen science biodiversity data collection: socio-economic factors, socio-cultural factors, and platform characteristics. These are further elaborated by their respective sub-variables. Combined, these different variables are hypothesised as factors that influence citizen science biodiversity data collection.

Figure 5: Research conceptual framework.



Source: (Author, 2022)

3. Research Design and Methodology

3.1 Introduction

This chapter outlines the research design and methods used to answer the research questions and achieve the research objectives. First, the research strategy is provided, in line with the overall aim of the study. Primary and secondary data types and collection methods are defined and described, followed by a description of the sample population. Indicators for the concepts outlined in Chapter 2 have been operationalized for quantification, followed by data analysis methods. Limitations and challenges of the study are provided at the end of the chapter.

3.2 Research Design and Methodology

3.2.1 Research Strategy

The main objective of this inductive research is to understand the factors that limit citizen science biodiversity data collection on iNaturalist. The nature of this study requires collecting and analysing demographic information as well as opinions and usage of the platform. This is then compared to actual citizen science data to study the relationship of the survey responses to data gaps and observation rates.

For this research, a survey questionnaire research strategy was chosen for data collection. Van Thiel (2014) acknowledges that a survey is suitable to study people's opinion and to gather large amounts of data. A survey strategy was chosen because it enables the collection of a wide range of new data and responses while allowing generalization of the results for the entire population. Also, global perceptions and behaviours are influenced by cultural and social factors, resulting in complex and nuanced responses. A survey questionnaire has the ability to quantify diverse indicators.

3.2.2 Data Types and Data Collection Methods

The research conducted herein consists of primary and secondary data sources. The main data collection method by means of the survey questionnaire was used to gather primary data about the data collection practices of users as well as their opinions about a range of statements. This resulted in quantitative data collection. The survey consists of mostly structured questions and statements organized on a Likert-scale. The questions and statements result from the operationalized indicators, discussed in Section 3.3, based on literature review. The survey is organised into three sections: habitual use of the platform; Likert-scale questions about user preferences, experiences, and opinions; and demographic questions. The survey is carried out using the general community forum on the iNaturalist platform.

Secondary iNaturalist data is used by accessing the GBIF and extracting research-grade occurrence data corresponding to the countries in which survey respondents live. The data is organized into two "groups": the United States of America (USA) and all countries within Europe (USA/Europe); and all countries outside USA and Europe (All Other Countries) in which respondents of the survey live. The formation of these groups was done to compare differences in survey responses between areas that contribute more observations (USA/Europe) with areas that contribute less observations (All Other Countries). This data is used to triangulate results from the survey questionnaire.

3.2.3 Sample Size and Selection

The research population for this study was chosen to be the active users on iNaturalist. An active user is someone that has contributed an identification, observation, comment, or post in the last 30 days. This was used, as opposed to the total registered users, because being a registered user does not necessary imply regular contribution of data. In this regard, the active users are a more accurate representation of the population that engages in data collection. According to iNaturalist Site Stats (2022) there were 250,420[2] average active users in the past year. Since the unit of analysis for this research is the individual user, the research population is the total number of active users on iNaturalist. Considering the predetermined timeline of the research, it was determined that the confidence level of the study be 95% with a confidence interval of ±9%. Based on a scientific sample size calculator[3], at the stated level of confidence and confidence interval, the calculated representative sample size is 119 users.

There are multiple ways for individual users to interact and communicate on iNaturalist. Some users identify and/or confirm the species observations of other users, some post journal entries for others to read, and some engage in discussions on project home pages. Users can even send private messages to each other. Reaching the entire userbase of iNaturalist, and many other citizen science platforms, however, is difficult if not impossible. The most realistic way to reach the most users, given time constraints of the study, was determined to be the community forum, where users can communicate with each other on different topics or concerns. The survey questionnaire was distributed online, through a link to the Qualtrics form on the "General" topic of the community forum, accessible to all iNaturalist users. This results in a voluntary/self-selection response sample, a type of non-probability sample.

3.3 Operationalization: Variables and Indicators

Operationalization involves taking concepts and turning them into measurable variables, which show what will be studied and/or measured (Van Thiel, 2014). To determine the factors leading

[2] Calculated using the data from the first day of each month from July 2021 to June 2022 on https://www.inaturalist.org/stats

[3] Sample size calculated using https://www.surveysystem.com/sscalc.htm

to spatial, temporal, and taxonomic data gaps, an operationalization of variables (Table 1) has been developed, resulting from the research questions and conceptual framework. This study outlines three independent variables in line with the research objectives: socio-economic factors, socio-cultural factors, and platform characteristics. These variables are expected to have an influence on the dependent variable, citizen science data collection. A number of sub-variables and indicators are deduced from these variables and the context in which they are used. The indicators were then used to formulate questions for the questionnaire in order to collect data to answer the research questions.

Table 1: Operationalization Table.

Research Questions	Variables		Indicators	Sample Questions	Type of Data/ Data Collection Method
What socio-economic factors limit citizen science biodiversity data collection?	Socio-Economic	Income	-Current socio-economic class -Ability to pay for special equipment	-What is your current socio-economic status? -Do you use any other equipment besides a smartphone for data collection?	Quantitative / Survey Questionnaire
		Phone/Internet Accessibility	-Access to phone and internet	-Do you experience phone or internet service disruptions because of inability to pay?	
		Education/ Experience	-Highest level of education -Experience with iNaturalist	-What is your highest level of education? -How long have you used iNaturalist?	
		Age	-Age	-What is your age?	
		Transport Accessibility	-Ability to travel to areas of interest -Accessibility to green spaces	-Do you have access to a variety of green spaces? -Are you able to travel to places of interest?	
What socio-cultural factors limit citizen science biodiversity data collection?	Socio-Cultural	Species Charisma/ Aesthetic Appeal	-Interest in visual aesthetic	-Do you record observations for one particular species? -Do you avoid certain taxa for being scary/ugly?	
		Values	-Perceived outcomes or personal goals	-What are your objectives for recording observations?	
		Time Availability	-Free time	-Would you collect more data if you had more free time? -How often do you record new observations? -Do you usually record new observations on weekdays, weekends, or both?	
		Attitude/Motivation	-Interest in volunteer work (projects)	-Are you interested in volunteering time to iNaturalist projects?	
What platform characteristics limit citizen science biodiversity data collection?	Platform Characteristics	Platform Design	-iNaturalist functionality	-Do you find the functionality of iNaturalist to be sufficient? -Has interacting on the forum of iNaturalist improved your species observation and/or identification skills?	
		Project Design	-Breadth of project subjects	-Are there species you want to study but there aren't projects/initiatives for?	
		Project Owners	-Interaction with project owners	-Do you feel supported by project owners?	

Source: (Author, 2022)

3.3.1 Data Analysis Methods

Quantitative data ascertained from the survey questionnaire was first organized and coded using excel software. Further analysis was conducted using statistical software. Descriptive statistics such as percentages and variance were used to summarize the results. Inferential statistics via a factor analysis was also performed to uncover socio-economic, socio-cultural, and/or platform characteristic factors present in the survey response data. Finally, the results of the factor analysis were then compared to survey data and iNaturalist observation data of the respondents.

3.4 Research Challenges and Limitations

As already mentioned, a general limitation of the study was the ability to contact citizen scientist communities. While exploring citizen scientist platforms, it was realized that very few had functionality that allowed mass communication to its users. Other avenues such as citizen science committees or forums similarly lacked ways to communicate with the community. iNaturalist was unique by having a forum where a survey could be distributed, yet another limitation is the bias of the voluntary sample. Respondents who participate in the forum are likely to have strong opinions and/or be more engaged in iNaturalist than the rest of the population.

Another challenge is the global scope of iNaturalist. The population on iNaturalist has no geographic boundaries and thus the respondents are influenced by unlimited socio-economic and socio-cultural factors. For this reason, the questionnaire needed to be kept broad, as to apply to all respondents. Income, for example, was evaluated by economic classes instead of annual household income to account for different currencies around the world. Similarly, acute cultural factors were excluded from the study, considering the vast nuanced cultural influences around the world.

Furthermore, the survey questionnaire was only distributed in the English language. This is a result of the time constraint of the study, but regardless a major limitation in collecting a diverse set of data.

Lastly, understanding and summarizing the taxonomic data profiles from the corresponding respondent countries would be extremely time consuming. For this reason, it is assumed that taxonomic bias exists throughout the areas studied. In reality, the extent to which these biases exist is extremely variable and context specific.

4. Results, Analysis, and Discussion

4.1 Introduction

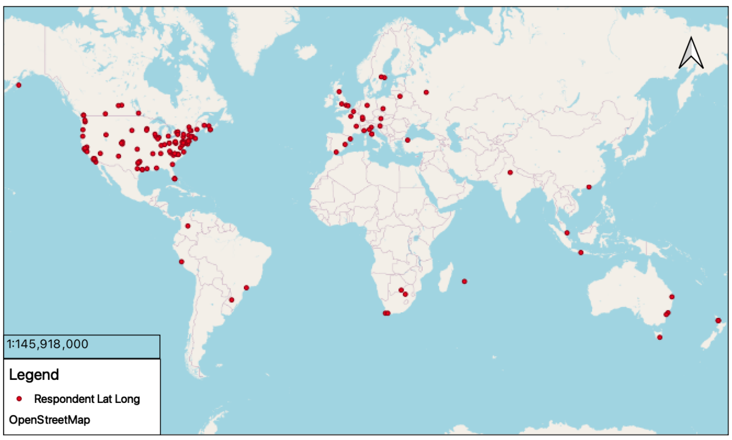
The following section provides the results of the study. First, an overview of the survey respondent characteristics, locations of residence, and a summary of their use of iNaturalist is provided. Next, results of a factor analysis are provided and interpreted. Lastly, iNaturalist data is collected from the countries in which the respondents live. This data is summarised and presented to provide insight into actual data collection. Finally, a discussion synthesises the results and key findings of these analyses.

4.2 Description of the Survey Population

4.2.1 Demographics

A demographic and socio-economic profile of survey respondents is provided in Table 2 including characteristics of the sample population such as age, gender, level of education and socio-economic status. This data represents responses of 149 iNaturalist users, exceeding the target sample population of 119 by 25%. Survey respondents are from 24 different countries across the globe (Figure 6), with 58% of responses from the United States of America. 49% of iNaturalist users describe their living area as urban while all others as peri-urban, rural, or not sure.

Figure 6: Respondent locations of residence.



Source: Author (2022)

Table 2: Socio-economic characteristics of respondents.

Characteristic	Values	Frequency	Percent
Age	18 or younger	14	9.3%
	19-29	36	24%
	30-65	86	57.3%
	66+	11	7.3%
Gender	Prefer not to say	5	3.3%
	Male	81	54.0%
	Female	55	36.7%
	Non-binary/ third gender	7	4.7%
Socio-Economic Class	Not Sure/Prefer to not answer	15	10%
	Lower Class/Poor	16	10.7%
	Average Middle Class	74	49.3%
	Upper Middle Class	36	24%
	Upper Class	8	5.3%
Highest Level of Education	Prefer to not answer	2	1.3%
	Less than High School	15	10%
	High School Degree	19	12.7%
	Undergraduate Degree	37	24.7%
	Graduate Degree	49	32.7%
	Doctoral Degree	27	18%
Type of Living Area	Not Sure	3	2%
	Rural	28	18.7%
	Peri-Urban (rural-urban transition zone)	44	29.3%
	Urban	74	49.3%
Employment Status (past 3 months)	Working full-time	55	36.7%
	Working part-time	19	12.7%
	Unemployed and looking for work	11	7.3%
	A homemaker or stay-at-home parent	1	0.7%
	Student	29	19.3%
	Retired	20	13.3%
	Other	13	8.7%

Source: Author (2022)

Table 2 shows that men (54%) were sampled relative to all other genders and 57.3% of respondents are between the age of 30-65. Only 7.3% of users are of retirement age (66 years of age or older) or and only 9.3% of users are 18 years of age or younger. 88% of respondents have a high school degree or higher.

73.3% of respondents belong to the average middle and upper middle class with lower class/poor and upper class representing only 10.7% and 5.3% of the total sample population, respectively. Lastly, 36.7% of respondents are full-time workers while 19.3% are students and 13.3% are retired.

4.2.2 Respondent Usage of iNaturalist

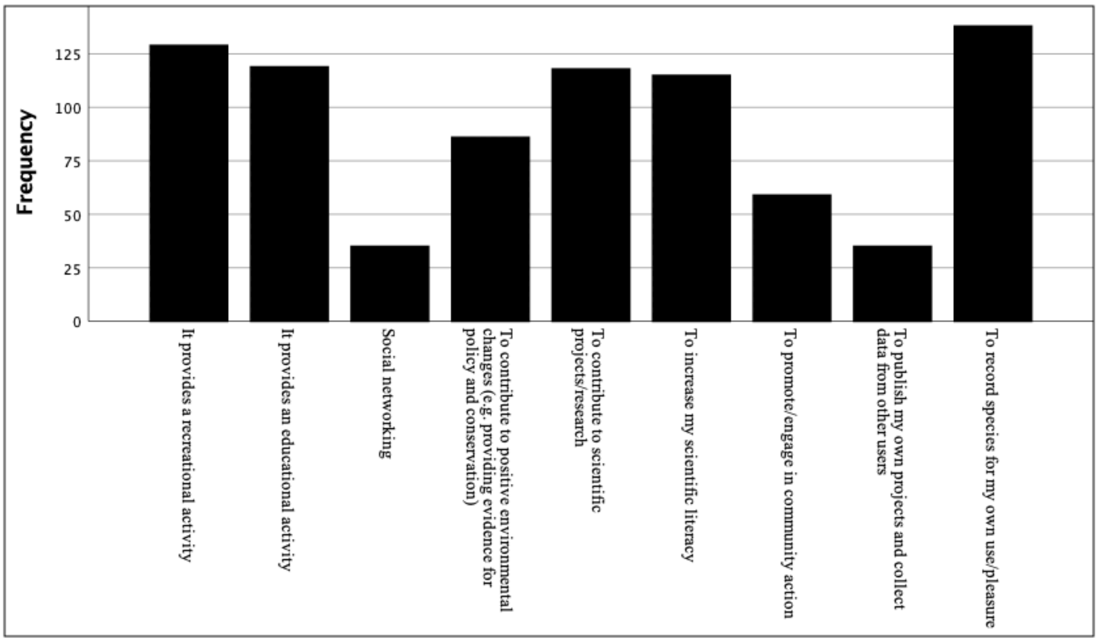
Table 3 provides a summary of how long respondents have used iNaturalist and a summary of how often and when they record observations. This data was recorded to provide insight into the experience of the user and to correlate the Likert-scale questions to frequency of observation. The majority of respondents have used iNaturalist between 1 and 5 years, record observations on a weekly basis, and record both on weekdays and weekends. Chart 1 summarises the reasons that respondents use iNaturalist (note: respondents were allowed to select all that applied to them). The top five most frequent answers were: to record data for their own use/pleasure (16.5%), for a recreational activity (15.5%), for an educational activity (14.3%), to contribute to scientific projects/research (14.1%), and to increase scientific literacy (13.8%).

Table 3: Respondent usage and experience of iNaturalist.

Characteristic	Values	Frequency	Percent
How long respondent has used iNaturalist	< 1 year	25	16.7%
	1-5 years	92	61.3%
	5-10 years	29	19.3%
	10+ years	3	2.0%
How often respondent records new observations	Varies/Not sure	37	24.7%
	Daily	39	26.0%
	Weekly	65	43.3%
	Monthly	7	4.7%
When respondent records new observations	Weekdays	4	3.3%
	Weekends	15	10.0%
	Both	130	86.7%

Source: Author (2022)

Chart 1: Reasons respondents use iNaturalist.



Source: Author (2022)

4.3 Survey Results

The majority of the survey questionnaire was designed using a Likert-scale. The questions were grouped into scales where the respondents could choose a level of agreement (from strongly disagree to strongly agree) or frequency of agreement (never to always). The Likert-scale results are included in Appendix 1. Of particular interest was the responses to taxonomic observations. Only 26.8% of respondents agreed or strongly agreed that it is more likely that they record for visually appealing species. On the contrary, 70.5% of respondents agreed or strongly agreed that they generally record observations for any species, regardless of visual appeal and 89.8% claimed that they never avoid recording observations for unattractive or scary taxa. This was surprising because species charisma/aesthetic appeal was considered to be a driver of taxonomic gaps, yet most users claim that this isn't a factor determining what they record. Some results, albeit intuitive, correspond to expectations. 69.1% of respondents said they would record more if they had more free time and 66.4% said they would record more if they had more specialized equipment.

Respondents were generally more neutral on questions relating to projects on iNaturalist although only 7.5% reported never having participated in them. 44.6% were neutral in regard to believing that project goals are achievable. This can be explained, in part, that not all projects have specific goals. In addition, 38.3% were neutral about projects covering an adequate range of species. Lastly, 47.7% were neutral about feeling supported/motivated by project creators/administrators.

Another interesting result was in regard to accessibility. 47.7% responded that they were able to travel to places of interest sometimes while 49.7% responded that they frequently record species observations for those that are easy to access. This indicates that mobility may be a limiting factor, while keeping in mind that "ease of access" could be interpreted differently.

4.4 Factor Analysis

Thiel (2014) explains that a factor analysis is used to find the underlying relationship between a group of variables and estimates the extent to which it co-varies with other variables, and thus whether the variable being tested forms part of the factor.

A factor analysis was chosen as part of the survey analysis to reduce the information gleaned from the survey to show the factors that explain the most variance in the data. These factors are shown in the result as groups of closely related values that correspond to similar questions/sub-variables. This result is the "factor loading" or the correlation between responses to a question. The expectation is that this analysis would reduce complexity and identify the prominent variables and sub-variables that form socio-economic, socio-cultural, and platform factors that limit citizen science data collection.

A principal component method was chosen because there were no preconceived expectations about the results (Van Thiel, 2014). In addition, the principal component analysis method reduces the number of observed variables to a smaller number of principle components which account for the most variance in the observed variables. A direct oblimin rotation was used to allow for correlation between the factors. The Kaiser-Meyer-Olkin measure of sampling adequacy is .595, above the recommended value of 0.5, and Bartlett's Test of Sphericity was significant (<.001). Coefficients in the pattern matrix (Table 4) were required to have a value of at least $|0.5|$, showing common variance among items. These metrics prove that factor analysis was suitable for the study.

The eigen values, above a value of 1, indicated that 12 factors explained 69% of the variance. Eight items (Q5, Q6, Q7, Q8, Q12, Q23, Q24, Q26) were eliminated because they did not contribute to a factor and did not meet the minimum coefficient level, $|0.5|$.

Table 4: Pattern Matrix of the PCA.

Pattern Matrix[a]

	Component											
Question	1	2	3	4	5	6	7	8	9	10	11	12
Q1											0.678	
Q2										0.783814		
Q4		-0.729638										
Q9								0.624				
Q10		-0.743										
Q11					0.829							
Q13										0.528		
Q14	0.794											
Q15	0.801											
Q16	0.831											
Q17	0.704											
Q18									0.854			
Q19											0.735	
Q20				0.649								
Q21				0.745								
Q22		-0.616										0.529
Q25						0.872						
Q29												0.864
Q30			0.842									
Q31								-0.662				
Q32			0.845									
Q33							-0.806					
Q34					-0.789							

Source: Author (2022)

Although the pattern matrix identified 12 components, not all of them presented contextual significance. Component three, for example loads age and highest level of education together, however, this phenomenon is not unique to is survey or population. Generally, education tends to increase with age. Components 10 and 12 did not clearly correlate to a sub-variable of the study. Also, only components with more than one factor loading were identified as significant. This resulted in omitting components 6, 7, and 9 from the analysis. In the end, six components, shown in Table 5, were considered to have contextual and statistical significance to the survey. The percentage of variance (Table 5) indicates the total amount of variability in the dataset that is explained by each factor.

Table 5: Summary of Factors.

Component/Factor	Representative Variable/Sub-variable	Percentage of Variance
1 (1 in Pattern Matrix)	Motivation	13%
2 (2 in Pattern Matrix)	Accessibility (Transport)	9%
3 (4 in Pattern Matrix)	Access to Technology	6%
4 (5 in Pattern Matrix)	Accessibility (Time)	5%
5 (8 in Pattern Matrix)	Income	4%
6 (11 in Pattern Matrix)	Experience	3%

Source: Author (2022)

4.5 iNaturalist Data

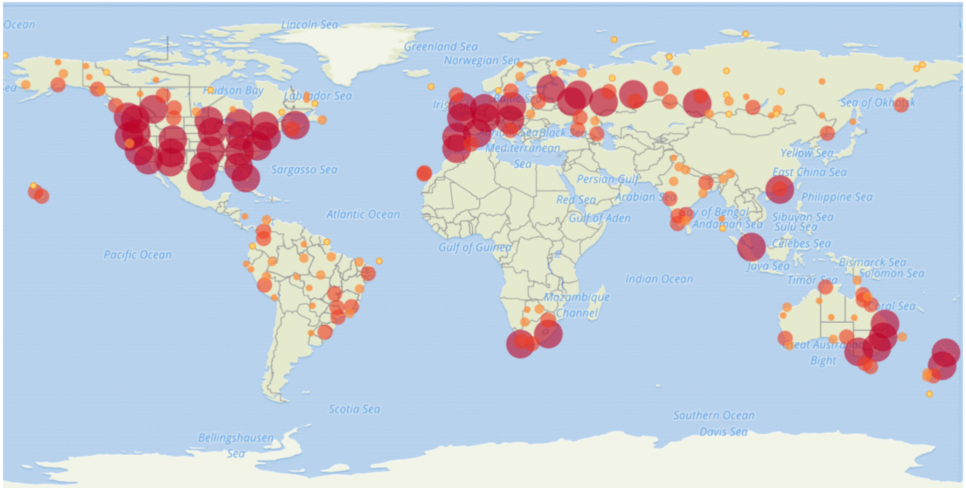
To evaluate the survey answers in relation to actual citizen science performance, iNaturalist data was extracted from the GBIF repository. Table 6 provides a snapshot of the research-grade observation rates recorded on iNaturalist from March, April, and May 2022 in USA/Europe, and All Other Countries. As discussed in Section 3.2.2, this is provided to not only illustrate the spatial variance in data, but to also compare the survey answers in relation to "data-rich" and "data-poor" areas. The purpose of this is to compare survey responses to actual observation rates. This was done by comparing the responses to questions that composed the factors outlined above, for USA/Europe and All Other Countries. Table 6 shows that observation rates are almost three times higher in USA/Europe than All Other Countries. Figure 7 shows the spatial distribution of occurrence data, with higher occurrences in North America and Europe than anywhere else.

Table 6: Summary of Respondent Observations (March-May 2022) on iNaturalist.

Area	Number of Observations	Percentage of Total	Average Observations Per Month
USA	1,566,872	73%	681,150
Europe (10 countries)	476,579		
All Other Countries (13 countries)	760,637	27%	253,545

Source: iNaturalist contributors, iNaturalist (2022).

Figure 7: Map of Respondent Observation Occurrences on iNaturalist from March to May 2022.



Source: GBIF.org (2022)

In addition, Appendix 2 includes two charts showing the quarterly distribution of iNaturalist observation data in USA/Europe and All Other Countries, across 24 taxonomic classes from July 2021 to June 2022. The taxonomic classes are the same as those chosen by Troudet (2017) as shown in Figure 1. This is provided to show the taxonomic, spatial, and temporal discrepancies in observation data. These charts show a great deal of variance between species class over time and space. For example, Appendix 2 shows that a similar total number of observations were recorded for birds (1,473,906 in USA/Europe; 765,293 in All Other Countries) as insects (1,840,966 in USA/Europe; 660,251 in All Other Countries). The number of observations for birds and insects are relatively similar, and thus supports the survey response that users generally record observations for any species, regardless of visual appeal and never avoid recording observations for unattractive or scary taxa, as mentioned in Section 4.2. It is important to note, however, that insect taxa vastly outnumber bird taxa. Although the number of observations is similar between the two groups, insects are highly underrepresented, due to there being many more species to observe.

4.6 Findings on Research Questions

4.6.1 Findings on Research Question One

Research question one seeks the socio-economic factors that limit biodiversity data collection. Factors 2 (transport accessibility), 3 (access to technology), 5 (income), and 6 (experience) reflect socio-economic sub-variables.

4.6.1.1 Factor 2: Transport Accessibility

Factor 2 is a socio-economic factor of transport accessibility. 83% of USA/Europe and 100% of All Other Countries reported that they record new observations on both weekdays and weekends. Although the majority of both groups strongly agreed that they have access to a variety of green spaces, 8.8% of USA/Europe users disagreed while 3% of All Other Countries disagreed with this statement. In addition, 51% of the USA/Europe group responded that they sometimes are able to travel to areas of interest, while 44% of All Other Countries said that they could frequently travel to areas of interest. These responses are interesting because respondents from All Other Countries responded that they have access to a variety of green spaces and are frequently able to travel to places of interest, while being in more data-deficient areas.

4.6.1.2 Factor 3: Access to Technology

Factor three addresses phone and internet access. In this case, USA/Europe and All Other Countries had similar responses. Over 90% of both groups reported that they never experienced gaps because they are unable to pay. Likewise, 63% of USA/Europe and 56% of All Other Countries sometimes experience gaps in phone/internet service because of service or other issues. These similar responses indicate that access to technology affects data collection similarly for both groups.

4.6.1.3 Factor 5: Income

Factor five is the socio-economic factor of income. As stated earlier, the majority of respondents identified themselves as in the average middle class (47% of USA/Europe; 59% of All Other Countries). Only 10% of USA/Europe respondents and 13% of All Other Countries identified as lower class/poor. This indicates that income is a barrier to participating on iNaturalist. In addition, 41% of USA/Europe and 38% of All Other Countries agreed that they would record more observations if they had more specialized equipment. Assuming that the only barrier to obtaining specialized equipment is cost, then income can be considered a limiting factor of data collection.

4.6.1.4 Factor 6: Experience

Factor six considers the experience of iNaturalist users, based on how long and how they've been using the platform. As highlighted earlier, most users have been using the platform between 1-5 years. The survey also asked whether interacting on the forum has improved species observation and/or identification skills. The majority of both groups indicated that they agree with this statement (31% in USA/Europe; 34% All Other Countries) but also had similar proportions neutral on the question (30% of USA/Europe; 25% of All Other Countries). Interestingly, 20% of USA/Europe and 16% of All Other Countries disagreed or strongly disagreed that interacting on the forum improved species observation and/or identification skills. This could be explained by different uses of the forum, or in other words, not all users need help improving their skills, or they use the forum for different topics.

4.6.2 Findings on Research Question Two

Research question two seeks to explore the socio-cultural factors that limit biodiversity data collection. Factors 1 (motivation) and 4 (time accessibility) correspond to socio-cultural sub-variables.

4.6.2.1 Factor 1: Motivation

Factor 1 includes questions that aim to indicate the user's level of motivation. Here, 46% of USA/Europe were neutral while 44% of All other Countries agreed that the desired goals of projects on iNaturalist are achievable. High proportions of both groups (>40%) agreed that the outcome of projects provide value to society. Lastly, the majority of both groups (>40%) were neutral about feeling supported/motivated by project creators/administrators. The high levels of neutrality indicate that interactions or experiences on projects are either still in infancy or that goals/outcomes/interaction with project administrators aren't a key feature of projects.

4.6.2.2 Factor 4: Time Accessibility

Factor four corresponds to the sub-variable of time availability. The expectation being that more free time would increase the opportunity to record observations, and thus more data would be collected. Not surprisingly, 71% of USA/Europe and 63% of All Other Countries agreed or agreed strongly that they would record more observations if they had more free time. This correlated somewhat to employment status. 41% of USA/Europe and 25% of All Other

Countries work full time. Interestingly, another 19% of users in All Other Countries are unemployed. Regardless, there are many other lifestyle and factors that occupy peoples time, besides employment. This factor, however, seemed to affect the global population similarly i.e. time is a limiting factor for data collection.

4.6.3 Findings on Research Question 3

Research question three seeks to explore the platform characteristics that limit biodiversity data collection. The results of the factor analysis, however, did not identify any clear platform related components such as platform or project design. As reflected upon earlier, users expressed much neutrality on questions relating to projects or platform characteristics. This may be due to different circumstances such as preferential use of the platform (126 users said they use iNaturalist to record observations for their own use/pleasure). Although they may contribute to projects, it is not the main objective of the user, and the opinions reflect that. Also, projects don't always have specific goals. There is one project on iNaturalist, for example, for people of the Lesbian, Gay, Bisexual, Transgender, and Queer (LGBTQ+) community to share any observations they wish and connect with each other, much like a social networking site. This project does not specify any guidelines and serves simply as a community resource. This would also explain some neutrality in terms of the goals, outcomes, and experience with project owners/administrators.

4.7 Discussion

This study was conducted to explore barriers to global biodiversity data collection on the citizen science platform, iNaturalist. A survey measured each sub-variable in the operationalization table by a set of questions. The survey yielded 149 responses, resulting in a confidence interval of 8%[4]. Although this exceeded the target sample size, responsivity to the survey was low in areas with low observations (e.g. India). The majority of respondents were from areas with high observation rates. A more evenly distributed sample, targeting areas with low observations, would better suit the research objectives of finding limitations of biodiversity data collection. As mentioned previously, this would require improving functionality of citizen science platforms to include the ability to contact users directly. This way, a purposive and more representative sample could be drawn. A factor analysis then identified six components, four socio-economic and two socio-cultural, that explain significant variance in the dataset. Factors relating to iNaturalist platform characteristics were not found to be significant in the factor analysis.

The factor analysis identified four socio-economic factors that limit data collection on iNaturalist: income, experience, access to technology, and transportation accessibility. First, survey results also showed that the majority of respondents belong to the average and upper middle class. Only 11% of respondents claimed belonging to the lower class/poor class, despite these classes comprising over 61% of the global population (Kochhar, 2021). This, along with the fact that over 60% of respondents answered that they would record more observations if they had more specialized equipment, indicates that income is a limiting factor.

Experience of users in this analysis was measured by the number of years with iNaturalist and whether they felt that using the forum had improved their observation and/or identification skills. It is worth pointing out that there are only nine forum topics on iNaturalist, none of which explicitly focus on observations or identification of species. In light of this, iNaturalist

[4] Calculated using: https://www.surveymonkey.com/mp/margin-of-error-calculator/

could expand its forum to include more topics. This gives more opportunity for the platform to motivate, educate, and engage users and therefore, increase observations.

Over half of USA/Europe and All Other Countries sometimes experience gaps in phone/internet service because of service or other issues. This presents a factor that may contribute to spatial or temporal data gaps, when users are unable to make observations. Citizen science platforms have little if not any control over this issue, however, it helps to identify this accessibility issue in order to understand causes of data gaps or devise plans to work around them.

Accessibility to parks is also an indicator on the City Biodiversity Index. Although phrased as “green spaces” in the survey, these indicators both measure mobility and the variety of green places a user has access to. Access to green spaces can be more crucial for the urban environment, where access to biologically diverse areas can be less ubiquitous than in some rural areas. Considering 50% of survey respondents were from urban areas, this presents an important factor in data collection. The responses pertaining to accessibility of green spaces and ability to travel were also unexpected. Put simply, respondents claim to have less mobility and green space access in areas with more observations. This appears contradictory, yet worth exploring further with more open questions in order to determine the validity and context of such responses.

Survey results also showed that the proportion of sample population increased as levels of education increased. This indicates that the platform attracts more highly educated individuals. Also, education is an indicator in the City Biodiversity Index. This supports that education is important to promoting biodiversity data collection and urban biodiversity. The availability of education, therefore, could be considered as a limitation for future studies.

Motivation and time accessibility were identified by the factor analysis as socio-cultural factors that limit data collection on iNaturalist. Sub-variables corresponding to personal values or attitudes were not recognized as factors limiting data collection. Dos Santos et al. (2020) has identified taxonomic bias caused by species charisma/aesthetic appeal as a source of data gaps. As identified in Section 4.2, however, respondents claimed to have little taxonomic bias in terms of attractiveness. As discussed in Section 4.5, iNaturalist data (Appendix 2) indeed showed that respondents from both groups (USA/Europe and All Other Countries) recorded a similar number of annual observations for Aves and Insecta classes. This illustrates that users were not avoiding species for the way that they look. It is important to note that the sample population were users of the forum, and likely more enthusiastic and specialized respondents, thus presenting sample bias. A larger sample population including those that do not use the forum on iNaturalist may exhibit different preferences. Nevertheless, these results indicate that species charisma is not a strong limitation of data collection, especially among more active and enthusiastic individuals.

Free time was another socio-cultural factor related to data collection. As previously mentioned, the majority of survey respondents said that they would record more if they had more free time. This may be helpful for researchers to know, especially for organized scientific studies that use citizen scientists. Awareness of time constraints of observers allows for careful design of research protocol and methodology.

Platform characteristics were considered as a possible limitation to biodiversity data collection. Platform design, project design, and project owner indicators, however, were not identified by the factor analysis as statistically significant. Still, the survey revealed interesting insights relating to the iNaturalist platform. The majority of survey respondents claimed to have at least some experience with projects yet remained neutral on many of the questions about projects. This indicates that relationships of users with projects are a limiting factor. Two observations aim to explain this neutrality. First, the survey failed to understand that some projects are designed without specific goals. This can explain some neutrality about goals or outcomes of the projects. In response, platforms such as iNaturalist could endeavour to promote or explain the benefits of projects with specific goals. This may increase participation and/or effectiveness of projects. Neutrality could also be explained by people not knowing about projects. The home page of iNaturalist briefly mentions projects but does not explain what they are or how to join them. Advertising projects on the platform may similarly increase participation and/or awareness.

In reflection, understanding limitations in respect to platform characteristics could be better understood by asking more detailed questions on awareness of projects as well as how people use the platform. It would be interesting to survey users who primarily contribute to projects, since users on the forum are largely opportunistic researchers interested in collecting observations for their own personal use.

5. Conclusion

The IPBES (2019) recognizes that biodiversity is declining faster now than ever before. This has dangerous implications for humankind, considering the associated ecosystem services that will continue to be negatively impacted by this crisis. Monitoring biodiversity is a cornerstone to protecting it, by enabling informed conservation strategies and understanding of the drivers of loss. Citizen science presents an important tool that helps scientists, government officials, and community action groups in gathering diverse biodiversity data that can be used in indices to evaluate changes. This can especially be helpful to assess the effectiveness or failures of different policies or interventions that aim to address biodiversity loss.

The purpose of this study was to explore various factors that limit biodiversity data collection. To do this, data was collected by a survey questionnaire and secondary data available on the GBIF. The survey was distributed on the forum of iNaturalist, a popular citizen scientist platform used by millions of users worldwide to record species observations. The purpose of studying limitations to data collection is to address and reduce the taxonomic, spatial, and temporal biodiversity data gaps identified in academic literature. This would lead to more accurate biodiversity measures and ecosystem assessments and therefore more targeted policies and strategies. Quantitative survey results were then analysed using descriptive and inferential statistics to reveal the most relevant factors that answer the research questions.

In spite of the difficulties of contacting citizen scientists directly to participate in a survey, it was possible to utilize the forum on iNaturalist to reach a large community. As discussed, this presented its own disadvantages such as a voluntary sample and respondent bias. One important recommendation in light of this study would be for citizen science platforms to increase communication functionality to carry out scientific studies involving its users. This would allow for more purposive samples and would allow for more in-depth questioning that reflects the users local cultural and economic environment. This would also allow researchers to target areas with low data collection, and where addressing data gaps may be more constructive. With this in mind, the following answers to the research questions are provided as a preliminary analysis of iNaturalist users and can be considered a framework or starting point for future research.

The first research question asked which socio-economic factors limited citizen science biodiversity data collection. Amano (2016) suggested that wealth, experience, and inaccessibility due to location were possible factors of spatial information gaps. Indeed, the factor analysis and reflection identified income, transportation accessibility, and experience as relevant factors. In addition, education and access to technology were also identified. These could, however, be considered as components of experience and income, respectively. In response, to reduce data gaps, it may be prudent to study and expand access to transportation, education, technology, and green spaces while promoting income equality especially in areas with observation deficiencies.

The second research question asked which socio-cultural factors limited citizen science biodiversity data collection. The factor analysis identified motivation and free time as limiting factors of data collection. Although citizen science platforms have little agency over how much free time users have, they could endeavour to enhance motivation by advertising local achievements and promoting projects with beneficial outcomes.

Dos Santos et al. (2020) identified that species charisma, or aesthetic appeal, may influence observation research. The survey therefore included visual appeal as a possible source of taxonomic bias. Surprisingly, survey respondents claimed that they record observations for any species they can, regardless of visual appeal. A study by Callaghan, Poore, Hofmann, Roberts & Pereira (2021) found no evidence that colorful species of birds were over-represented in unstructured citizen science (such as iNaturalist) data. Instead, large-bodied, common, and species in large groups were found with strong or moderate evidence of over-representation. This agreement among studies supports that charisma is less responsible for taxonomic data gaps. Troudet et al., (2017) instead hypothesised that societal preferences were a major factor in taxonomic bias, and therefore a cause of taxonomic data gaps. Further research should aim to study the extent of other public preferences and may help to address the causes of taxonomic bias.

Finally, the study included some questions on platform characteristics such as platform design, project design, and project owners, but did not identify any that may limit biodiversity data collection. It is worth reflecting that the questions asked were a small set to gain insight on possible limitations. A more in-depth survey and analysis may reveal different results. In terms of reducing biodiversity data gaps, citizen science platforms may consider increasing public awareness, especially among underrepresented taxa. This can inform users about data gaps while encouraging users to bridge them. Citizen science platforms may also improve user awareness of functionality, including the availability and purpose of projects, since many users expressed indifference towards them.

The methodology for this research and subsequent results seek to provide a jumping off point for future research, where more targeted questions may be asked in order to understand the factors in more detail. Understanding the causes of taxonomic, spatial, and temporal biodiversity data gaps is the first step to overcoming them. Better data quality will improve the reliability and effectiveness of conservation strategies, thus protecting biodiversity and the ecosystem services we all depend on.

Bibliography

Amano, T., Lamming, J. D. L., & Sutherland, W. J. (2016). Spatial Gaps in Global Biodiversity Information and the Role of Citizen Science. *BioScience*, *66*(5), 393–400. https://doi.org/10.1093/biosci/biw022

Aronson, M. F. J., Lepczyk, C. A., Evans, K. L., Goddard, M. A., Lerman, S. B., MacIvor, J. S., … Vargo, T. (2017). Biodiversity in the city: key challenges for urban green space management. *Frontiers in Ecology and the Environment*, Vol. 15, pp. 189–196. Wiley Blackwell. https://doi.org/10.1002/fee.1480

Balázs, B., Mooney, P., Nováková, E., Bastin, L., & Arsanjani, J. J. (2021). Data quality in citizen science. *The science of citizen science* (pp. 139-157) Springer Nature. doi:10.1007/978-3-030-58278-4

Braat, L. C., & de Groot, R. (2012). The ecosystem services agenda:bridging the worlds of natural science and economics, conservation and development, and public and private policy. *Ecosystem Services*, Vol. 1, pp. 4–15. https://doi.org/10.1016/j.ecoser.2012.07.011

Callaghan, C. T., Ozeroff, I., Hitchcock, C., & Chandler, M. (2020). Capitalizing on opportunistic citizen science data to monitor urban biodiversity: A multi-taxa framework. *Biological Conservation*, *251*. https://doi.org/10.1016/j.biocon.2020.108753

Ceballos, G., Ehrlich, P. R., & Raven, P. H. (2020). Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proceedings of the National Academy of Sciences*, *117*(24). https://doi.org/10.1073/pnas.1922686117/-/DCSupplemental

Ceccaroni, L., Bibby, J., Roger, E., Flemons, P., Michael, K., Fagan, L., & Oliver, J. L. (2019). Opportunities and Risks for Citizen Science in the Age of Artificial Intelligence. *Citizen Science: Theory and Practice*, *4*(1). https://doi.org/10.5334/cstp.241

Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., … Turak, E. (2016). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, *213*, 280–294. https://doi.org/10.1016/j.biocon.2016.09.004

Costanza, R., de Groot, R., Farberll, S., Grassot, M., Hannon, B., Limburg, K., … Suttonllll, P. (1997). *The value of the world's ecosystem services and natural capital*.

Czúcz, B., Arany, I., Potschin-Young, M., Bereczki, K., Kertész, M., Kiss, M., … Haines-Young, R. (2018). Where concepts meet the real world: A systematic review of ecosystem service indicators and their classification using CICES. *Ecosystem Services*, Vol. 29, pp. 145–157. Elsevier B.V. https://doi.org/10.1016/j.ecoser.2017.11.018

Daily, G. (1997). *Nature's services: societal dependence on natural ecosystems*. Washington DC.

dos Santos, J. W., Correia, R. A., Malhado, A. C. M., Campos-Silva, J. v., Teles, D., Jepson, P., & Ladle, R. J. (2020). Drivers of taxonomic bias in conservation research: a global analysis of terrestrial mammals. *Animal Conservation*, *23*(6), 679–688. https://doi.org/10.1111/acv.12586

Duelli, P., & Obrist, M. K. (2003). Biodiversity indicators: The choice of values and measures. *Agriculture, Ecosystems and Environment*, *98*(1–3), 87–98. https://doi.org/10.1016/S0167-8809(03)00072-0

Faith, D. P., Collen, B., Ariño, A. H., Koleff, P., Guinotte, J., Kerr, J., & Chavan, V. (2013). Bridging biodiversity data gaps: Recommendations to meet users' data needs. In *Biodiversity Informatics* (Vol. 8). Retrieved from http://data.gbif.org/

Fedor, P., & Zvaríková, M. (2019). Conservation ecology. In B. Fath (Ed.), *Encyclopedia of ecology* (pp. 337-346) Oliver Walter.

Frigerio, D., Richter, A., Per, E., Pruse, B., & Vohland, K. (2021). Citizen science in the natural sciences. *The science of citizen science* (pp. 79-96) Springer Nature. doi:10.1007/978-3-030-58278-4

Gan, J. (2021). Urban biodiversity and built environment: Case study of Shanghai. In *Urban Biodiversity and Built Environment* (pp. 1–17). Springer Singapore. https://doi.org/10.1007/978-981-16-0949-7_1

Gaston, K. J., & Spicer, J. I. (2004). *Biodiversity an introduction* (2nd ed.) Blackwell Publishing.

GBIF.org (2022) *Occurrence Map*. Available from https://www.gbif.org/occurrence/map [July 2022]

Haase, D., Larondelle, N., Andersson, E., Artmann, M., Borgström, S., Breuste, J., … Elmqvist, T. (2014). A quantitative review of urban ecosystem service assessments: Concepts, models, and implementation. *Ambio*, Vol. 43, pp. 413–433. Kluwer Academic Publishers. https://doi.org/10.1007/s13280-014-0504-0

iNaturalist contributors, iNaturalist (2022). iNaturalist Research-grade Observations. iNaturalist.org. Occurrence dataset https://doi.org/10.15468/ab3s5x accessed via GBIF.org on 2022-08-03.

iNaturalist site stats. (2022). Retrieved from https://www.inaturalist.org/stats

IPBES. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (S. Díaz, J. Settele, E. Brondízio, H. T. Ngo, M. Guèze, J. Agard, … C. Zayas, Eds.). Bonn, Germany: IPBES secretariat. https://doi.org/10.5281/zenodo.3553579

Kelling, S., Johnston, A., Bonn, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., … Guralnick, R. (2019). Using Semistructured Surveys to Improve Citizen Science Data for Monitoring Biodiversity. *BioScience*, *69*(3), 170–179. https://doi.org/10.1093/biosci/biz010

Kobori, H., Dickinson, J. L., Washitani, I., Sakurai, R., Amano, T., Komatsu, N., … Miller-Rushing, A. J. (2016). Citizen science: a new approach to advance ecology, education, and conservation. *Ecological Research*, *31*(1), 1–19. https://doi.org/10.1007/s11284-015-1314-y

Kochhar, R. (2021). *Are you in the global middle class? find out with our income calculator* Pew Research Center.

Kolbert, E. (2014). *The sixth extinction*. Great Britain: Bloomsbury Publishing plc.

Kowarik, I. (2011). Novel urban ecosystems, biodiversity, and conservation. *Environmental Pollution, 159*(8), 1974-1983. doi:10.1016/j.envpol.2011.02.022

Liu, H., Dörler, D., Heigl, F., & Grossberndt, S. (2021). Citizen science platforms. *The science of citizen science* (pp. 439-459) Springer Nature. doi:10.1007/978-3-030-58278-4

Millennium Ecosystem Assessment (MEA). (2005). *Ecosystems and Human Well-being: Biodiversity Synthesis*. Washington DC.

Nolan, K. A., & Callahan, J. E. (2006). Beachcomber biology: The shannon-weiner species diversity index. In M. A. O'Donnell (Ed.), *Tested studies for laboratory teaching* (pp. 334-338) Proceedings of the 27th Workshop/Conference of the Association for Biology Laboratory Education (ABLE).

Njue, N., Stenfert Kroese, J., Gräf, J., Jacobs, S. R., Weeser, B., Breuer, L., & Rufino, M. C. (2019). Citizen science in hydrological monitoring and ecosystem services management: State of the art and future prospects. *Science of the Total Environment*, Vol. 693. Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2019.07.337

Seto, K. C., Güneralp, B., & Hutyra, L. R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(40), 16083–16088. https://doi.org/10.1073/pnas.1211658109

Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., … Parrish, J. K. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation*, *181*, 236–244. https://doi.org/10.1016/j.biocon.2014.10.021

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-09084-6

Uchida, K., Blakey, R. v., Burger, J. R., Cooper, D. S., Niesner, C. A., & Blumstein, D. T. (2021). Urban Biodiversity and the Importance of Scale. *Trends in Ecology and Evolution*, Vol. 36, pp. 123–131. Elsevier Ltd. https://doi.org/10.1016/j.tree.2020.10.011

United Nations. (2012). *World Urbanization Prospects*.

Van Noordwijk, Toos (C. G. E.), Bishop, I., Staunton-Lamb, S., Oldfield, A., Loiselle, S., Geoghegan, H., & Ceccaroni, L. (2021). Creating positive environmental impact through citizen science. *The science of citizen science* (pp. 373-395) Springer Nature. doi:10.1007/978-3-030-58278-4

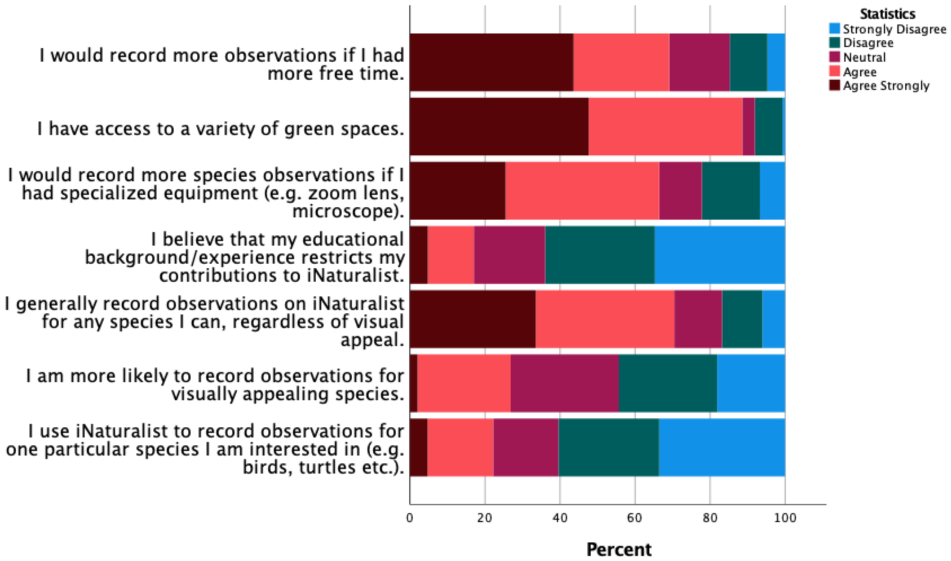
Van Thiel, S. (2014). *Research methods in public administration and public management* (1st ed.). London: Routledge. doi:https://doi.org/10.4324/9780203078525

Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., … Wagenknecht, K. (2021). *The Science of Citizen Science*.

Yesson, C., Brewer, P. W., Sutton, T., Caithness, N., Pahwa, J. S., Burgess, M., … Culham, A. (2007). How global is the global biodiversity information facility? *PLoS ONE*, *2*(11). https://doi.org/10.1371/journal.pone.0001124

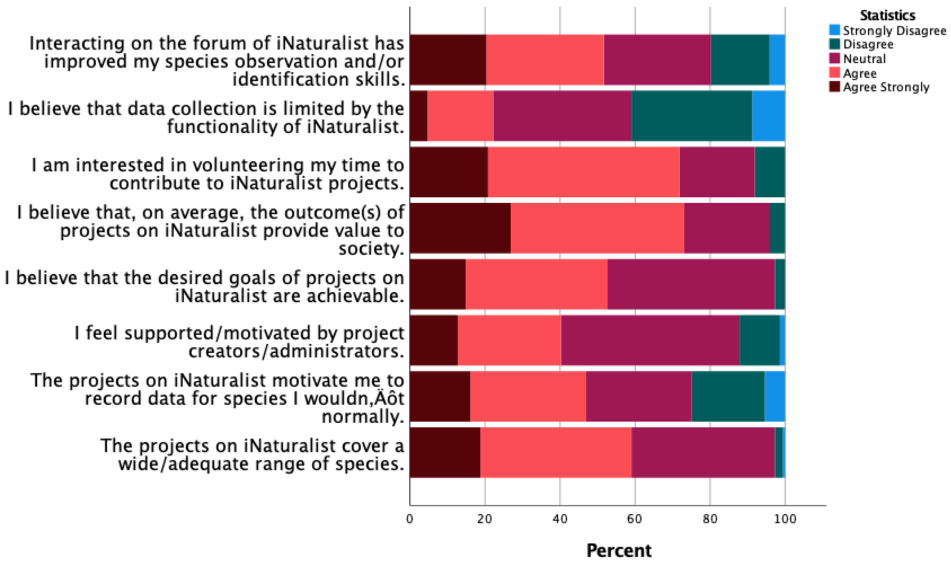
Appendix 1: Likert-Scale Survey Results

Figure 8: Level of Agreement Questions 1-7.



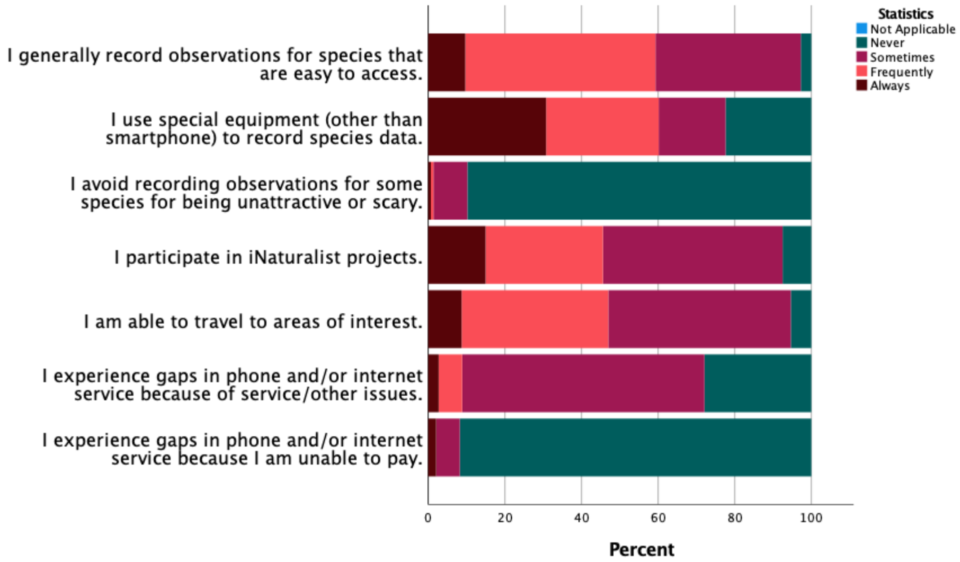
Source: Author (2022)

Figure 9: Level of Agreement Questions 8-15.



Source: Author (2022)

Figure 10: Frequency of Agreement.



Source: Author (2022)

Appendix 2: iNaturalist Observation Data

Table 7: Number of species occurrences among 24 taxonomic classes in USA/Europe group from July 2021 to June 2022.

Taxonomic Class	July-Sept ('21)	Oct-Dec ('21)	Jan-Mar ('22)	Apr-Jun ('22)	Total Observations/ Class
Aves	283,580	306,610	407,560	476156	1,473,906
Liliopsida	90,413	37,325	49,885	176094	353,717
Mammalia	53,379	40,089	43,896	65259	202,623
Actinopterygii	20,849	11,255	10,243	21,158	63,505
Amphibia	39,702	19,483	25,838	42217	127,240
Pinopsida	16,154	13,329	11,862	15,709	57,054
Globothalamea	21	1	7	9	38
Magnoliopsida	724,172	288,432	293,201	902,679	2,208,484
Reptilia	61,359	29,517	28,911	99705	219,492
Polypodiopsida	22,601	16,286	14,402	29873	83,162
Jungermanniopsida	613	688	1,210	854	3365
Florideophyceae	794	344	383	446	1967
Bryopsida	3,793	5,212	0	0	9,005
Anthozoa	2,964	2,161	2,583	2950	10,658
Polychaeta	363	493	427	428	1711
Lecanoromycetes	5523	6,925	9936	8074	30458
Bivalvia	6169	4,742	4232	5102	20245
Maxillopoda	639	391	577	780	2387
Bacillariophyceae	45	37	72	188	342
Malacostraca	13,295	7,954	9088	16238	46,575
Agaricomycetes	58,039	67,491	20400	22,375	168,305
Gastropoda	22,226	16,220	15066	26,184	79,696
Arachnida	56,937	23,263	11309	44,723	136,232
Insecta	872,226	226,299	117,070	625,371	1,840,966
Total	2,355,856	1,124,547	1,078,158	2,582,572	7,141,133

Source: iNaturalist contributors, iNaturalist (2022).

Table 8: Number of species occurrences among 24 taxonomic classes among All Other Countries group from July 2021 to June 2022.

Taxonomic Class	July-Sept ('21)	Oct-Dec ('21)	Jan-Mar ('22)	Ap-Jun ('22)	Total Observations/ Class
Aves	171,465	153,532	162,664	277,632	765,293
Liliopsida	88,036	45,674	19,799	65,026	218,535
Mammalia	21,151	16,070	15,176	24,877	77,274
Actinopterygii	8,204	9,106	10,092	10,920	38,322
Amphibia	15,335	8,815	6,594	12,967	43,711
Pinopsida	13,104	7,308	6,105	11,899	38,416
Globothalamea	0	0	3	1	4
Magnoliopsida	511,528	180,717	86,811	347,427	1,126,483
Reptilia	15,578	17,289	11,397	19,384	63,648
Polypodiopsida	22,143	11,471	6,966	18,737	59,317
Jungermanniopsida	539	562	508	991	2,600
Florideophyceae	271	187	230	514	1,202
Bryopsida	4,180	3,073	3,072	5,304	15,629
Anthozoa	959	1,251	1,192	2,174	5,576
Polychaeta	165	186	148	254	753
Lecanoromycetes	4,002	3,201	2,486	4,888	14,577
Bivalvia	2,709	1,618	1,456	2,894	8,677
Maxillopoda	265	174	240	446	1,125
Bacillariophyceae	25	7	12	34	78
Malacostraca	3,376	2,308	2,825	4,770	13,279
Agaricomycetes	19,793	16,212	5,887	19,819	61,711
Gastropoda	9,304	9,774	8,603	13,340	41,021
Arachnida	17,879	12,923	11,476	13,708	55,986
Insecta	267,039	120,876	105,014	167,322	660,251
Total	1,197,050	622,334	468,756	1,025,328	3,313,468

Source: iNaturalist contributors, iNaturalist (2022).

Appendix 3: Survey Questionnaire

iNaturalist Survey

Start of Block: Introduction

Q21 Thank you for taking the time to participate in this survey about global biodiversity data gaps. The survey will ask you about your experience with and opinions of the iNaturalist platform.

This survey is being conducted as thesis research for a Master's of Science degree in Urban Management and Development at the Institute for Housing and Urban Development Studies at Erasmus University Rotterdam.

This survey has one section on background information, two sections on experience, and one demographic section. We expect this survey to take about 10 minutes to complete. Thank you for your time and contribution!

End of Block: Introduction

Start of Block: User Profile

Q20 Do you consent to participate in this survey?

- ○ Yes (1)
- ○ No (2)

Q1 How long have you been using iNaturalist?

- ○ Less than 1 year (1)
- ○ 1-5 years (2)
- ○ 5-10 years (3)
- ○ 10+ years (4)

Q2 How often do you record new observations on iNaturalist?

- Daily (1)
- Weekly (2)
- Monthly (3)
- Varies/Not Sure (4)

Q3 Why do you use iNaturalist? (select all that apply)

- To contribute to scientific projects/research (1)
- To publish my own projects and collect data from other users (7)
- To record species for my own use/pleasure (2)
- It provides a recreational activity (4)
- It provides an educational activity (5)
- To contribute to positive environmental changes (e.g. providing evidence for policy and conservation) (6)
- Social networking (8)
- To promote/engage in community action (9)
- To increase my scientific literacy (10)

Q4 When do you usually record new observations?

- ◯ Weekdays (1)
- ◯ Weekends (2)
- ◯ Both (3)

End of Block: User Profile

Start of Block: Agreement Questions

Q5 Please select your the level of agreement with the following statements.

	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Agree Strongly (5)
I use iNaturalist to record observations for one particular species I'm interested in (e.g. birds, turtles etc.). (1)	○	○	○	○	○
I'm more likely to record observations for visually appealing species. (2)	○	○	○	○	○
I generally record observations on iNaturalist for any species I can, regardless of visual appeal. (3)	○	○	○	○	○
I believe that my educational background/experience restricts my contributions to iNaturalist. (4)	○	○	○	○	○
I would record more species observations if I had specialized equipment (e.g. zoom lens, microscope). (5)	○	○	○	○	○
I have access to a variety of green spaces. (6)	○	○	○	○	○
I would record more observations if I had more free time. (7)	○	○	○	○	○
The projects on iNaturalist cover a wide/adequate range of species. (8)	○	○	○	○	○
The projects on iNaturalist motivate me to record data for species I wouldn't normally. (10)	○	○	○	○	○

I feel supported/motivated by project creators/administrators. (11)	○	○	○	○	○
I believe that the desired goals of projects on iNaturalist are achievable. (12)	○	○	○	○	○
I believe that, on average, the outcome(s) of projects on iNaturalist provide value to society. (13)	○	○	○	○	○
I am interested in volunteering my time to contribute to iNaturalist projects. (15)	○	○	○	○	○
I believe that data collection is limited by the functionality of iNaturalist. (16)	○	○	○	○	○
Interacting on the forum of iNaturalist has improved my species observation and/or identification skills. (17)	○	○	○	○	○

End of Block: Agreement Questions

Start of Block: Frequency Questions

Q6 Please select your experience with the following statements.

	Never (1)	Sometimes (2)	Frequently (3)	Always (4)	Not Applicable (5)
I experience gaps in phone and/or internet service because I’m unable to pay. (1)	○	○	○	○	○
I experience gaps in phone and/or internet service because of service/other issues. (7)	○	○	○	○	○
I am able to travel to areas of interest. (2)	○	○	○	○	○
I participate in iNaturalist projects. (3)	○	○	○	○	○
I avoid recording observations for some species for being unattractive or scary. (4)	○	○	○	○	○
I use special equipment (other than smartphone) to record species data. (5)	○	○	○	○	○
I generally record observations for species that are easy to access. (6)	○	○	○	○	○

End of Block: Frequency Questions

Start of Block: Demographic Questions

Q7 In which country do you currently reside?

▼ Afghanistan (1) ... Zimbabwe (1357)

Q8 In what city/region do you usually record observations for iNaturalist? (please write in the text box below)

____

Q9 What type of area do you currently live in?

- ◯ Rural (1)
- ◯ Peri-Urban (rural-urban transition zone) (2)
- ◯ Urban (3)
- ◯ Not Sure (4)

Q10 What is your Age?

- ◯ 18 or younger (1)
- ◯ 19-29 (2)
- ◯ 30-65 (3)
- ◯ 66+ (4)

Q11 What is your current socioeconomic status?

- Lower Class/Poor (1)
- Average Middle Class (4)
- Upper Middle Class (5)
- Upper Class (2)
- Not Sure/Prefer to not answer (3)

Q12 What is your highest level of education?

- Less than High School (5)
- High School Degree (1)
- Undergraduate Degree (2)
- Graduate Degree (3)
- Doctoral Degree (4)
- Prefer to not answer (6)

Q23 How do you describe yourself?

- Male (1)
- Female (2)
- Non-binary / third gender (3)
- Prefer to self-describe (4) ____
- Prefer not to say (5)

Q25 What best describes your employment status over the last three months?

- Working full-time (1)
- Working part-time (2)
- Unemployed and looking for work (3)
- A homemaker or stay-at-home parent (4)
- Student (5)
- Retired (6)
- Other (7)

End of Block: Demographic Questions

Appendix 4: IHS copyright form

In order to allow the IHS Research Committee to select and publish the best UMD theses, students need to sign and hand in this copyright form to the course bureau together with their final thesis.

By signing this form, you agree that you are the sole author(s) of the work and that you have the right to transfer copyright to IHS, except for those items clearly cited or quoted in your work.

Criteria for publishing:

1. A summary of 400 words must be included in the thesis.
2. The number of pages for the thesis does not exceed the maximum word count.
3. The thesis is edited for English.

Please consider the length restrictions for the thesis. The Research Committee may elect not to publish very long and/or poorly written theses.

I grant IHS, or its successors, all copyright to the work listed above, so that IHS may publish the work in the IHS Thesis Series, on the IHS web site, in an electronic publication or in any other medium.

IHS is granted the right to approve reprinting.

The author retains the rights to create derivative works and to distribute the work cited above within the institution that employs the author.

Please note that IHS copyrighted material from the IHS Thesis Series may be reproduced, up to ten copies for educational (excluding course packs purchased by students), non-commercial purposes, provided a full acknowledgement and a copyright notice appear on all reproductions.

Thank you for your contribution to IHS.

Date : 7 August 2022

Your Name(s) : Alicia DeMaio

Your Signature(s) : Alicia DeMaio

Please direct this form and all questions regarding this form or IHS copyright policy to:

Academic Director Burg. Oudlaan 50, T-Building 14th floor, 3062 PA Rotterdam, The Netherlands	gerrits@Ihs.nl Tel. +31 10 4089825