IDENTIFYING, EXTRACTING, AND GEOREFERENCING TOPONYMS FROM A NINETEENTH-CENTURY TEXT

by

Eric Dorman

A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Geographic and Cartographic Science

Committee:

_____	Dr. Matthew Rice, Thesis Chair
_____	Dr. Maction Komwa, Committee Member
_____	Dr. Arie Croitoru, Committee Member
_____	Dr. Dieter Pfoser, Department Chairperson
_____	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
_____	Dr. Fernando R. Miralles-Wilhelm, Dean, College of Science
Date: _____	Summer Semester 2020 George Mason University Fairfax, VA

Identifying, Extracting, and Georeferencing Toponyms from a Nineteenth-Century Text

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Eric Dorman
Bachelor of Science
University of Mary Washington, 2018

Director: Matt Rice, Associate Professor
Department of Geography and Geoinformation Science

Summer Semester 2020
George Mason University
Fairfax, VA

DEDICATION

I dedicate this thesis to my family, who supported me throughout this journey

ACKNOWLEDGEMENTS

I would like to thank the many friends, relatives, and supporters who have made this happen. I especially would like to thank the faculty of the University of Mary Washington and George Mason University who helped me along this GIS journey, namely: Brian Rizzo, Matthew Rice, and Dieter Pfoser.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GIS ----------Geographic Information Systems
GEOINT----------Geointelligence
TGN ----------Thesaurus of Geographic Names
NGA ----------National Geospatial Agency
GNS ----------GEOnet NAMES Server
US BGN ----------United States Board of Geographic Names
NER----------Named Entity Recognition

ABSTRACT

IDENTIFYING, EXTRACTING, AND GEOREFERENCING TOPONYMS FROM A NINETEENTH-CENTURY TEXT

Eric Dorman, M.S.

George Mason University, 2020

Thesis Director: Dr. Matt Rice

This thesis explores the process of recreating and mapping a journey from an 1863 work of realistic fiction, Five Weeks in a Balloon, by Jules Verne. As with many of Verne's works, the novel is a geographically-rich story set in 19th-century Africa, with locations provided through coordinate-based metric georeferences, relative georeferences, and place names, many of which are outdated or archaic. This work answers the question: Is it possible to map a journey documented with antiquated place names, and if not, what should be done to improve the system? Using three different published editions of Fives Weeks in a Balloon as a source for place names, as well as multiple digital gazetteer resources, and period-specific maps; this research demonstrates that it is possible to reconstruct the literary journey in a geographic information system (GIS). Where the journey cannot be reconstructed or only partially so, this thesis discusses ways to improve the method. In the end, charting journeys with obsolete place names is possible, but current gazetteers struggle to resolve all of the place names

contained in the text, underscoring the need for temporal gazetteer development, focusing on antiquated place names and abandoned locations. Future development of historical gazetteers will be useful for research in the humanities and social sciences, where historical source documents are often used.

CHAPTER 1: INTRODUCTION

In the world of geographic information science (GIS), georeferencing is the process of taking geographic information, particularly place names and anything associated with it, and deducing its location on a map (Kemp, 2008). In recent years, methods and even algorithms have been developed for extracting geographic information from unstructured text, which is text that lacks indexing or metadata. Simply put, unstructured text is usually common written or typed text. Almost all the text in this thesis is unstructured text. Unstructured text is found in everything from tweets to novels to cuneiform tablets. In much of the business world, the study of, often unwittingly, crowdfunded masses of data, called data mining, has become extremely important as a means to study trends. One of the most difficult yet rewarding forms of data mining is extracting information from unstructured texts. The research presented in this thesis uses georeferencing of unstructured text to create new information, both tabular and geographic, of the trajectory associated with a novel. This geographic data mining activity is an important way to integrate archives of textual information with other media, and is described by authors such as Carroll (2006) as a challenging, but critical activity.

1.1 Extracting Geographic Information From Text

It has been determined by Carroll (2006) and others that developing methods of extracting place names and geographic information from unstructured texts and

georeferencing are excellent tools for the GIS community. The potential there is from extracting and mapping unstructured texts ranges from mapping novels to tracking targets of interest. Previous works, noted in the Literature Review to follow, demonstrate this potential through various feats and experiments by my predecessors. However, there is a difficult frontier in this field to be braved that may ultimately require cross disciplinary work with the linguistic, historical, and archeological communities. Georeferencing modern locations is relatively easy due to most tools being familiar with the names of modern places. However, as we go back in time to older texts, we come across places that have either long been abandoned and forgotten by humans or places that, for linguistic or cultural reasons, exist under different names then they do today; some just re-lettered slightly, e.g. *Peking* and *Beijing*, and others made unrecognizable. This relatively uncharted field in mapping literature, due to its particular conditions that differ it from the rest, requires unique and maybe even entirely new methods and tools for mining and georeferencing the unstructured text in these antiquated documents. This master's thesis is an attempt to make baby steps into this field that can become a foundation for works that builds on its findings. It shall also be a scout that discovers what tools and resources we lack to ultimately gain the ability to automate the mining and georeferencing of geographic, unstructured text from antiquated documents.

The goal of this thesis is to test the waters to see if it is possible to create a method for extracting and georeferencing geographic information from unstructured text that contains outdated place names for modern locations or names for places that have since been abandoned. The chief hurdle of this project is figuring out how to overcome

temporal changes of place names, which usually occur for linguistic, geopolitical, or cultural reasons. This thesis uses the Jules Verne novel *Five Weeks in a Balloon* as the main source material.

Names change and places, both towns and nations, come and go over time. What was once Byzantium, then Constantinople, is now Istanbul and where there was once the Roman Empire is now an array of dozens of nations with a variety of religions, languages, ethnicities, and cultures. These temporal changes to place names occur over time due to either linguistic or cultural reasons. Linguistic changes occur either due to evolution of the language of those who possess the place or because of the adaptation of a place name by another language to be more compatible with said language. In the latter instance whether or not the owners of the adopting language possess the place or not is irrelevant. For example, the Roman city of Londinium became Lundenwic and then London as possession of the city changed. In Turkish, a language belonging to a people who never possessed London, the city is called Londra. Cultural changes occur when possession of a place changes culture, the present culture changed radically, or because the original name of the place is so incompatible with the language of a foreign culture or outright unknown to it that said culture created a name for it. For example, the Dutch colony of New Amsterdam was taken by the British who then renamed it New York. When the Soviet Union replaced the Romanov Tsars in Russia, the new secular government carried out name changes both to purge the old ways and honor the new system, two famous examples being St. Petersburg becoming Leningrad and Tsaritsyn became Stalingrad. After the Soviet Union collapsed, Leningrad returned to being St.

Petersburg and Stalingrad became Volgograd. In ancient times, the Chinese, due to the limitations of ancient Mandarin, referred to the Roman Empire as “Daqin”, which meant “Great China”. Consequently, cultural changes to place names tend to be the most radical. Time changes places not just in name but in physical composition as well. Londinium was a Roman settlement, Ludenwic was an Anglo-Saxon trade hub, and modern London is the capital of Britain. Without undergoing a name change London has physically changed over the centuries, either due to immediate events like the Great Fire of 1666 or gradual ones like the Industrial Revolution.

These temporal changes present a complication in Named Entity Recognition (NER) and in mining geographic data from unstructured texts. Current GIS tools are designed for recognizing current names for current places. How then, shall we go about NER and geographic data mining of unstructured text? Ultimately, we will want to create a tool that can automate the process, but first we need to create a feasible method by hand, if we have the means to create it. This thesis, therefore, seeks to take these first steps; to attempt to manually create a method for georeferencing antiquated place names and determining what, if anything, needs to be made to improve the system.

Outside of GIS, this field holds considerable potential for contributing to fields such as archeology, urban planning, literature studies, and intelligence analysis. In archeology, the ability to georeferenced abandoned places from ancient texts can ease the ability to provide parameters for locating cultural sites. In urban planning, mapping antiquated texts can aid planners in understanding previous city layouts and knowing the locations of long gone buildings and businesses. In literature studies, such a tool can be

used for mapping ancient journeys. In intelligence analysis, a tool that can identify linguistic variations of a place name can help them analyze geographic intelligence in a little known language.

1.2 Jules Verne, Five Weeks in a Balloon, and the age of African Exploration

Jules Verne was born in 1828 in Nantes, France. Growing up with the Industrial Revolution occurring around him, Verne lived in an environment that fueled his imagination which in turn gave significant influence to his writings. Meanwhile beyond Europe, the Age of Exploration raged on. In particular, at this time, the exploration of the African continent had become the center of interest. Until recently, the interior of the continent was perceived to be unexplorable due to dangerous diseases, animals, and natives. But now, thanks to advances in medicine, logistics, and technology, explorers were making inroads. By the 1860s, Jules Verne was a struggling writer, living day to day on slim profits made off of short stories and plays. During this time, a major undertaking was underway in Africa. Two explorers, John Hanning Speke and Sir Richard Burton had returned from East Africa announcing the discovery of two great lakes in the region, Lake Victoria and Lake Tanganyika. The purpose of their journey, as were almost all expeditions in the region, was to discover the source of the Nile, which had eluded the European/Mediterranean world since Egyptian times. Explorers, Speke, who found Lake Victoria, and Burton, who found Lake Tanganyika, each contended that the lake they found was the source. And now, Speke had gone so far as to return to Africa with companion James Augustus Grant to acquire certain proof that Lake Victoria is the source. Consequently, public interest in Africa was high and Verne smelled opportunity.

Using his understanding of technology and the notes of Speke, Burton, and other explorers, Verne authored the novel *Five Weeks in a Balloon*.

Five Weeks in a Balloon became Jules Verne's breakthrough novel and the novel where he found his stride and theme as an author. First published in French in 1863 and then retranslated to English in 1869, the novel's success was a result of both riding a wave of public interest in African exploration and Verne discovering his writing formula. Like the novels to come, Verne discovered here the power of using technology and exploration to spice up the story, as well as the use of character trope combinations to provide awe and humor to the readers.

In this novel, a trio of explorers, consisting of a jack-of-all-trades scientist, his loyal assistant, and his initially skeptical hunter friend, seek to cross the entire African continent from East to West via balloon, linking the ranges of non-fictional exploration in the process. Demonstrating his technical understanding, Verne explained thoroughly how the balloon accomplished the journey. It was kept aloft by hydrogen gas that was expanded or contracted as needed through indirect heating by flame which was fueled by hydrogen gas made by electrolyzing water. It is then propelled by the trade winds. During this journey, the group casts off from Zanzibar, an island off the coast of modern Tanzania, confirms Lake Victoria as the source of the Nile and proceed through unexplored territory to West Africa, visiting major locations like Lake Chad and Timbuktu before reaching the French colony of St. Louis in modern Senegal.

1.3 Using Historical Accounts to Construct the Setting

For the setting, Jules Verne was heavily reliant on the notes of explorers who journeyed through the continent. At this time, the explored continent of Africa could be divided into four spheres: north, south, east, and west. The southern sphere consisted of people like David Livingston who, according to the book, recently found Lake Malawi while pushing into the interior from South Africa. The western sphere included the regions where Europeans engaged in trade of raw material, colonization, and slavery. It also includes the Niger River and the lands of the Ghana, Mali, and Songhai empires. Its most successful explorers are Henry Barth and Mungo Park. The northern sphere consists of people trying to reach the Nile's source by coming up the river itself. When the book was authored, the farthest-reaching explorer in that sphere was Andrea Debono, a Maltese merchant. The eastern sphere consists of explorers seeking to approach the Nile's source from the east coast in either modern Somalia or Tanzania. Speke and Burton are a part of this sphere. The goal of the explorers in *Five Weeks* is to link the eastern, northern, and western spheres. During the journey, the expeditions of Speke, Burton, Park, and Barth are the most heavily referenced sources for the setting. Their notes have been acquired for this research.

During their joint expedition, John Speke and Sir Richard Burton attempted to approach the Nile's source from the east, travelling through modern Tanzania. When Burton fell ill, Speke set out to the north. Burton recovered and continued west. The result was that Speke found Lake Victoria and Burton found Lake Tanganyika. They later rejoined and spent the trip home arguing over which was the source of the Nile. Heinrich

(Henry) Barth was a Prussian working for England. Multilingual and well trained in West African culture, Barth approached the region through the Sahara in 1850 and then extensively explored the region, visiting Timbuktu, Lake Chad, Agadez, and Adamawa. To this day, his account of his journey is considered an invaluable source of information concerning pre-colonial West Africa. Mungo Park is considered to be the first successful European explorer of Africa. In 1795, he entered the region via the Gambia and became the first European to reach the Niger River. He then explored two-thirds of it by making his way to its mouth.

1.4 Geographical Descriptions and Georeferences in Five Week in a Balloon

Jules Verne's novels are useful for testing new methods of georeferencing and NER due to their tendency to be geographically rich in nature. The location of his adventurers are indicated by either direct reference or indirect reference. A direct or precise reference is when the adventurer is in a very particular location relative to the geographic context, either a named place as a point (i.e. in London in the context of England) or a place marked by geographic coordinates (i.e. 51.5074° N, 0.1278° W). An indirect or relative reference occurs when the character is near but not in a particular place (i.e. 10 miles east of London) or within a polygon place (i.e. in London, but London is the context and a direct reference would be 'at Buckingham Palace'). *Five Weeks in a Balloon* is particularly useful for this experiment for two reasons. First, the setting is pre-colonial Africa. In this setting, the adventurers travel through towns and countries that existed prior to the mass colonization of the continent, which resulted in many places changing names or being abandoned. Second, a series of constant factors improve the

ability to focus on what is important. The travel range is confined to the African continent, unlike the globetrotting *Around the World in 80 Days* or *20,000 Leagues under the Sea*. Also, the means of travel is also consistent, a balloon. These constants help improve the focus on georeferencing place names.

As a secondary challenge, the distinction between direct and indirect references is important due to the issue of granularity. With direct references, the placement of an individual is very certain within a geographic context. Direct references to point features create the backbone of the locational context of this journey, and facilitate mapping, as in other studies of travel literature, e.g., McDermott (2017) and Cave (2016). Indirect references are considerably more uncertain as they either entail the individual being in a polygon in the geographic context or near, but not at, a point or polygon feature. Nonetheless, indirect references are still important because even if they cause an uncertainty of place, they still provide enough sense of place relative to direct references to tell us the direction of a journey. Furthermore, antiquated geographical entities, like former countries and regions, are just as important for this research as antiquated points, like towns.

This is the present question this thesis seeks to answer: *Is it possible to map a journey documented with antiquated place names, and if not, what should be done to improve the system?* This thesis shall use the Jules Verne novel *Five Weeks in a Balloon* as the source material to test a method on, drawing from gazetteers to reference locations. Once done, the results of the method are reviewed and features that should be improved to make automation easier are noted.

CHAPTER 2: LITERATURE REVIEW

Forays into the world of literature mapping have already been done and with results that are necessary to mention due to their findings supplying precedents to this research. These previous forays include several informal efforts, including a regularly cited blog post from Atlas Obscura (Figure 1), which manually reconstructed the trajectories of several American travel and adventure works (Kreitner and Melendez, 2015), and a more relevant Wikipedia map of the Jules Verne Five Weeks in a Balloon trajectory (Figure 2). Both efforts, as far as can be ascertained, were constructed manually, without the help of GIS, gazetteers, or programming. Nevertheless, they serve as examples and precedent for this work. The majority of attention in this chapter focuses on traditional academic works, using more sophisticated literary and geotechnical methods.

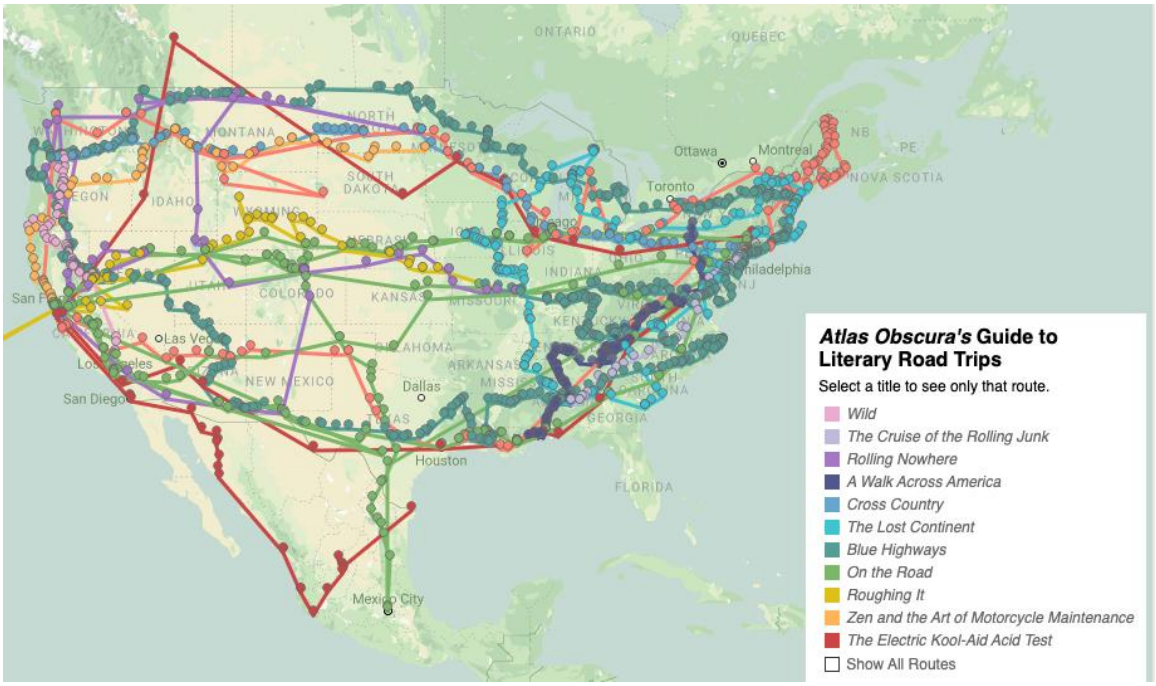


Figure 1: Kreitner and Melendez (2015) map of literary journeys

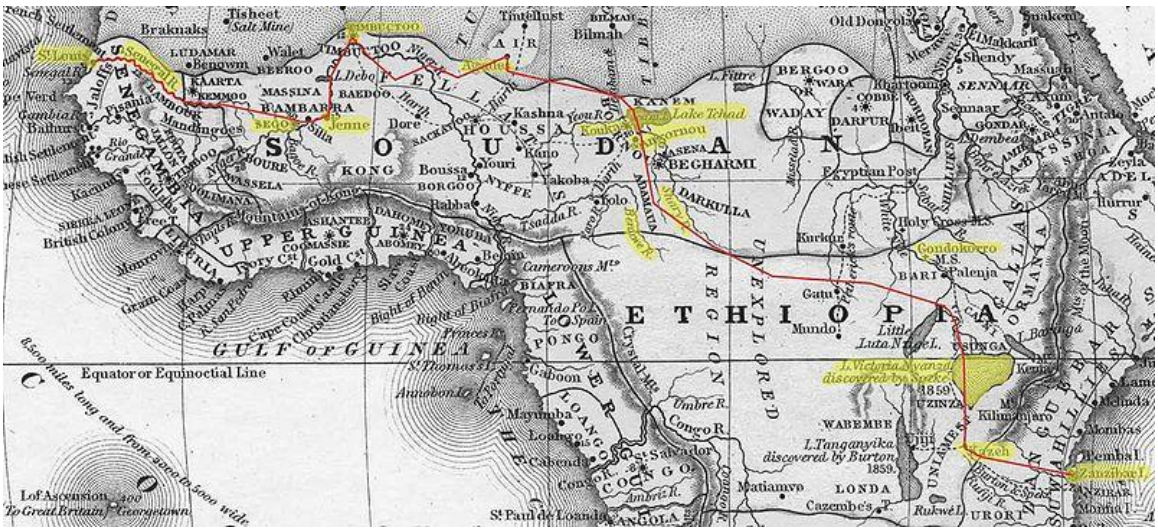


Figure 2: Roke's manually produced trajectory of Five Weeks in a Balloon, Wikipedia

While this thesis seeks to solve a certain aspect of the aforementioned scientific problem, several predecessors have made progress in other areas of the field to narrow down the focus. Their works and inquiries covering issues such as: gazetteer cooperation, georeferencing across language barriers, geographical ambiguities, automated georeferencing, mapping relative references, and analyzing georeferences in social media while dealing with colloquial information. This literature review summarizes the achievements of previous writers in this field and explain the importance of their results to this thesis.

2.1 International Workshop on Digital Gazetteer Research and Practice (2006)

In December of 2006, many of the foremost minds and pioneers of GIS gathered in Santa Barbara, California for the International Workshop on Digital Gazetteer Research and Practice to discuss the potential uses, practical uses, and shortcomings of gazetteers in their current state (Goodchild and Hill, 2006). People desiring to enter the workshop had to submit a position paper demonstrating their grasp of concepts and what they could contribute to the conversation. This workshop followed a similar successful workshop on new forms of spatial data sharing infrastructure, including gazetteers (Goodchild et al., 2005).

Of particular interest is the submission by Krzysztof Janowicz, then from the University of Muenster, currently University of California, Santa Barbara. In his submission, “Similarity-based Identity Assumptions for Historical Places”, Janowicz believes that global gazetteers need to strengthen their relationship with local gazetteers. He believes that a common language is not enough and that using all linguistic variations

of a place name can eliminate similarity problems. As things stand, scholars struggle to rely on local gazetteers due to their tendency to have vague or incomplete information, non-unique place names, place names that refer to many different kinds of features, misinterpreted names, names that refer to mobile political forces and artificial landmarks, and place names that are outdated. In short, he believed that global gazetteers need to draw data, including place names, from local, niche gazetteers to improve the quality of their own gazetteers.

2.2 Reconstructing and Mapping Historical Events using Multi-lingual Gazetteer Resources

In a 2014 thesis, Bekisz tackled the linguistic deviation of place names through the analysis and comparison of Lithuanian and German place names for the same or related features in a series of World War 2 Holocaust documents, including the Jäger Report. They wondered, as his mission statement notes "to what degree does toponym deviation as listed in the Jäger Report, hinder the research process with regards to place name georeferencing and discovering spatial associations?" In this Nazi report, known as the Jager report because it was written by a commander of the same name, it details, over the course of 5 months in 1941, the time, place, and number of people, mostly Jews, executed during the Nazi occupation of Lithuania. They cross-referenced this report with a period Lithuanian map and acquired coordinates from NGA. They divided the process into: extract place names from the Jager report, extract place names from the Lithuanian map, cross reference them with the NGA database, and associate them with acquired coordinates and non-geographic information. They also noted, on a scale of 0 to 2 the

range of place name deviation for German and Lithuanian names of the same place and so was able to deduce the degrees of linguistic deviation that occurred throughout the region, as well as spatial deviations that occurred between the Jager report and the geographical agencies of Lithuania.

Bekisz's work is significant because he produced a method for tackling linguistic deviations of place names. It is also significant because the work focuses on a dynamic, changing environment, where geopolitical events may have been a confusing factor in the linguistic specification of locations.

2.3 Creating, Mapping, and Analyzing a Travel Trajectory from a Novel

In a 2016 master's thesis, Cave recreated the journey of another Verne classic *Around the World in Eighty Days*. In this novel, the main character, Phileas Fogg, follows up on a bet that, using the infrastructure of the time, he can go around the world within 80 days, all while a detective tracks him, suspecting him of bank robbery. The purpose of her thesis was to determine if a journey, set in the real world, but contained in a fictional novel can be geoparsed and georeferenced while addressing the issues of ambiguity and relevance. An ambiguity occurs when a proper noun can be applied to multiple things. For her thesis, Cave broke it down to three forms of ambiguity; geo/geo, geo/non-geo, and dual. A geo/geo ambiguity occurs when a place shares a name with another place (i.e. Alexandria, Virginia vs Alexandria, Egypt). A geo/non-geo ambiguity occurs when a place shares a name with a non-place (i.e. John York vs York, England). A dual ambiguity is a combination of the previous two (i.e. George Washington vs Washington DC vs Washington State). With this criteria, Cave was able to separate useful information

from other details by identifying whether a proper noun was attributed to a geographic place and what place it was particularly referring to in the event was shared by other places. Another feature of her thesis was the analysis of the relevance of place name mentions. This was broken down into four categories: spatial relevance, temporal relevance, thematic relevance, and overall relevance. Spatial relevance occurs when the person mentioning a location is at said location. As Cave put it, "A reference is only considered to be spatially relevant for this project if the character the reference applies to is physically present at the location at that moment of the story." A location is spatially irrelevant before a character reaches it, after they leave it, and if they are never there. However, one must be mindful of reminiscences, where a character reflects on a past event that occurred in the referred place. Temporal relevance again occurs when the character is at the place at the time of its referencing. Consequently, spatial relevance and temporal relevance occur at the same time. Thematic relevance occurs if the place mentioned is relevant to the journey, usually a place to be visited or a place that has been visited. Overall relevance is the sum of the previous three relevancies. A place must be spatially, temporally, and thematically relevant to be relevant overall. By these relevancies criteria, Cave was able to phase out places mentioned in the book that were irrelevant to the journey.

2.4 Automated Geoparsing and Georeferencing with Frequency and Proximity Clustering Analysis

In the 2017 dissertation, *Frequency and Proximity Clustering Analyses for Georeferencing Toponyms and Points-of-Interest Names From a Travel Journal,* McDermott developed an algorithm capable of extracting place names and point-of-interest (POI) and georeferencing them onto a map and tested on a 1987 travel journal novel titled *The Lost Continent*. The algorithm uses clustering analysis, a method based on Tobler's First Law of Geography, to eliminate ambiguities. They build on the work done by Cave, stating that georeferencing travel journals was important for creating geographic information applications as well as intelligence and terrorism analysis. They believed the ambiguity issues she dealt with could be reduced if integrated into an automated process. They used clustering analysis to reduce data noise because of its ability to remove irrelevant references due to distance from clusters of places.

The contributions in McDermott's dissertation are important to this thesis because they demonstrate that with the appropriate analytical and conceptual framework, automating place name extraction and georeferencing is possible, even in a very large document. The difference between McDermott's work and this thesis however is that McDermott used a fairly recent novel with modern place names in contrast to this research's source material which has outdated place names. One of the main challenges is overcoming the antiquated place names issue and so create a method that can be automated. McDermott, however, documented the difficulty of geoparsing and georeferencing points-of-interest names, which differ from authoritative or official place

names in that they are not gathered in traditional gazetteer documents and change frequently.

2.5 Using CSA and LCP to chart travels in the Lake District

This team from the University of Chester (Murrieta-Flores et. al., 2017) did a project using cost-surface analysis (CSA) and least-cost-path (LCP) analysis to help utilize relative geographical references for charting out a journey. They opened by noting that while some humanities like Archeology have been using GIS for decades, others, like Literary Studies have only turned to it recently. This was because GIS was well suited for mapping precise references, which fields like Archeology were well acquainted with, while most literature tended to be full of relative references. The rigidity of point-based maps as they noted "are suitable for structuring the analysis of quantitative geographical phenomena, such as proximity and scale; however, they are less adequate for structuring the analysis of qualitative human phenomena, such as the experience of visiting a place or of travelling from one place to another." Using travel journals of a poet, a naturalist, and an agriculturalist through England's Lake District to provide a diverse set of viewpoints of the same area, the team digitized each account that concerned the Lakeland region. Each extraction was then analyzed for place names which were in turn organized using XML. They were then georeferenced using the Old Cumbria Gazetteer. This provided a dataset by which the team could implement a cost-surface analysis and a least-cost-path analysis that takes into account the topography and thus reconstruct the paths of travel.

This project is significant because the team created a method for overcoming the issue of relative references that are often the common geographic reference of literature. Unlike this research, they fortunately did not have to worry as much about temporal or cultural deviations of place names as the sources were 18th century accounts of English places written in the English language.

2.6 Analyzing Georeferences from Thematic Social Media

A more recent thesis by Phillips (2019) tackles the issue of colloquial place names while using Named Entity Recognition on tweets to track cartel activity in Central America. When selecting gazetteers for referencing place names, Phillips found region specific gazetteers to be the most useful as it rounded down the number of reference diversions that can cause geo/geo ambiguities. But, "Unfortunately, due to the extensive time and resources required to create a region-specific gazetteer, there is currently no gazetteer encompassing the Northern Triangle region. Thus, this presents a challenge that this study overcomes by using geo-crowdsourced information from Wikipedia and DBpedia." The ambiguity problem is intensified because the limited text of a tweet means limited geographic context is given to the place names. Colloquial names complicate the place names resolution of a place as they tend to be things like "village square" and "movie theater" rather than proper place names. Research by Rice et al. (2011, 2012, 2013) document the process of building a gazetteer with colloquial, informal, and foreign language place name variants for use in georeferencing crowdsourced data, and Aburizaiza et al. (2016) demonstrates how informal and indirect

locational references can be used to form spatial footprints. These techniques, and those in McDermott, are useful in informing this thesis research.

2.7 Mapping Literature: The Prototype of "A Literary Atlas of Europe"

Piatti et. al. (2009) experimented in methods for developing literary cartography while creating a prototype for their 'Literary Atlas of Europe'. They appear to understand that mapping a fictional world involves breaking it down into unique components. "To map a fictional world, the spatial structure of the prose has to be broken down into single elements and their respective functions." (pg. 2)

In the case of the Literary Atlas, the geography of literature can be broken down into five categories: setting, zone of action, projected space, route, and marker. A setting is where an action takes place, the zone of action is where several settings overlap, projected space is a place a character thinks about but is not in, a route is a path along which characters move, and a marker is a mentioned place used to help indicate the geographical scope and range of the story.

They tested the method on Lake Lucerne, North Frisia, and Prague for a diversity of environments. They concluded that while a second set of information layers needs to be created. They believe that the test successfully proved that cartographically representing literary settings is possible. This can be doubly so if geographic coordinates are provided by the book for obscure locations as *Five Weeks in a Balloon* does.

2.8 The Slipperiness of Literary Maps: Critical Cartography and Literary Cartography

In this article for Cartographica, S. Bushell, then of Lancaster University, discusses the important differences between the traditional maps of Critical Cartography and the narrative bound maps of Literary Cartography.

According to Bushell (2012), Critical Cartography covers maps in the traditional sense. This field has a dual nature. Its maps can be artistic or scientific in nature. They can be "decorative or precise, objective or subjective". Literary Cartography, on the other hand, is a strange customer in that while it is bound by the text of a literature it can also be chaotic to map based on the level of fictionality that occurs in said literature. Maps provided by literature can help us visualize the setting, but we must keep in mind that such maps are only telling us what they want us to know.

Five Weeks in a Balloon may be centered in the very real continent of Africa, but there are also locations that are fictional, especially as the explorers pass through regions that were unexplored in Verne's time. While reconstructing the journey it must be understood that there are things and places the book claims to exist at locations where there are not. The notional or realism and fictionalism in geographic locations are captured during this thesis research as an item in the database used for this work. This material is presented in Chapter 3.

2.9 Chapter Summary

Prior to this thesis, there has been many meditations and forays into the multidisciplinary field of Literary GIS. Janowicz discussed the necessity for global

gazetteers to cooperate with local gazetteers and include all linguistic variations of place names (2.1). Bekisz created a method for noting linguistic deviations that occur over multiple documents from different cultures (2.2). Cave figured out how to remove ambiguities and measure a geographic reference by its relevance to the present moment in a story (2.3). McDermott was able to automate the process of text extraction and path construction in a way that can be applied to modern texts with modern locations (2.4). Murietta-Flores et. al. achieved a significant inroad for mapping literature by creating a method for mapping relative geographic references (2.5). Phillips created a method for overcoming the issue of colloquial information that often comes from volunteered geographic information (2.6). Rice, Aburizaiza, and others build gazetteers with locally-specific place name variants for georeferencing and mapping crowdsourced data.

In spite of all this, none of these works inquire into georeferencing potentially abandoned places and Bekisz's research largely did not face linguistic deviations as severe as some of the locations in *Five Weeks* which have undergone fundamental changes since the writing of the novel. McDermott's work explored the problems with points-of-interest names, which do change frequently, but have different dynamics than the deviations in this research. There are countless documents made over the millennia of civilization that refer to places no longer occupied or are referred to by names or in languages that are no longer used by any modern people. How then do we georeference these antiquated texts? The goal of this thesis is to create such a method and take note what may be required to improve the method. Once more, the guiding research question for this work is as follows: *Is it possible to map a journey documented with antiquated*

place names, and if not, what should be done to improve the system? This thesis answers that question affirmatively in the following chapters, which present data collection and methodology employed in this research, as well as results, conclusions, and future work.

CHAPTER 3: DATA AND METHODOLOGY

This chapter covers the sources of data and the methods that are used in this thesis. This chapter also discusses the data sources and means by which data is extracted, sorted, and stored. It also discusses how third-party sources are used to improve the quality of the data so that the resulting map is more refined. Finally, this section shall discuss the means by which the results shall be displayed.

3.1 Data sources and storage

The source corpora for this thesis are three versions of *Five Weeks in a Balloon* by Jules Verne. Being not as well-known as *Around the World in 80 Days* or *20000 Leagues Under the Sea*, finding multiple editions was a bit challenging, even though the book is public domain due to copyright expiration. The first edition is a Wesleyan University Press edition made in 2015, it is supposed to be the most modern translation, at 291 pages. (Figure 3) The second edition is a 2016 copyright of the 1869 translation by William Lackland. It is copyrighted to Jules Verne and includes the 1867 illustrations by Riou and de Montaut at 145 pages. The Wesleyan Edition also has these illustrations. The last edition was published by Odin Library Classics, it is a Works of the Public Domain and is 249 pages long. However, these differences are mostly a result of text size and compression. Linguistically, they differ very little. *Five Weeks in a Balloon* is not as publicly well received in the present as compared to the likes of *Twenty Thousand*

Leagues Under the Sea or *Around the World in 80 Days*. Consequently, there are less editions of this novel and many of its forms are just the same edition with different covers. The Wesleyan edition is the version that is chiefly different while the Odin and Lackley are pretty much the same. In fact, they are probably the same version. The difference is largely that of sentence organization. There are no sections of the story that are exclusive to one edition. Most place names are the same across editions. The differences that are present are the usually the matter of a letter. The only major linguistic differences in place names is the spelling of Timbuktu. The Wesleyan edition spells it in its current English form listed prior. The other two spell it 'Timbuctoo'. Jules Verne was French, Timbuktu is located in Mali, a former French colony, and the current French spelling for the city is 'Tombouctou'. The use of 'Timbuctoo' in the Odin and Lackley editions, a name that differs from both the modern English and French versions of the name, suggests that this was the form of the name used by the English or French of the time. The similarity and general concordance of the three corpora sources provides a rich environment in which to extract, map, and analyze location data.

'"There's Kabra!" exclaimed the doctor, joyously; "there is the harbor of Timbuctoo, and the city is not five miles from here!" "Then, sir, you are satisfied?" half queried Joe. "Delighted, my boy!" "Very good; then every thing's for the best!"'–Odin edition p.130

'"It's Kabra!" the doctor exclaimed delightedly. "It's the harbor for Timbuktu; the town isn't five mile from here!" "So you're happy, sir?" Joe asked. "Ecstatic, my boy." "Well, this too shall pass"'—Wesleyan edition p. 259

‘“There’s Kabra!” exclaimed the doctor, joyously; “there is the harbor of Timbuctoo, and the city is not five miles from here!” “Then, sir, you are satisfied?” half queried Joe. “Delighted, my boy!” “Very good; then every thing’s for the best!”’—Lackley edition p. 222

To qualify as an entry, a place requires one of two things, a coordinate or a proper place name. If a place is given its own coordinate by the text, then it can be placed on a map regardless of whether it has a proper name or not. If a place has a name, then this name can be processed through a gazetteer by which one can get coordinates attributed to the place thus allowing it to be mapped. Without a name or coordinates, finding the location of a place is impossible in of itself. “I stood in a clearing in the woods” This example could mean any clearing in any forest. Even if the name of the forest is known there is another temporally unique problem which is that environments in a place change over time. What was a clearing may now be covered in forest, where there was not a creek there may be now, where there were once a series of sand dunes there may now be a sand flat thanks to the storm surge of a hurricane. If a coordinate pair is given for the clearing, then this can be used as a point that affects the line of trajectory for the journey, regardless of whether the environment of the place has changed and regardless of whether or not the place has a name. This criteria allows the establishment of ‘anchor points’ which can help show the direction of the journey.

If one must enter a place name to a gazetteer to acquire a coordinate, they must be prepared for the possibility that they will get multiple returns. That is, that there are many places that possess that name. This is a geo-geo ambiguity like the ones addressed by

Cave (2016). How does one deduce which is the place, and in turn the coordinates, they are looking for?

"Everything is related to everything else, but near things are more related than distant things." -Waldo Tobler (1970)

This is the first law of geography and the solution to the problem, one also employed in the works of McDermott (2017) and Phillips (2019). Knowing the sequence of locations in a journey is important because it gives context to a place of otherwise unknown location. For a place of unknown location to be placed between two places of known location, or even near one known location, gives it a relative position that narrows down the candidates that have the same place name. For example, there are many Springfield's in the United States. If a text said "from Springfield, east to Boston" it could refer to any Springfield west of Boston, Springfield, Massachusetts, Springfield, Illinois, etc. However due to relative position, it most likely refers to Springfield Massachusetts. It only become certain though when the place is between two known locations. "From New York City we went to Springfield, and from Springfield, east to Boston". So, if this research searches for a place's coordinates via gazetteer and gets multiple returns, the most likely candidate is deduced by looking at the locations of the places that came before and after it.

Figure 3: The front covers of *Five Weeks in a Balloon* used. From the left: Odin Library Classics, Wesleyan, and Illustrated (Lackley translation)

3.2 Data collection and storage

As each book was read, each time a place name or geographic coordinate as well as any information associated with the reference occurred, it was highlighted. Each reference and their information was then entered into excel, with a table for each book. Each table has a series of columns for information as follows. Table 1 contains a complete list of data collection items, definitions, and examples.

Table 1: Section of the excel sheet for storage, all full tables can be found in the Appendix

Section title	Section description	Example
Primary	Numbers uniquely attributed to an entry for searches and joins	1, 2, 3, etc.
Name	Name of a place. In the Combined table, this is spread into columns for each edition and the Johnson map, plus an 'other' column	Kaole
Lat	Latitude provided by the book	-4.28333 From G26 of Combined table
Long	Longitude provided by the book	29.25 From H30
Lat from gazetteer	Latitude provided by gazetteer	-1.735278 From I34
Long from gazetteer	Longitude provided by gazetteer	6.7630 From J53
Lat from Johnson	Latitude from the Johnson map	-7.083 From K18
Long from Johnson	Longitude from the Johnson map	37.393 From L18
Lat combined	All latitudes are in this column for map making	8.692 From M57
Long combined	All longitudes are in this column for map making	14.576 From N62
Coord Source	Where did the coordinate come from	Book, Gazetteer, or other
Gazetteer used	Tells what gazetteer was used if any	Getty, Google, or NGA
info	Miscellaneous, nom-geographic information associated with the place	part of the third and highest range of the Usagara system From Q20
factual, fictional, or both	Whether is real, is not, or is a mix of both. If both, why	both, Waterloo Place is a real street, not far from the Travellers club, but there does not appear to be a current place with the number three From R2
location type	If represented in shapefile	Point, line, polygon,

	form, what feature type would be best for this location in the context of Africa?	multipoint
relative or precise	Whether the reference was a precise of relative one	Relative or Precise
relationship with other locations	The relationship of a place with other places	northern shore of Lake Chad, between villages Lari and Ingemini From U63
address	Address of place if given	3 Waterloo Place From V2
chapter	Chapter in which place was reached	18 From W33-40
month	Month in which place was reached	4 From X10-48
day or last day	Day or last day of journey leg or being at a place	20 From Y23-28
Journey day	Day of journey in non-calendar form	33 From Z85-86
year	Year	1862
time	Hour of day	11 From AB64
am/pm	Period of day when time was given	AM or PM
view	View from the balloon of area if described	Low houses on the curves of the river long lines of donkeys and camels seen From AD83
elevation	Elevation of balloon in feet if given	1500 From AE11
foreign	Foreign key used for joins	1, 2, 3, etc.

‘Primary’ (row 1) is for the primary key of each entry on a table. ‘Name’ (row 2) is for the name of a location, if given. In the table where all data shall be combined, there are five columns: one for each edition, plus one for names from a period map, and one

‘other’ column for names acquired by tertiary means. ‘Lat’ (row 3) is for latitude when given by the book and ‘long’ (row 4) is for longitude when given by the book. ‘Lat from gazetteer’ (row 5) is for latitude provided by a gazetteer and ‘long from gazetteer’ (row 6) is for longitude provided by gazetteer. ‘Lat from Johnson’ (Row 7) and ‘Long from Johnson’ (Row 8) contain coordinates from a rectified period map. ‘Lat combined’ and long combined’ (rows 9 and 10) are the latitudes and longitudes of the previous four columns combined to make it easier to display on a mapping software. ‘Coords source’ recounts what the source of the coordinates are, either from the book, gazetteer, or other (Row 11). The ‘Gazetteer’ column tells what gazetteer was used, if required (Row 12). If a coordinate was not provided in the book, the gazetteers mentioned in 3.3 are used. If a gazetteer can find the place, the coordinate is recorded and the gazetteer that succeeded is recorded with that entry. The ‘Info’ column is for miscellaneous information associated with a place, like the events that happened there (Row 13). The ‘factual, fictional’ column holds the determination as to whether the location is a real place or a fictional location original to the book (Row 14). ‘Location type’ is for whether the place is a point or an area (Row 15). The ‘Relative or Precise’ column is for whether the reference is relative or precise in nature (Row 16). The column ‘Relationship with other Locations’ describes a place position relative to another noted location (Row17). ‘Address’ is if an address is given, a largely unnecessary column given the setting but every detail was considered as a possibility during the data collection phase (Row 18). ‘Chapter’ notes the chapter in which the travelers reach their destination (Row 19). ‘Month’, ‘day’, and ‘year’ collectively store the date of the initial arrival and the time and am/pm columns

contain the time of arrival (Rows 20, 21, and 23). ‘Journey day’ is the day of the journey in non-calendar form, starting at Koumbeni Island near Zanzibar where the balloon is launched and at day 0 (Row 22). This is as important as highlighting the geographic coordinates because it is important to retain the temporal order of events to reconstruct the journey, especially if the group splits up. ‘Time’ (Row 24) and ‘am/pm’ (Row 25) note the hour the balloon arrives at a place and whether it was in the morning or evening. ‘View’ contains a description of what was seen on the earth below from the balloon (Row 26). What is seen may differ based on how high the balloon was. ‘Elevation’ contains the elevation of the balloon in feet (Row 27). This column is tricky in that elevations relative to both ground level and sea level are present. ‘Foreign’ holds the foreign key of an entry, which is used for joins (Row 28).

3.3 Gazetteers: finding coordinates for place names and non-metric georeferences

The novel provided geographic coordinates at certain points, especially when the team entered unexplored territory where there were no discovered landmarks. However, there are not enough specific coordinate data to satisfactorily map out the journey. Verne generally did not give coordinates for major places like towns, making it necessary to turn to gazetteers.

Another method for acquiring geographic information about a place is gazetteers. A gazetteer is a geographical dictionary, sometimes used as an index for a map or atlas. It contains geographical information associated with a place within the scope of the gazetteer, which can range from the world to a defined region of it. “Gazetteers have

traditionally been known as dictionaries of place names, and they are familiar as reference volumes containing short descriptions of named geographic places or as indices at the back of atlases containing lists of place names providing the page number and map grid where each place can be found." (Hill, 2008) Regional gazetteers often describe their region in more detail than a global one and some of the more detailed gazetteers reach over linguistic and even temporal boundaries to ensure maximum association of places with appropriate positions. As noted in the previous chapter, the spatial granularity of a gazetteer is key. Rice et al. (2011, 2012) created a gazetteer with highly detailed local place names, which helped provide support for geo-crowdsourcing.

For this project, the following gazetteers were used: Getty Institute, the National Geospatial Intelligence Agency (NGA), and Google Maps. Getty is a museum, research, and conservation institute with the following mission statement: "Getty advances and shares the world's visual art and cultural heritage for the benefit of all." While primarily interested in art, Getty understands that art in its creation is influenced by the geography of its conception. With this understanding, Getty has created the Thesaurus of Geographic Names. While it does not consider itself a GIS, the Getty Thesaurus of Geographic Names (TGN) exists as a resource for finding geographic coordinates for places "both current and historical" (about TGN). NGA is an agency of the United States government that specializes in geospatial intelligence, that is, extracting information of importance to the military and to first responders from geographic imagery. To aid them in their mission, NGA has created the GEOnet Names Server (GNS). The GNS "is the official repository of standard spellings of all foreign geographic names, sanctioned by

the United States Board on Geographic Names (US BGN). The database also contains variant spellings (cross-references), which are useful for finding purposes, as well as non-Roman script spellings of many of these names. All the geographic features in the database contain information about location, administrative division, and quality. The database can be used for a variety of purposes, including establishing official spellings of foreign place names, cartography, GIS, GEOINT, and finding places." The NGA uses the GNS to locate foreign places and all of their linguistic variations. However, since this gazetteer is not culturally focused it has limitations, particularly on historical matters. The web resource www.maplandia.com is a gazetteer developed from Google Maps. Its scope is in the present, but there is also in-depth information, cataloging all manner of obscure places in the world. At their respective sites, one can enter the name of a place and get a result or results suggesting the possible location of the place. One can then find the most likely place, depending on the context of the location and acquire the geographic coordinates. Then, the coordinates are entered in the appropriate columns in the Excel worksheet that has the information acquired from the novels.

3.4 Maps: finding period works of cartography

Although they do not provide exact coordinates like a gazetteer does, a historical map of a region can provide a general context of locations, especially forgotten locations that a gazetteer, most of which are focused on modern locations, does not provide. Knowing *Five Weeks in a Balloon* was published in 1863, a quick Google search found the image shown below. The image is a 3563x5000 pixel 1863 map of Africa published by Alvin Jewett Johnson, owner of Johnson and Ward, in the 1863 edition of *Johnson's*

New Illustrated Family Atlas. (Figure 7) Johnson maps are prized to this day for their detail, accuracy, and coloring. Close examination has found many locations from the novel that the gazetteers failed to locate. And by providing a visual context for these unfound places one can more easily find their approximate location via a map software like Google Maps or Google Earth and get the coordinates. The map possesses a projection, either Mercator or something related to it. It appears to use the equator as its line of tangent and warping occurs the farther one gets from the equator

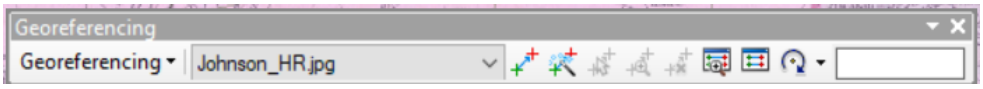
To account for these distortions and to project the map in a way that coordinates can be extracted from it, this thesis will use georectification. To do this the map is imported to ESRI ArcMap, version 10.6.1 released in 2018 is used for this thesis. This software is used because the Georeferencing toolbar in that program is useful for georectification. The Georeferencing toolbar is located in Customize > Toolbars. (Figure 4) When a certain location possesses a determined geographic coordinate, such as 0, 0 off the coast of West Africa, it is marked with a control point. This control point is then right-clicked and the appropriate x, y coordinates entered. This process is repeated until ideally at least 30 spatially-distributed control points are entered (Figure 8). Points can be considered spatially-distributed, not when they are clustered in a single area, but rather when the points are as spread out as possible, with points in the northern, southern, eastern, western, and center areas of the map. As these points are entered, the map is rectified using a third-order polynomial transformation based on the points, creating a unique projection (Figure 8, 9). The transformation is necessary because the Johnson map is projected, most likely with a Mercator cousin that uses the equator as a line of tangent.

The third-order polynomial transformation is ideal because with so many control points affecting the warp, the ability to find coordinates is maximized without using a more extensive time-consuming transformation. The third-order polynomial transformation also allows more complex geometric adjustments beyond simple translation and rotation. Comparisons with geometric transformations of other types, such as those using triangulated irregular networks, e.g., Camelli et al. (2012) were explored, but the third-order polynomial transformation seemed more appropriate for this application and for a map at this scale.

With the map geo-rectified, one can examine the map and find locations, both point and area, move the mouse over them, and record said locations in an excel table. The contemporary political boundaries (including coastlines) are overlain on the rectified image, and as can be deduced from the overlay, the rectification procedure is qualitatively good and acceptable for the purposes of this research, which seeks to reproduce the trajectory contained in the novel. It is expected for some points to have a large residual error, especially away from the equator. (Figure 5, 8) This is because of the warping caused by the projection. Since the image is a scan of an original map, it is also possible that stretching could have occurred as a result of the scanning, which can cause residual errors.

The results of the rectification procedure (Figure 6) indicate that the accuracy of the third-order polynomial transformation is approximately 0.0616364 digital degrees, which translated to ground units (kilometers) is 6.854 km. For Figures 9 and 10, a layer

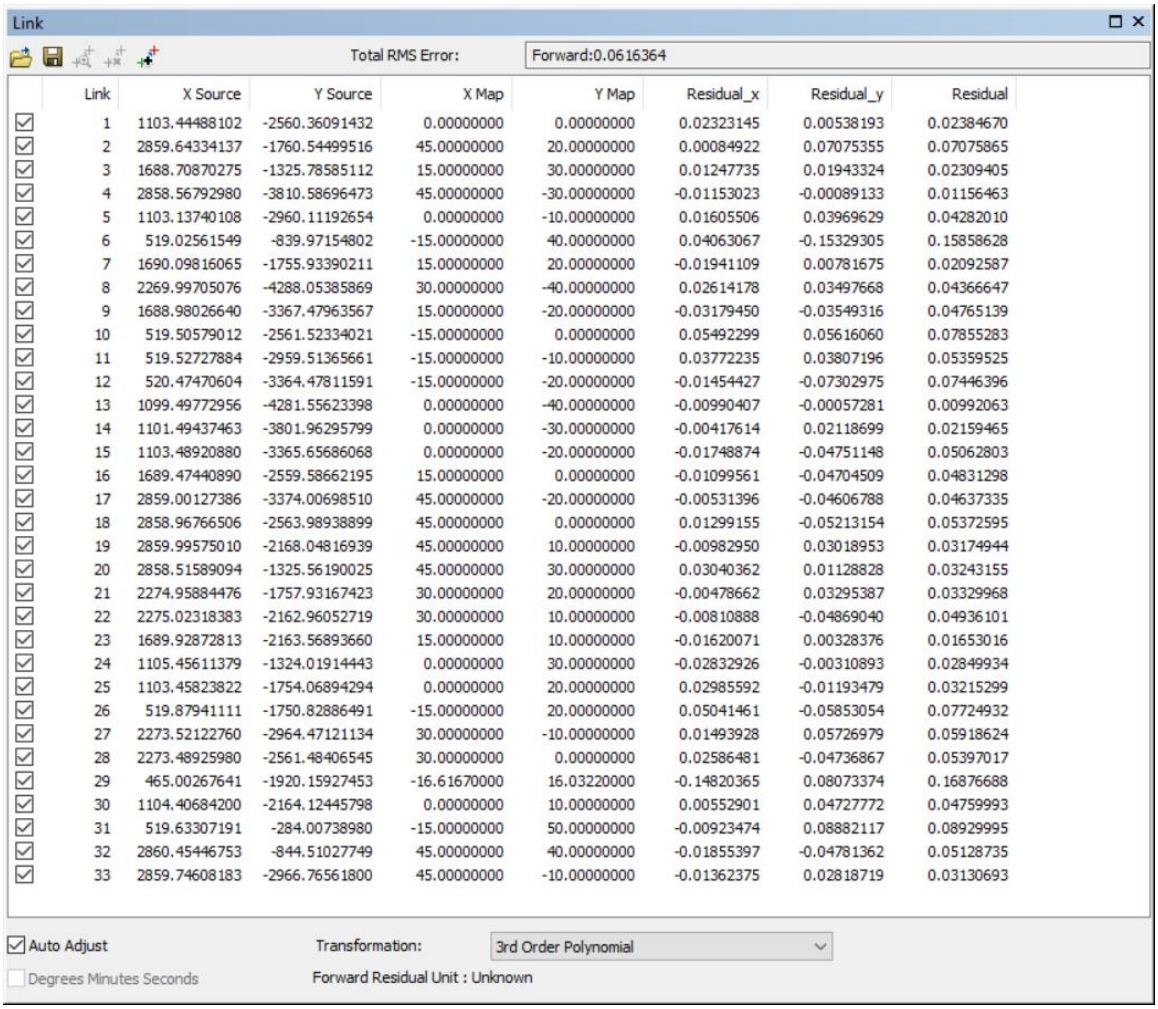
containing modern country borders has been added to demonstrate both the accuracy of the map itself and of the rectification.



Rectification Control Points (Johnson Map, 1863)

Image X	Image Y	Latitude	Longitude
1103.444881	-2560.360914	0	0
2859.643341	-1760.544995	45	20
1688.708703	-1325.785851	15	30
2858.56793	-3810.586965	45	-30
1103.137401	-2960.111927	0	-10
519.0256155	-839.971548	-15	40
1690.098161	-1755.933902	15	20
2269.997051	-4288.053859	30	-40
1688.980266	-3367.479636	15	-20
519.5057901	-2561.52334	-15	0
519.5272788	-2959.513657	-15	-10
520.474706	-3364.478116	-15	-20
1099.49773	-4281.556234	0	-40
1101.494375	-3801.962958	0	-30
1103.489209	-3365.656861	0	-20
1689.474409	-2559.586622	15	0
2859.001274	-3374.006985	45	-20
2858.967665	-2563.989389	45	0
2859.99575	-2168.048169	45	10
2858.515891	-1325.5619	45	30
2274.958845	-1757.931674	30	20
2275.023184	-2162.960527	30	10
1689.928728	-2163.568937	15	10
1105.456114	-1324.019144	0	30
1103.458238	-1754.068943	0	20
519.8794111	-1750.828865	-15	20
2273.521228	-2964.471211	30	-10
2273.48926	-2561.484065	30	0
465.0026764	-1920.159275	-16.6167	16.0322
1104.406842	-2164.124458	0	10
519.6330719	-284.0073898	-15	50
2860.454468	-844.5102775	45	40
2859.746082	-2966.765618	45	-10

Figure 4: Georeferencing toolbar, and 30 control points used for rectification



Link	X Source	Y Source	X Map	Y Map	Residual_x	Residual_y	Residual
1	1103.44488102	-2560.36091432	0.00000000	0.00000000	0.02323145	0.00538193	0.02384670
2	2859.64334137	-1760.54499516	45.00000000	20.00000000	0.00084922	0.07075355	0.07075865
3	1688.70870275	-1325.78585112	15.00000000	30.00000000	0.01247735	0.01943324	0.02309405
4	2858.56792980	-3810.58696473	45.00000000	-30.00000000	-0.01153023	-0.00089133	0.01156463
5	1103.13740108	-2960.11192654	0.00000000	-10.00000000	0.01605506	0.03969629	0.04282010
6	519.02561549	-839.97154802	-15.00000000	40.00000000	0.04063067	-0.15329305	0.15858628
7	1690.09816065	-1755.93390211	15.00000000	20.00000000	-0.01941109	0.00781675	0.02092587
8	2269.99705076	-4288.05385869	30.00000000	-40.00000000	0.02614178	0.03497668	0.04366647
9	1688.98026640	-3367.47963567	15.00000000	-20.00000000	-0.03179450	-0.03549316	0.04765139
10	519.50579012	-2561.52334021	-15.00000000	0.00000000	0.05492299	0.05616060	0.07855283
11	519.52727884	-2959.51365661	-15.00000000	-10.00000000	0.03772235	0.03807196	0.05359525
12	520.47470604	-3364.47811591	-15.00000000	-20.00000000	-0.01454427	-0.07302975	0.07446396
13	1099.49772956	-4281.55623398	0.00000000	-40.00000000	-0.00990407	-0.00057281	0.00992063
14	1101.49437463	-3801.96295799	0.00000000	-30.00000000	-0.00417614	0.02118699	0.02159465
15	1103.48920880	-3365.65686068	0.00000000	-20.00000000	-0.01748874	-0.04751148	0.05062803
16	1689.47440890	-2559.58662195	15.00000000	0.00000000	-0.01099561	-0.04704509	0.04831298
17	2859.00127386	-3374.00698510	45.00000000	-20.00000000	-0.00531396	-0.04606788	0.04637335
18	2858.96766506	-2563.98938899	45.00000000	0.00000000	0.01299155	-0.05213154	0.05372595
19	2859.99575010	-2168.04816939	45.00000000	10.00000000	-0.00982950	0.03018953	0.03174944
20	2858.51589094	-1325.56190025	45.00000000	30.00000000	0.03040362	0.01128828	0.03243155
21	2274.95884476	-1757.93167423	30.00000000	20.00000000	-0.00478662	0.03295387	0.03329968
22	2275.02318383	-2162.96052719	30.00000000	10.00000000	-0.00810888	-0.04869040	0.04936101
23	1689.92872813	-2163.56893660	15.00000000	10.00000000	-0.01620071	0.00328376	0.01653016
24	1105.45611379	-1324.01914443	0.00000000	30.00000000	-0.02832926	-0.00310893	0.02849934
25	1103.45823822	-1754.06894294	0.00000000	20.00000000	0.02985592	-0.01193479	0.03215299
26	519.87941111	-1750.82886491	-15.00000000	20.00000000	0.05041461	-0.05853054	0.07724932
27	2273.52122760	-2964.47121134	30.00000000	-10.00000000	0.01493928	0.05726979	0.05918624
28	2273.48925980	-2561.48406545	30.00000000	0.00000000	0.02586481	-0.04736867	0.05397017
29	465.00267641	-1920.15927453	-16.61670000	16.03220000	-0.14820365	0.08073374	0.16876688
30	1104.40684200	-2164.12445798	0.00000000	10.00000000	0.00552901	0.04727772	0.04759993
31	519.63307191	-284.00738980	-15.00000000	50.00000000	-0.00923474	0.08882117	0.08929995
32	2860.45446753	-844.51027749	45.00000000	40.00000000	-0.01855397	-0.04781362	0.05128735
33	2859.74608183	-2966.76561800	45.00000000	-10.00000000	-0.01362375	0.02818719	0.03130693

Figure 5: Full Control Points with Residual Error

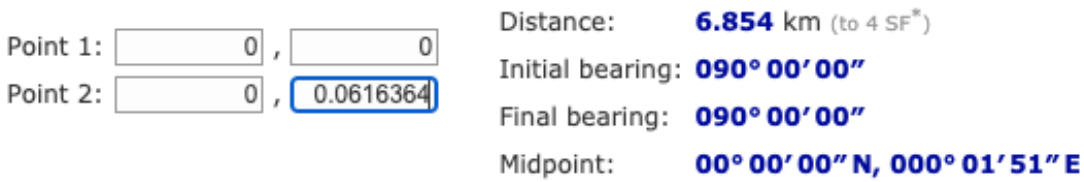


Figure 6: Rectification calculation and RMS Error statistics

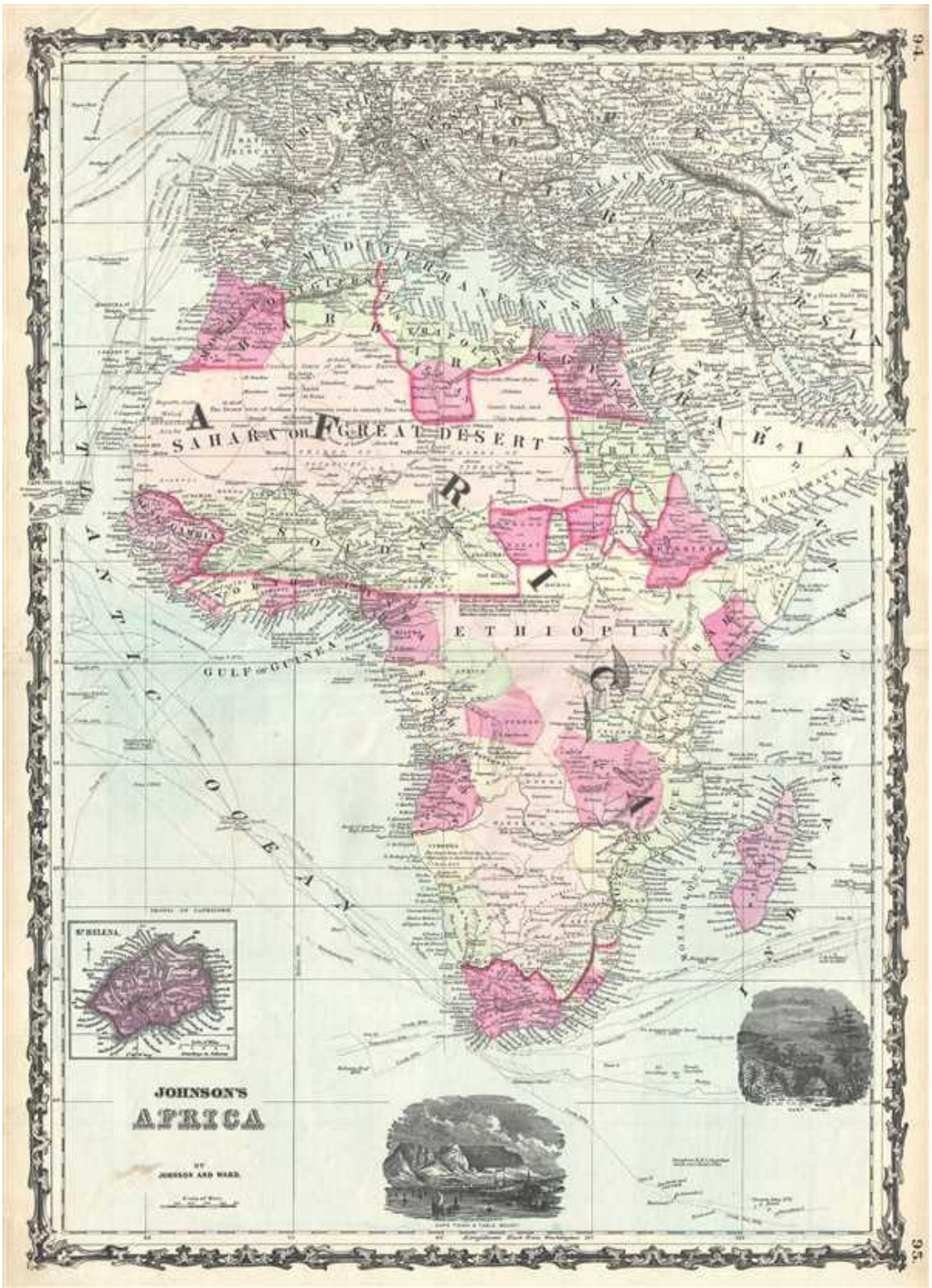


Figure 7: 1863 map of Africa

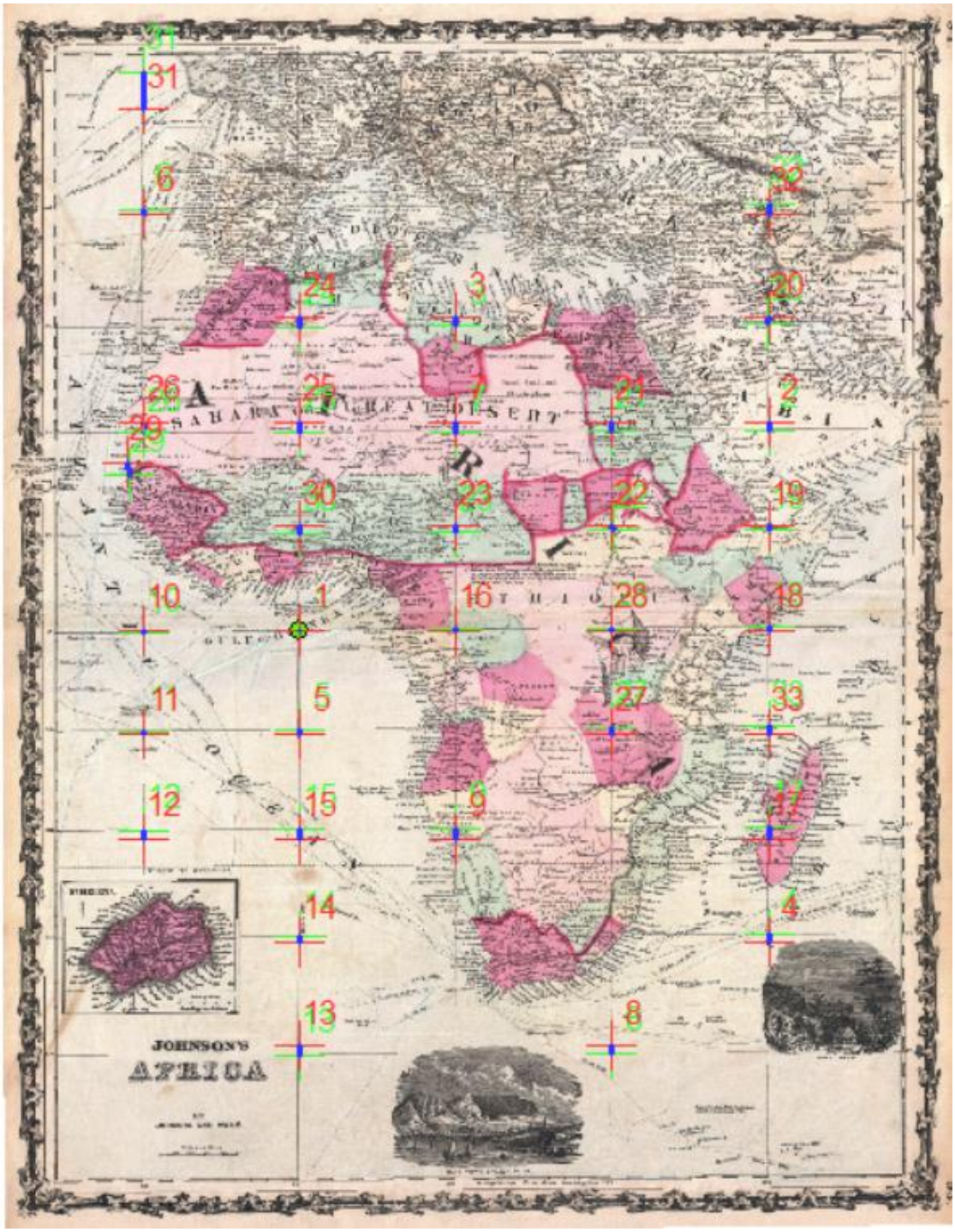


Figure 8: Map with control points

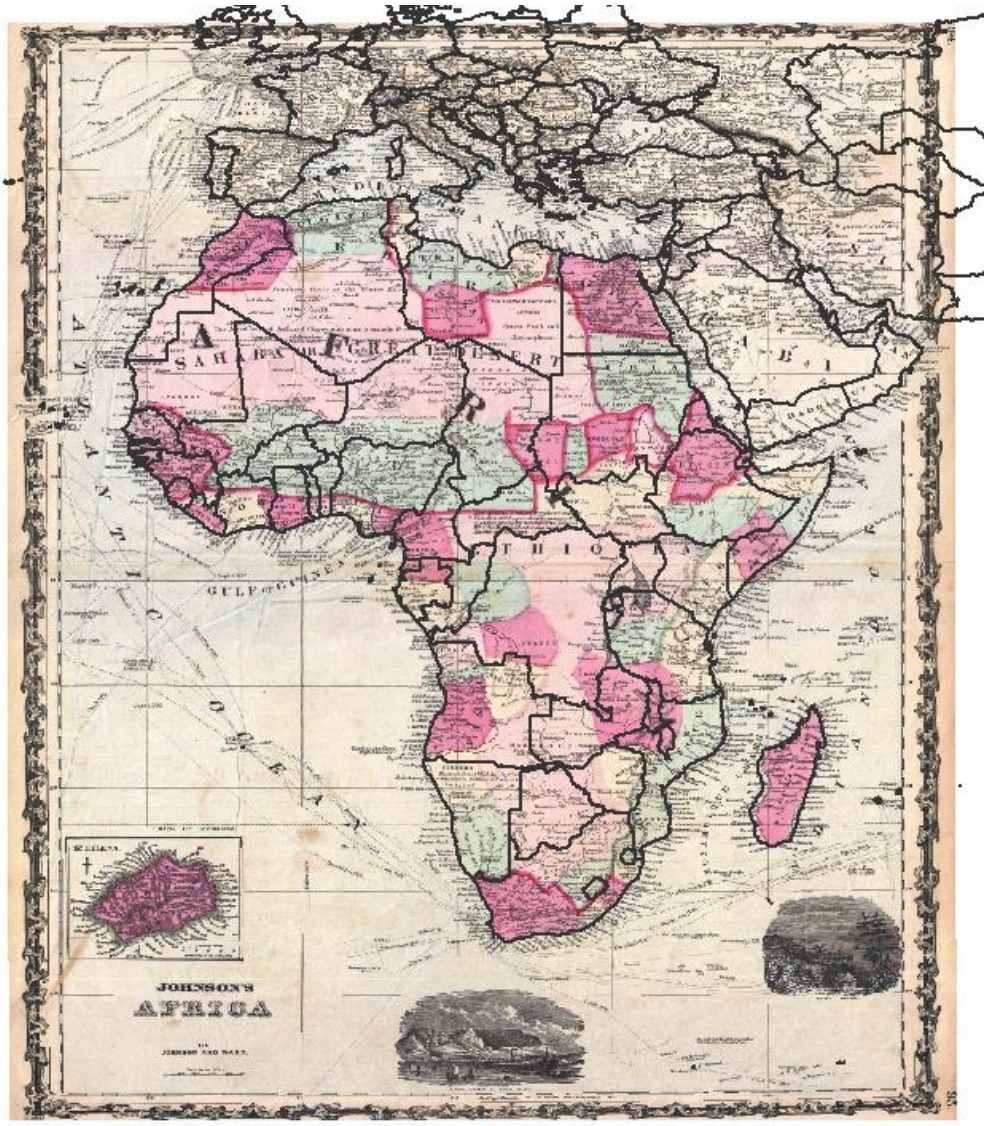


Figure 9: Rectified Johnson map (1863) with country boundaries in black, showing good qualitative result of the rectification process.

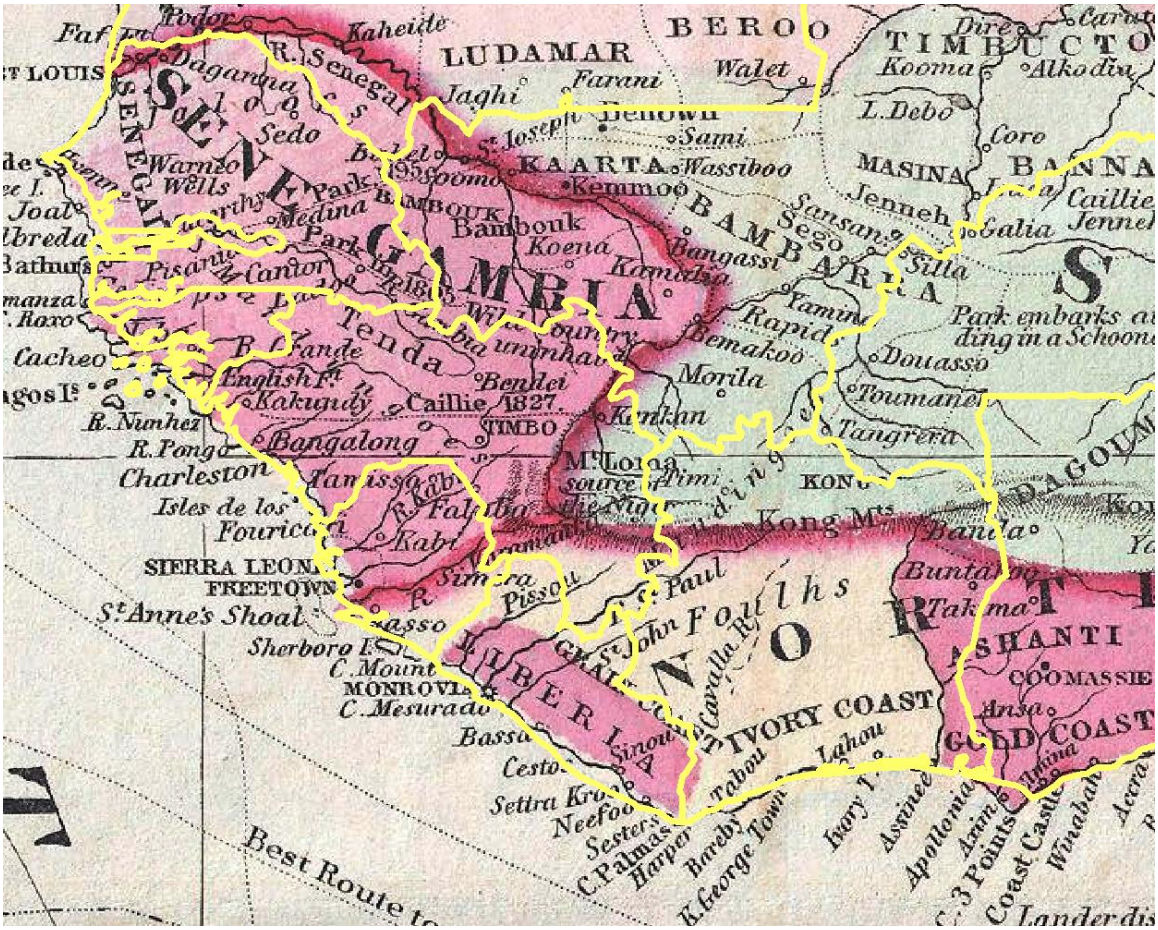


Figure 10: Johnson 1863 map and country outlines in yellow, with detail of West Africa

3.5 <u>Putting it together: A table combining all the data</u>

All data gathered from the books, the gazetteers, the Johnson map, etc. is accumulated into a single excel table, titled ‘Combined’. This is to streamline the data and to ease the observation of place name deviations between books. Data from this table is processed into graphs for analysis via a separate statistics table. Putting the data altogether helps make it easier to see the big picture and answer the research question of

this thesis. A complete copy of this combined data table is contained in the Appendix (Tables 2-6) at the end of this thesis, as well as at a URL contained therein.

3.6 QGIS: Displaying results

With all data gathered from the novels, gazetteers, and period maps, the final result can now be displayed in a mapping software. For this thesis, QGIS is used. QGIS is a free and open source GIS, similar in function to ArcMap. If the excel tables are saved individually as .csv comma delimited format, they can be imported to QGIS and be displayed as layers, via Layer, Add Layer, Add Delimited Text Layer. The version being used is QGIS 3.6.1 with GRASS 7.6.1. The table is displayed as a point layer, a series of anchor points made possible by coordinates acquired. If a coordinate was not acquired for an entry, the entry cannot be displayed. A line layer is made to manually link the points in chronological order to show the route of the journey. Since the entries were made chronologically, one can use the identify tool to look at a points attributes, namely its number from the ‘primary’ column, to know the order of locations and thus how to draw the line.

This thesis uses QGIS because initial data gathering and map making was not only done for this thesis but also for a semester project for GGS 664 which mandated the use of QGIS over ArcMap. After the class, continuing to use QGIS for map making was easier then transferring everything over to ArcMap.

3.7 Chapter Summary

Geographically-relevant data in the book is highlighted and entered into an excel sheet, one sheet for each book. Additional geographic coordinates and other relevant information, if possible, is drawn from the Getty, NGA, and Maplandia gazetteers, as well as from a map made in the year of the story. Each book's table is saved individually as a comma delimited csv filed and be imported as a point layer to QGIS. Then the data is combined into a single table to see the big picture; observe place name deviations, analyze data, and create a point layer that is the summary of all information acquired from books, gazetteers, and other sources of geographic information. A line layer is then be manually made, connecting the points chronologically to reconstruct the journey in the software and determine the accuracy of the method. *Is it possible to map a journey documented with antiquated place names, and if not, what should be done to improve the system?*

CHAPTER 4: RESULTS AND ANALYSIS

This chapter covers the results of the methods applied and an analysis of said results. Maps produced from the extracted data is covered first, showing the results of each edition, followed by a map summary of all the data from the books, gazetteers, maps, and other data sources. This is followed by graphical and chart-based summaries of the data and a discussion and analysis of the results.

4.1 Mapping the Travel Trajectories

This section covers the map results. The data gathered onto Excel spreadsheets from each edition is imported to QGIS and exported as maps. There is one map for each of the three editions. There is also a fourth map made from the table that combines the data from the three editions plus data gathered from sources like the Johnson map. In addition, there is a map addressing the temporal flow of the journey and a map showing all routes. The results are as follows: Wesleyan, Odin, Illustrated/Lackley, Combined, Temporal, and All.

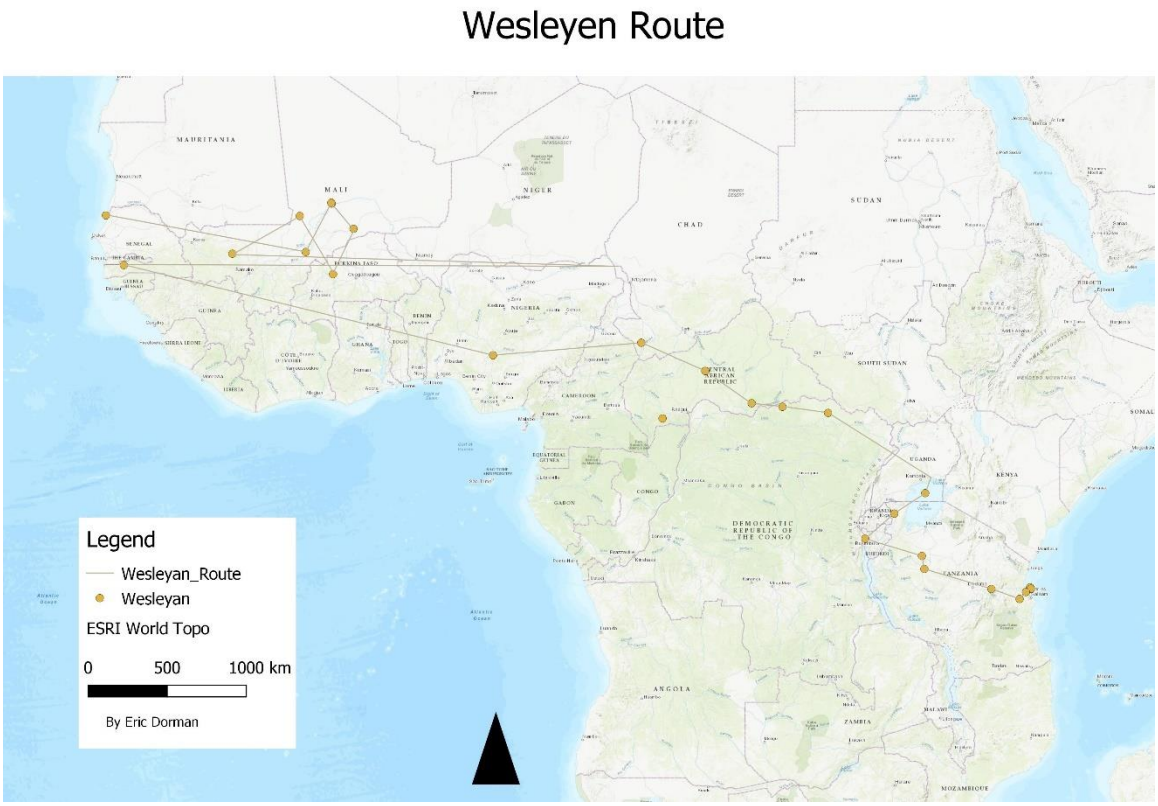


Figure 11:Route from the Wesleyan book.

This map (Figure 11) is from the first edition read, the Wesleyan edition. There is one chief error that occurred when getting coordinates. The location of the entry that was acquired for "Kouka" via gazetteer is in West Africa, when the book clearly states that the town in question is near Lake Chad. This error was retained because the error was dealt with as the other editions were reviewed and to serve as an example that the geo/geo ambiguities Cave dealt with in their thesis are just as capable of occurring in a place as culturally and linguistically unique as Africa as anywhere else. Otherwise, the other oddity is the team seeming to go in a loop in central West Africa. This error too was reexamined and addressed when georeferencing the other editions. There is one point on

this map that is not connected with the rest, this marked the southernmost point of Barth's journey and was not a point reached by the fictional adventurers.

Figure 12: Odin Route

This map (Figure 12) is based off of the route of the Odin Library Classics edition. A reexamination of gazetteer results helped find locations that were more sensible, creating the route above. One thing that should be noted. The Kouka in the modern nation of Chad is referred to in the book as a town near Lake Chad. The Kouka

found via Maplandia that was nearest to Lake Chad was empty desert to its east. This fit the narrative, but suggested that the site of Kouka has been abandoned.

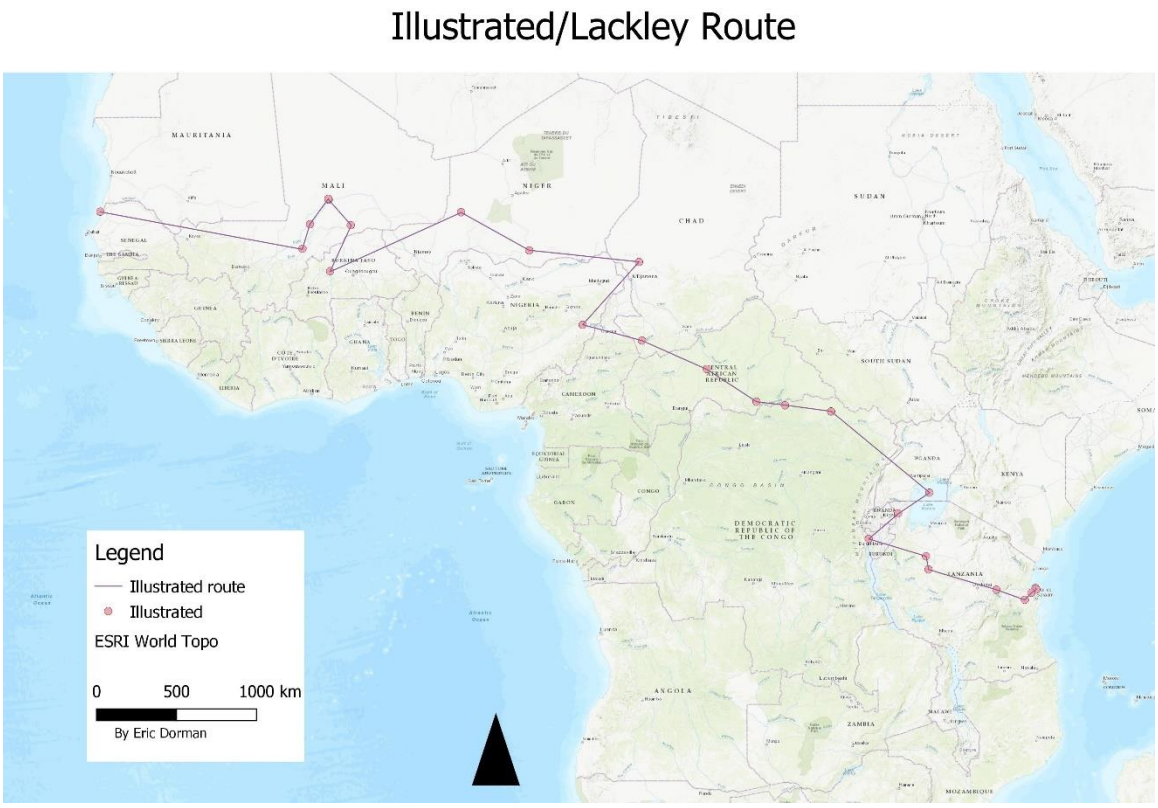


Figure 13: Illustrated/Lackley Route

The georeferencing of the Illustrated/Lackley edition (Figure 13) turned out much the same as the Odin route. As this is this is the most refined result, here shall begin the

analysis of the route and the data behind it. The logic used behind exploring Africa by balloon and by launching from an Eastern location like Zanzibar was based on the idea that the balloon would ride the trade winds, which usually go from east to west. The balloon generally does just that except on three occasions: when going from the Malagarazi River to Lake Victoria, when going from the Benue River to the region of Lake Chad, and when going from Goa to the Hombori Mountains and Timbuktu beyond. One should keep in mind that this is a fictional journey that relies on the adventurers seeing the major sights of real explorers, so if one was to attempt to recreate the journey, there is no promise that the winds will do as they so conveniently did in the book.

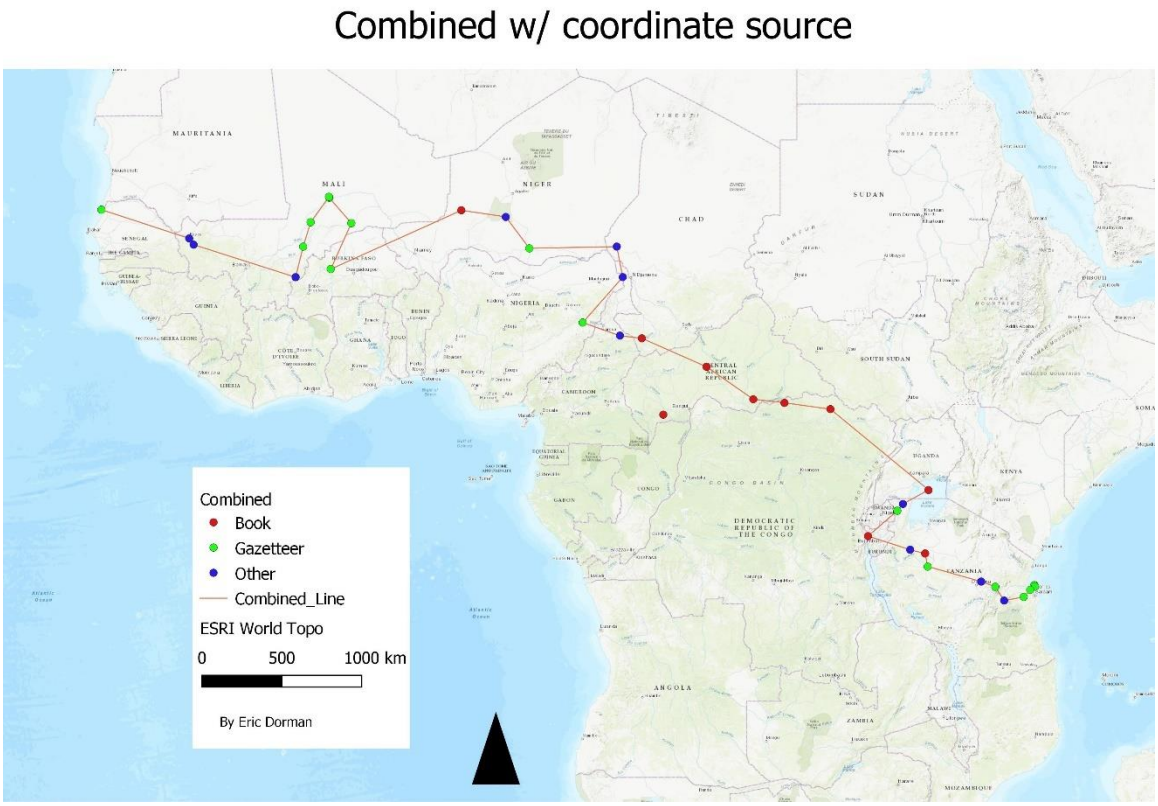


Figure 14: Map of all data combined

On its own, the Johnson map did not provide enough locations to get its own map. However, it can be used to augment existing data to provide a better map. With all the data from the books, gazetteers, and tertiary sources accumulated onto a single excel table, this map (Figure 14) was produced showing the journey at its most detailed. The balloon briefly goes southwest before turning northwest to the northern tip of Lake Tanganyika. It then turns northeast to Lake Victoria. There was a place on the Nile afterwards that the adventurers stopped at that was allegedly reached by the explorer Andrea Debono, but without a coordinate, it could not be charted. Now begins the purely

fictional leg where the adventurers pass through what Verne says is a desert but is actually the Congo basin, which is largely rainforest. They enter the explored sphere of West Africa through the Adamowa region in the southeast and make their way north to Lake Chad. After their adventures there, they went hard west along the edge of the Sahara, though Verne does not mention desert, seeing towns like Zinder and Agadez before going southwest across the Niger. Starting at the Hombori Mountains, they do a u-shaped leg going up to Timbuktu and down past Lake Debo to Jenne. Finally, they go northwest to French Senegal. The balloon fails them at Gouina Falls, where a French squadron finds them and takes them to Fort Medina and then to St. Louis.

In this map the points are classified by the origin of the coordinates; book, gazetteer, or other. All of the points from the gazetteer or other category mimic the Eastern and Western spheres of exploration of the time. This is because the coordinates for these categories are based off of names provided by the book or from period maps. Neither Verne nor Johnson would know of any place names from outside the explored region because no one has been there to record those places. Meanwhile most of the points from the book are from the uncharted region. Verne provided these coordinates heavily there to compensate for the lack of any place names to anchor the journey in that region. This was not because there were no places there, but because no one, or at least no European, had been there.

Figure 15: Map showing temporal flow of journey

This map (Figure 15) studies the flow of time during the journey. They lost the balloon at Gouina Falls on day 40 and reached St. Louis on day 53. They spent a little over five weeks in the balloon and crossed the whole continent in a little over seven. Discounting St. Louis and Medina since they were not reached by balloon, the median day is 20. At this point in the story, the travelers were in Adamowa, having just left uncharted territory. Distance-wise, they have covered more then half of the distance between the beginning at Zanzibar and the end at Gouina Falls. This means that the average speed of the first leg of the journey was greater than that of the second. In particular, the leg prior to Lake Victoria was covered in only five days.

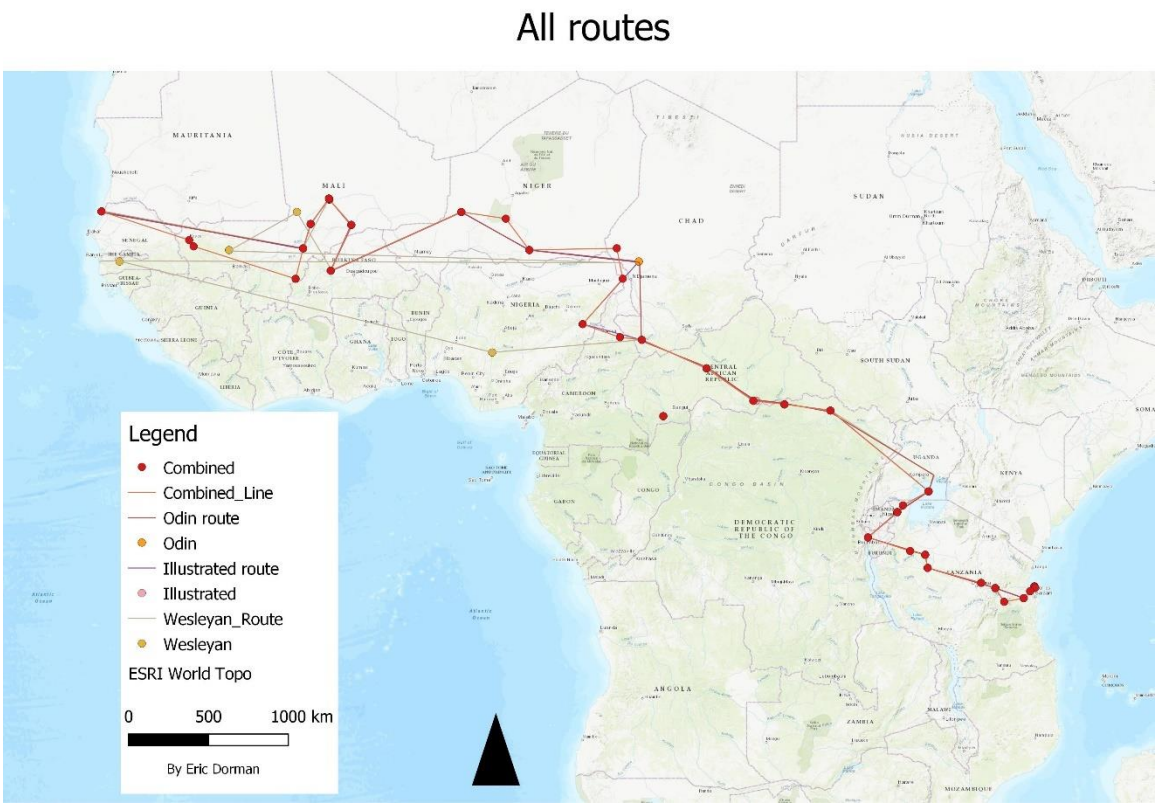


Figure 16: Map displaying all routes for comparison

This map (Figure 16) shows all routes from the three editions and the Combined table. The purpose of this table is to show any variation of route results. Obviously, the one most unlike the other is the Wesleyan version, while the other editions are generally more aligned with the combined result. The one major exception is the location of Kouka whose position provided by the gazetteers was different from the one provided by the Johnson map.

After some initial errors, the reconstructions became more and more similar. Gazetteers definitely aided in journey reconstruction. The addition of the Johnson map further enhanced the displaying of the route. However, not all places were able to be georeferenced, as this next section shall discuss.

4.2 Charts and other perspectives of data

This section covers charts displaying different perspectives of the data and the analysis thereof. This includes the number of coordinates acquired relative to entries, number of coordinates from each source, number of coordinates from each gazetteer, and number of places that are relative or precise references.

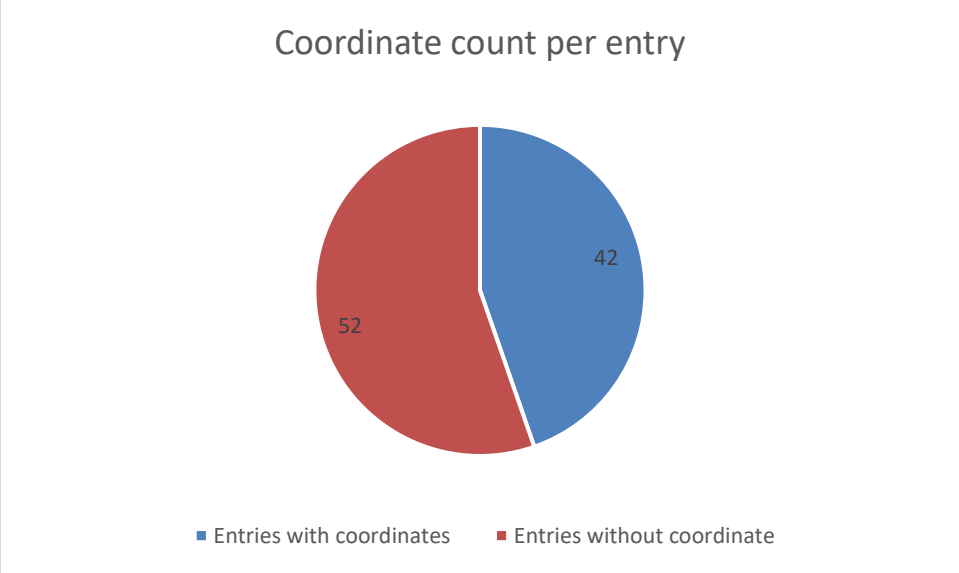


Figure 17:Pie chart of whether and entry could be georeferenced or not.

This pie chart (Figure 17) shows how many locations were georeferenced by either the book or by a gazetteer and those that could not be georeferenced. Altogether, 55.3% of coordinates could not be georeferenced (Keep in mind that there are 94 entries).

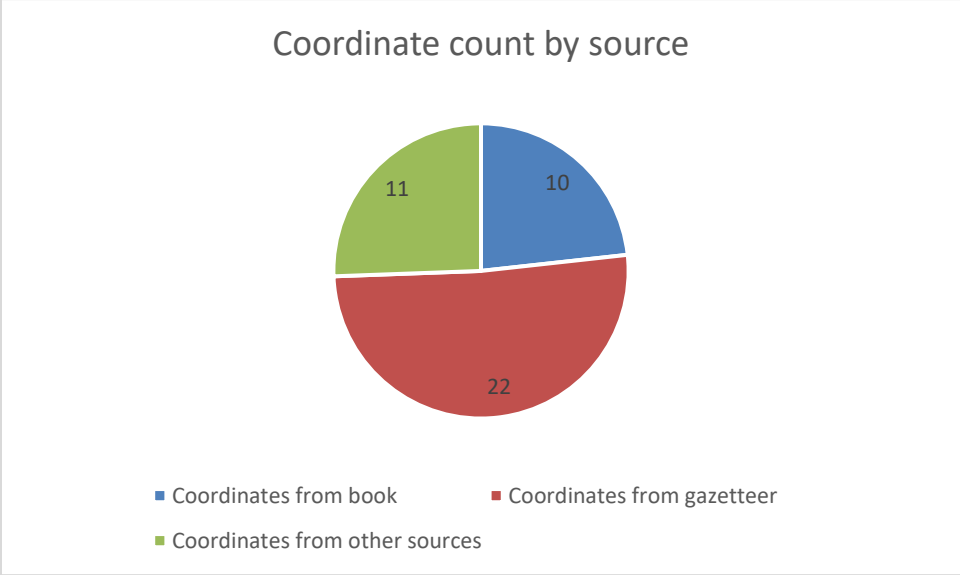


Figure 18: Pie chart of average number of coordinates by origin

Referencing the pie chart in Figure 18, 52.3% of the coordinates came from gazetteer. Only ten complete coordinates came from the book. The remaining quarter of coordinates came from tertiary sources, mainly the Johnson map. So even with a book as geographically detailed as a Jules Verne novel, it is possible for gazetteers and other

period sources to make even greater contributions to reconstructing a journey. There is an extra coordinate here as Kouka got cords from both gazetteer and other. The other coordinate was used but the gazetteer was retained.

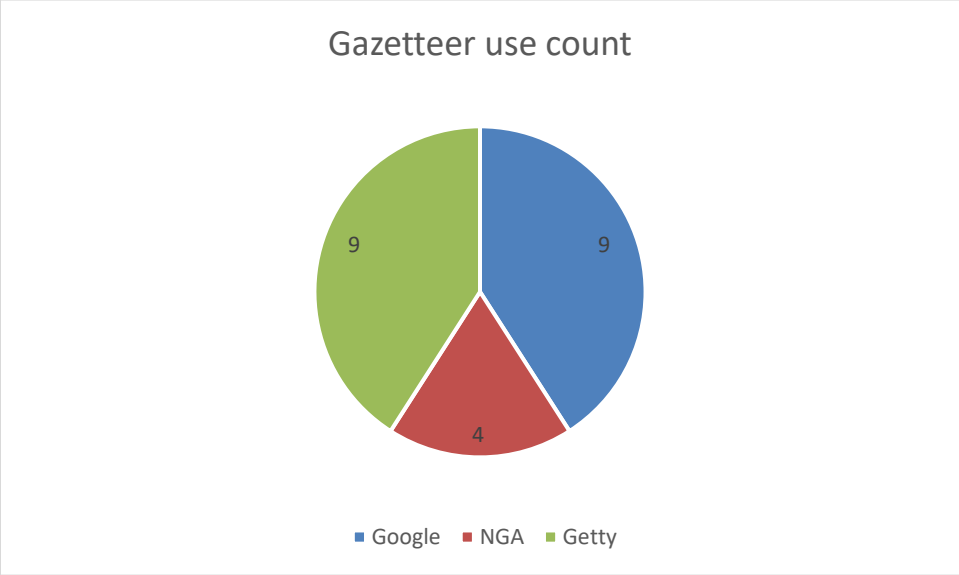


Figure 19: Number of times each gazetteer was used.

Figure 19 shows that three gazetteers were employed to get coordinates based on the place names provided. Of these, Getty and Maplandia were the most useful, each providing nine of the total 22 coordinates (41%) given by gazetteers. NGA contributed the least, only 4 (18%). It may be that the NGA gazetteer is not as well purposed for

georeferencing antiquated place names as its counterparts, which is consistent with its purpose and usage in contemporary geointelligence activities.

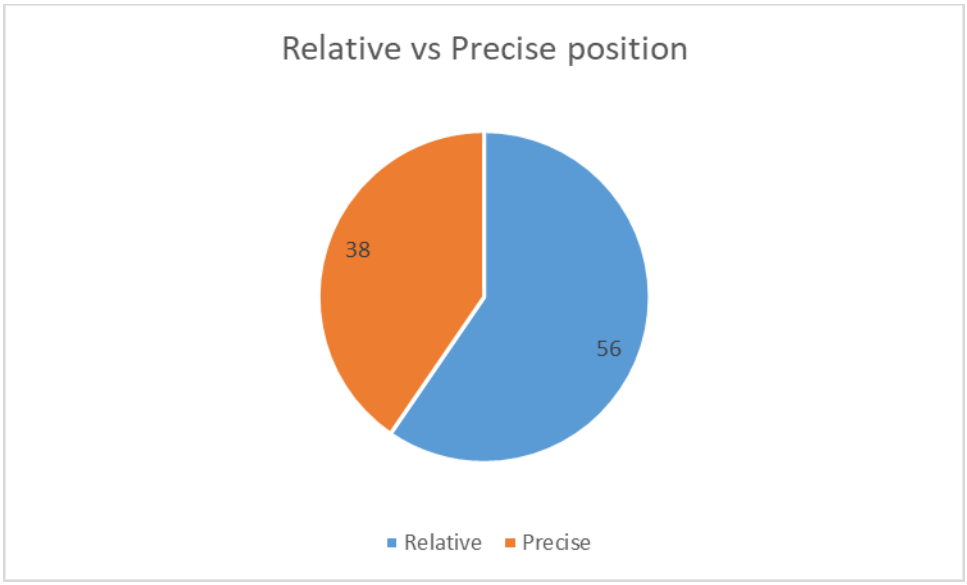


Figure 20: Relative vs Precise references

The majority of references (60% or 56/94) were relative georeferences, as shown in Figure 20, with the remainder (40%) being precise. The majority of places the travelers went through or by were polygon or line features passed through without a specific point given or were point features the travelers came near but did not actually reach. Unless a specific coordinate can be given for line or polygon features, they cannot aid in providing

anchor point. However, knowing the extent of the region or line feature can affect the trajectory of a journey if it is particularly out of the way of the direct route between the two anchor points that the feature is between. If the feature is physically between two anchor points, then the trajectory is unaffected since it can simply pass over it without need for adjustment.

By showing the full extent of the data's nature and not just the data that was mapped, it is shown how much of the data could not be mapped, and thus how much room for improvement there is for mapping archaic literature.

CHAPTER 5: CONCLUSIONS AND FUTURE WORK

Is it possible to map a journey documented with antiquated place names, and if not, what should be done to improve the system? The results of this thesis lead to the conclusion that georeferencing a journey with antiquated names is possible. However, between the book itself and the contributions of three gazetteers, not even half of the locations provided in the book were able to be precisely located with geographic coordinates. And this is in spite of the journey happening less than two centuries ago and in spite of the author being very generous with the supply of geographic coordinates compared to other authors of the time. How then can we map even older journeys? How do we georeference ancient and medieval places from accounts that predate our modern form of measuring the earth? If, at present, no gazetteer can be fully used to locate abandoned places, how then can any old journey or map be georeferenced thoroughly? How can we georeference John Smith's map of the Chesapeake Area or Leif Erickson's voyage to North America? If, by some happy miracle, we got a complete account of Pythia's voyage to northern Europe, how can we reconstruct the journey?

What should be noted is that no one gazetteer was particularly helpful in getting locations compared to the rest, but rather all three had to be relied on to get as many points georeferenced as possible. Even then, most locations, including many towns like Farram, Tounda, and Kernak, places ranging from minor villages to regional capitals,

could not be georeferenced via these gazetteers. This is most likely because these places were abandoned by our time. The Getty Thesaurus of Geographic Names (TGN) gazetteer exists mainly to support Getty as an institution for the arts and literature. As this thesis is centered around a piece of literature, one written by a famous author, Getty should have been capable of providing the locations of these places, but it has not. It was capable of showing the locations of famous sites of antiquity like the ruins of Carthage and Catalhuyuk. NGA's gazetteer is dedicated to NGA's role in geointelligence on behalf of the United States, so it is unlikely that an organization focused on modern, active locations would show interest in extinct locations; at least until the places become active in a way that interests them. Similarly, Maplandia, the gazetteer of Google Maps, which mostly aids the public with travel, would not be interested in abandoned locations that have not been turned into a museum or cultural site. To map antiquated locations, a GIS dedicated to these locations should be established, and resources from past efforts such as the Electronic Cultural Atlas Initiative should be mined to create a temporal gazetteer of historic place names.

5.1 A Gazetteer for the Antiquities

Unless already created, a map project, with a corresponding gazetteer, should be created; one dedicated to archeological sites and abandoned places. When an institution, such as a university or other cultural service, completes the excavation of a site like an abandoned town, they can add its location and information to the map and its gazetteer. Such a map can be used to reconstruct the lost worlds of previous centuries and geographically show both the accomplishments of the archeological field and the regions

that have gaps to fill. The locations can also have images of artifacts found there associated with them, providing an easy means for scholars and enthusiasts to see the distribution of similar cultural artifacts over an area. Such a geographic resource could have benefitted this thesis by providing the locations of now-abandoned locations that were active, even thriving, in pre-colonial Africa where this journey took place. Since a single gazetteer is all one now needs to rely of mapping historic places, this would streamline the georeferencing process. This, in turn, would make automating the process that much easier

5.2 Enhancing the Antiquities gazetteer through period maps

Asides from other modern gazetteers, this geographic resource would benefit greatly by drawing from period maps, like the Johnson map, and the gazetteers associated with them. To create this resource, period maps need to be actively sought out to incorporate their locations into this universal gazetteer. The two most major sources for such maps are the Library of Congress and the David Rumsey collection. The Library of Congress is one of the largest libraries in the world, whose contents includes a comprehensive cartographic collection with 5 million maps catalogued. David Rumsey is an avid map collector who made his collection a public resource via website and by housing his collection in the David Rumsey Map Center in Stanford University. His collection consists of over 150,000 maps from all different sections of the world ranging from the 16th to 21st centuries.

5.3 Mapping Reminiscences

This thesis investigates mapping journeys from the past via literature, but what about mapping journeys that occur before the journey that are talked about during said journey. One future thesis can look into reminiscences. A reminiscence is when someone thinks about the past. In *Five Weeks in a Balloon,* the main character Dr. Ferguson reminiscences often to describe the journeys of various explorers who have ventured through the regions in the years and decades prior to this fictional journey. A future thesis can discuss how to chart journeys discussed by characters during their own journey and develop a program to help distinguish between the main the journey and reminisced journeys. This future work might benefit from the work of Scott McDermott (2017), Robert Sack (as summarized in Holt-Jensen, 2018) and Piatti et al. (2009), all of whom addressed this issue in unique ways.

5.4 Antiquated Journey Georeferencing through People

Five Weeks in a Balloon is based on the journeys of dozens of European explorers who went into Africa. However, European travelers are not the only people with a tale to tell. Every journey is something that can be georeferenced and mapped and every person is a journey. In Table 2, which covers the first few columns of data in the Appendix, there is the entry ‘Medine’ in F92. When they concluded their journey by reaching French troops at Gouina Falls, the travelers proceeded to an outpost called Medina from which they proceeded to the colony of St. Louis and left Africa. Neither the gazetteers nor the Johnson map could provide a position for the place. However, I was given a lead by the books. Apparently, years prior to the journey, the outpost was besieged by a conqueror

and jihadist called El-Hajj (A title for those who completed the sacred hajj pilgrimage to Mecca) Umar Tall. El-Hajj Umar Tall, also known as Omar Saidou Tall was a real person who led a series of conquests in West Africa which included a failed siege of a Fort Medina. Picking up the tip of Omar Tall's life and journey was able to lead me to coordinates for the Medine settlement when the gazetteers and Johnson map could not. Further research into mapping the antiquated journeys of people as they lived their lives is warranted. A historical approach that focuses on 19th century knowledge and thought would be useful for mapping a journey such as the one contained in *Five Weeks in a Balloon*, and connections that might have been made by Jules Verne can be considered when dealing with locations that are otherwise difficult to resolve.

5.5 Improved Realism for *Five Weeks*

Though perhaps not as incredible in its time as the *Nautilus,* the balloon in this story, named *Victoria,* was very well thought out by Verne. In fact, the early chapters of the book describes in detail how the *Victoria* flies for long periods of time and how it can cross Africa without any steering or propulsion. The *Victoria* is both a hot air and hydrogen balloon. Water is electrolyzed into Oxygen and Hydrogen, the latter of which is burned to heat a metal system that circulates the heat into the balloon, which is airtight to contain the hydrogen. The hydrogen absorbs the heat and expands, causing a rise. If the heat is cut off, the hydrogen compresses and the balloon sinks. So as long as the hydrogen is contained and the balloon refuels on water every now and then, the balloon can fly indefinitely. As for crossing Africa, the balloon simply rides the trade winds that generally goes from east to west. Verne created this from the scientific knowledge of his

time, but can we reconstruct this using our more advanced knowledge of the elements and the winds? It would be a fun experiment for an adventurer or myth-buster to try to recreate the *Victoria* as Verne described her and see if it truly is possible to cross Africa east to west using our modern knowledge of the winds, both yearly and seasonal winds. They can also build on what this thesis' data suggests and measure the speed of the balloon and time of the crossing to see if the speed and time it took for the *Victoria* to cross Africa are realistic.

5.6 Discussing the Effect of Elevation on View

One unique feature of *Five Week in a Balloon* is that the vehicle is a flying one which goes back and forth between the ground and high elevations. When one's position changes vertically, their view of their surroundings is very different whether they are on the ground or high in the air. Though this thesis is not centered around the effect of elevation on view, the views taken from the balloon, and the elevation if given were recorded. If someone were to do a thesis on this subject, they can use the data, recorded in the appendix, to aid their research. The purpose of this future work would be an analysis of the style of geographical descriptions, which are certainly different for humans viewing the Earth from above and humans viewing the Earth from the surface. In the former, one assumes a more holistic and complete geographical description would emerge, but with less specific detail. In the latter, more detail of specific features may be given but a synthesis of their joint nature might not be possible. A deep literary analysis tied to location above the Earth's surface would be interesting and worthwhile, but as the

author of this thesis has discovered, this analysis would require some background in literary analysis that this author lacks.

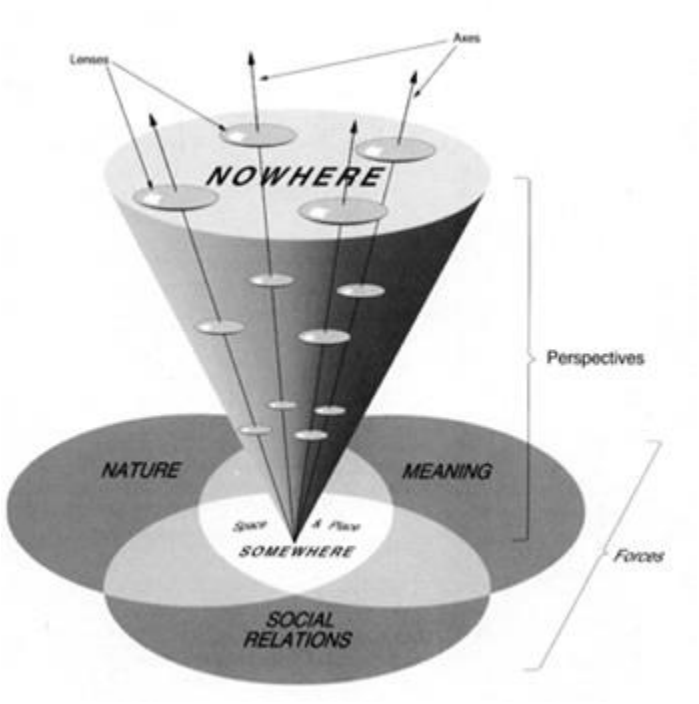


Figure 21: Sack's diagram of place, as presented and reviewed in Holt-Jensen (2018)

This is the Sack diagram (Figure 21). What it explains is how a place is a combination of physical features, social relations, and meaning. Physical features are what the place is made of. Social relations are how the place relates to other places. Meaning is the purpose or function of the place. The more dissociated one is with the three things associated with a place, the less the place means to them, which affects their

perception of the place. When this thesis defines a relative or precise reference, an example could be that ‘in London’ is a precise reference when it is a point in the context of Britain or Europe, but a relative one when it itself is the context, as a polygon. What changed how this thesis would view ‘in London’ as a relative or precise reference was the zoom or elevation, which changed the context. Elevation can affect our view or conception of a place. A future work can look into how a relatively minor change in elevation, such as via balloon, can affect our perception of a large place like a city compared to the perception we have when we are on the ground.

5.7 Final Remarks

Prior to this thesis, it has been well established that mapping literature is possible. This thesis has inquired into whether or not mapping antiquated places is possible. It is possible. But major improvements have to be done, including updating present places association with their past names, and finding the locations of now abandoned places. These improvements and other will require cooperation between the GIS, Literary, and Archeological communities to reach a desirable state and condition by which a future researcher can create a program to automate the charting of journeys from long ago.

APPENDIX

A complete copy of this data table can be accessed at:
https://github.com/egsdorman/Thesis_Data

Table 2: Part 1 of Appendix

Primary	Illustrated	Wesleyan	Odin	Johnson	Other
1	Royal Geographical Society session	Royal Geographical Society	Royal Geographical Society		
2	Travellers Club	Traveler's Club	Traveller's Club		
3	silk factories	factories of Lyon	factories of Lyons		
4	Leith	Leith	Leith		
5	Dr. Fergusson's home	the doctor's humble abode	doctor's modest dwelling		
6	NULL	NULL	null		
7	Greenwich	Greenwich	Greenwich		
8	Cape City	Cape Town	Cape City		
9	Zanzibar	Zanzibar	Zanzibar		
10	Koumbeni island	Koumbeni Island	Koumbeni Island		
11	Mrima	Mrima	Mrima country		
12	Kaole	Kaole	Kaole		
13	Uzaramo country	Uzaramo	Uzaramo		
14	Tounda	Tounda	Tounda		
15	Deje-la-Mhora	Deje-la-Mhora	Deje-la-Mhora		
16	Mt Duthumi	Mt. Dutumi	Mt Duthumi		
17	Zungomero country	Zungomero	Zungomero		
18	Imenge basin	Imenge Basin	Imenge basin		
19	Mt Rubeho	Mt. Rubeho	Mt Rubeho		
20	Ugogo country		Ugogo		
21	Kanyeme basin	Kanyeme basin	Kanyeme basin		
22	Mabunguru	Mabanguru	Mabunguru		
23	Jihoue-la-Mkoa	Jihoue-la-Mkoa	Jihoue-la-Mkoa		
24	Kazeh	Kazeh	Kazeh		
25	plains of Mfuto	Plains of Mfuto	plains o Mftuo		
26	Malagazeri river	Malagarasi River	Malagazeri		

27	Msene	Msene	Msene		
28		mountains of the Moon			
29	Uyofu	Uyofu	Uyofu		
30	Karagwah	Karagwah	Karagwah		
31	null		null		
32		Mt. Rubemhe			
33	Kafuro	Kafuro	Kafuro		
34	null	NULL			
35	Lake Ukereoue	Lake Victoria	Lake Victoria		
36	deserted island	Deserted Island	island on the lake		
37	Nile headwaters	Nile's headwater's			
38		waterfall			
39	Bengal island	Benga Island	Bengal Island		
40	cataracts of Makedo	Makedo Falls	cataracts of Makedo		
41	Mount Longwek	Mount Logwek			
42	Usoga	Usoga	Usogo		
43		cannibal village			
44	null	Null	null		
45	volcano	null	null		
46	null	gold field	null		
47	desert		desert		
48		null	null		
49	oasis	oasis	oasis		
50	Adamova	Adamawa	Adamova	Adamana	
51	Nigritia	Nigritia	Nigritia		
52	Benoue river	Benue River	Benoue River	R. Benue	
53	Yola	Yola	Yola		
54	Bagele		Bagele		
55	Mt Mendif	Mt. Mendif	Mt Mendif		
56	Mosfeia	Mosfeia	Mosfeia	Musfeia	
57	Mandara country				
58	Shari River	Chari	Shari river		
59	Kernak	Kernak	Kernak		
60	Lake Tchad		Lake Tchad		
61	Kouka	Kouka	Kouka		
62	Northern swamp				

63	Farram	Farram	Farram		
64	Tangalia	Tangalia	Tangalia		
65		Lari	Lari		
66			Tibbous		
67	Belad el Djerid	Belad-el-Djerid	Belad el Djerid		
68	Vanishing island	island of Biddiomahs	island of Biddiomahs		
69	snake tree	null	north marsh		
70	null				
71	Damerghou	Damergou	Damerghou		
72	Zinder	Zinder	Zinder		
73	Tagelai	Tagelel	Tagelei		
74	country of Kailouas	country of the Kailouas	land of the Kailouas		
75	Aghades	Agadez	Aghades	Agadez	
76	null	NULL	null		
77	null	NULL	null		
78	Sonray	Songhai Empire	Sonray		
79	Goa	Gao	Goa		
80	Niger River	Niger River	Niger River		
81	Hombori Mountains	Hombori Mountains	Hombori Mountains		
82	Kabra	Kabra	Kabra		
83	Timbuctoo	Timbuktu	Timbuctoo		
84	lake Debo	lake Debo	Niger near Debo		
85	Jenne	Djenne	Jenne		
86	Sego	Sego	Sego		
87	mountains	NULL	mountains		
88	last stop				
89	Falls of Gouina	Gouina Falls	Falls of Gouina		
90		Gouina Outpost			
91	Medina	Medina	Medina		Medine
92	St Louis	St Louis	St. Louis		
93	Portsmouth	Portsmouth	Portsmouth		
94	London	London	London		

Table 3: Appendix Part 2

lat	long	lat from gazetteer	long from gazetteer	Lat from Johnson	Long from Johnson

		55.95	-3.1667		
4	17				
		51.4667	0.0333		
		-33.9333	18.4667		
		-6.1667	39.2		
		-6.25	39.25		
		-6.45	38.95		
	38.33333				
		-6.866667	38.566667		
				-7.083	37.393
		-6.266667	36.866667		
				-5.958	36.017
		-5.0667	32.8167		
-4.28333	32.66667				
				-4.06	31.771
-3.25	29.25				
				-1.331	31.338
-2.66667					
		-1.735278	31.000833		
-1.75					
-0.5	32.86667				
2 40 N					
2					

4.33	27				
4.7	24.25				
4.916667	22.38333				
6.85	19.58333				
8.533333	15.71667				
		7.786	6.7630		
		9.460887	12.174273		
				8.692	14.406
		13.133333	15.533333	12.128	14.576
				13.905	14.217
	10				
		13.8	8.9833333		
				15.616	7.577
16	4.916667				
	2.333333				
		12.6	-2.9		
		15.256837	-1.667038		
		16.716667	-2.983333		

		16.770456	-3.005588		
		15.3132	-4.1008		
		13.9	-4.55		
				12.125	-5
				14.37685	-11.3683
		16.0322	-16.6167		
		50.7667	-1.0833		

Table 4: Appendix Part 3

Combined lat	Combined Long	Coord used	gazetteer used	info	factual, fictional, or both	location type
				Dr. Fergusson and his plan is introduced to the society and is well received	both, Waterloo Place is a real street, not far from the Travelers club, but there does not appear to be a current place with the number three	point
				feast held in Fergusson's honor	factual	point
				received order of silk for balloon	both, factories may or may not have existed but Lyon is a real place	multi point
55.95	-3.1667	Gazetteer	Getty	home town of Dick Kennedy, suburb of Edinburgh	factual	point
					both, home is fictional, but Greek Street and Soho Square are real and adjacent	point
	7	Book		height of Barth's expedition	factual	point
51.46	0.033	Gaze	Gett	port where Resolute	factual	point

67	3	tteer	y	docked		
-33.9333	18.4667	Gazetteer	Getty	Resolute stops to restock coal	factual	point
-6.1667	39.2	Gazetteer	Getty	town on island of same name, hostile natives force move to Koumbeni island	factual	point
-6.25	39.25	Gazetteer	Google	balloon is built, christened, and launched	factual?	point
				stretch of east African coast	factual	area
-6.45	38.95	Gazetteer	Google	hostile town	factual	point
				fertile land, hostile natives	factual	area
				village	factual	point
				village where explorer Maizan was tortured and executed	factual	point
-6.866667	38.566667	Gazetteer	Google	mountain of Ourizara range	factual	point
-7.083	37.393	Other		damp, unhealthy region	factual	area
					factual	area
-6.266667	36.866667	Gazetteer	Google	part of the third and highest range of the Usagara system	factual	point
-5.958	36.017	Other			factual	area
					factual	area
				stony country	factual	area
				large stone outcrop with nearby water deposits, balloon stops to refill on water	factual	point
-5.0667	32.8167	Gazetteer	Getty	trade hub, initially receptive but turns hostile	factual	point
-4.28333	2.66667	Book			factual	area
				tributary of Lake	factual	line

				Tanganyika		
-4.06	31.771	Other		cluster of villages, storm strikes	factual	point
				semicircle range on lower tip of Tanganyika	factual	area
-3.25	29.25	Book		northern limit of Unyamwezi country	factual	point
-1.331	31.338	Other		relatively civilized region	factual	area
				elephant killed	fictional	point
					factual	point
-1.735278	31.000833	Gazetteer	NGA	trade hub, Lake Ukereoue spotted	factual	point
					fictional	point
				source of the Nile	factual	area
-0.5	32.86667	Book		high mosquito population	fictional	point
					both, Nile's headwaters exist, but not like the book imagines	point
				found by Debono	factual	point
				farthest point of explorer Andrea Debono	factual	point
					factual	point
				the trembling mountain	factual	point
				home to Nyam-Nyam cannibal tribes	factual	area
				evidence of cannibalism present	fictional	point
4.33	27	Book		end of a 150 mile day trip, missionary rescued	fictional	point
4.7	24.25	Book		near a volcano	fictional	point
4.916667	22.38333	Book		Rescued missionary buried here, site of gold field	fictional	point
					fictional	area
6.85	19.58333	Book		balloon runs out of water	fictional	point
8.533333	15.71667	Book			fictional	point
				abundant wildlife, farthest	factual	area

				reach of Barth		
				home to Arab Chouas herdsmen, Atlantika Mountains seen on horizon.	factual	area
7.786	6.7630	Gazetteer	Getty	Niger tributary	factual	line
9.460887	12.174273	Gazetteer	NGA	city Victoria comes within 40 miles of	factual	point
				cluster of 18 villages	factual	point
				separates Niger and Chad watersheds	factual	point
8.692	14.406	Other		city between two mountains, sultan startled by Fergusson's Arabic	factual	point
				fertile forest	factual	area
				tributary of Lake Chad	factual	line
				capital of Loggoum country, locals try to destroy the balloon with incendiary pigeons.	factual	point
				Lake Chad, approached from southern shore, a crossroads of winds, Joe throws himself into the lake to save Victoria from condors	factual	area
12.128	14.576	Other	Google	capital of Bornou	factual	point
				Fergusson and Kennedy recuperate and remove torn outer layer of the balloon, after a search, they return to this point, but an anchor gets stuck, it is disconnected and saves Joe from quickmud	factual	point
				capital of Biddiomahs. Fergusson and Kennedy pass by it looking for Joe	factual	point
				Fergusson and Kennedy pass by it looking for Joe	factual	point
13.90	14.21	Othe		village	factual	point

5	7	r				
				Fergusson and Dick blown over by hurricane	factual	area
				desert of briars that border Soudan country	factual	area
				Joe swims here and is captured by natives, flees the island when the tide comes in	fictional	point
				Joe rests here unaware of his new bedfellows	fictional	point
				crossed in the evening after rescuing Joe	factual	line
				fertile undulating country	factual	area
13.8	8.983 3333	Gaze tteer	Goo gle	town with famous execution square and gallows tree	factual	point
				village reached by Barth	factual	point
				vegetated mountains neighboring Toureg territory	factual	area
15.61 6	7.577	Othe r		town in decline, visited by Barth	factual	point
16	4.916 667	Book		end of May 17 leg	fictional	point
				passed at end of May 18	fictional	line
				fertile land with lots of rain	factual	area
12.6	-2.9	Gaze tteer	Goo gle	petty town, former capital, reached by Barth	factual	point
				major waterway of West Africa	factual	line
15.25 6837	- 1.667 038	Gaze tteer	NG A	alien in appearance, seen at night, mesas not unlike those in Monument valley	factual	area
16.71 6667	- 2.983 333	Gaze tteer	Goo gle	port town of Timbuctoo	factual	point
16.77 0456	- 3.005 588	Gaze tteer	NG A	triangular famous city of West Africa, in decline, aerial view is a city of billiard balls and thimbles	factual	point
15.31	-	Gaze	Gett	lake crossed on the 21	factual	area

32	4.1008	tteer	y			
13.9	-4.55	Gazetteer	Google	city on large island, commercially active	factual	point
12.125	-5	Other		capital of Bambarra	factual	point
				splits Niger and Senegal watersheds	maybe	area
				Victoria stops for the night and the heating mechanism is disposed, is ambushed by local warlord	fictional	point
14.015	-11.1025	Other		Victoria reflated with hot to send her over the Senegal river where she is spotted by a Frenh patrol	factual	point
				French outpost	fictional	point
14.37685	-11.3683	Other		French outpost, reached by foot.	factual	point
16.0322	-16.6167	Gazetteer	Getty	French colonial capital of Senegal, reached by steamboat	factual	point
50.7667	-1.0833	Gazetteer	Getty	English port city, reached by frigate, they reach London the next day	factual	point
				English capital	factual	point

Table 5: Appendix Part 4

relative or precise	relationship with other locations	address	chapter	month	day or last day	Journey day	year
precise	in London	3 Waterloo Place	1	1	14		1862
precise	in London, on Pall Mall		1	1	14		1862
relative	in Lyon		2				
precise	near Edinburgh		3				
precise	on Greek Street near Soho square		3				

precise	somewhere in West Africa		4				
precise	near London		8	2	21		1862
precise	Cape of Good Hope, near Table Mountain		9	3	30		
precise	off coast of East Africa		11	4	15		
precise	off coast of Zanzibar		11	4	18	0	
relative	across channel from Zanzibar		12	4	18	0	
relative	in Mrima		12	4	18	0	
relative	in East Africa		12	4	18	0	
relative	in Uzaramo		12	4	18	0	
relative	east of Ourizara range		12	4	18	0	
relative	west of Deje-la-Mhora		12	4	18	0	
relative	west of Ourizara mountains		13	4	19	1	
relative	east of mt Rubeho		13	4	19	1	
relative	eastern border of Ugogo		13	4	19	1	
relative			13	4	19	1	
relative			14	4	19	1	
relative			14	4	20	2	
precise			14	4	20	2	
precise			14	4	20	2	
precise			16	4	20	2	
relative			16	4	20	2	
relati			16	4	20	2	

ve							
relative	south of Lake Tanganyika		17				
relative			17	4	21	3	
relative	west of Lake Ukereoue		17	4	21	3	
precise	in Karagwah		17	4	21	3	
relative			18				
relative	in Karagwah		18	4	22	4	
precise			18	4	22	4	
relative			18	4			
precise	in Lake Victoria		18	4	22	4	
relative			18	4	23	5	
relative			18	4	23	5	
precise	on the Nile		18	4	23	5	
relative	on the Nile		19	4	23	5	
relative			19	4	23	5	
relative	west of Mount Longwek		19	4	23	5	
relative			20				
precise			20				
precise			22				
precise	west of the Nile, uncharted territory		23				
relative	on May 1st, about 300 miles from Gulf of Guinea, May 3 balloon runs out of water		24	4	28/29/30/1/2/3/4/5/6	10	
preci			26				

se							
precise			28	5	7	19	
relative			29	5	9	21	
relative	west of Adamova		29	5	9	21	
relative	in Nigritia		29	5	9	21	
relative			29	5	9	21	
relative			29	5	10	22	
relative			29	5	10	22	
relative			30	5	11	23	
relative			30	5	11	23	
relative			30	5	11	23	
precise			30	5	11	23	
precise			31	5	12	24	
relative			32	5	12	24	
precise	northern shore of Lake Chad, between villages Lari and Ingemini		33	5	12/13/14/15	24	
relative	in Lake Chad		33	5	14	26	
relative	on eastern shore of Lake Chad		33	5	14	26	
relative	Northern Shore of Chad		33				
relative	north of Lake Chad		34				
relative	west or northwest of Lake Chad		34	5	15	27	
precise	6 miles from condor attack, on lake Chad		35	5	12/13/	24	
precise	northern shore of Lake Chad, between villages Lari and Ingemini		35	5	13	25	

precise			36	5	15	27	
relative	west of Lake Chad		37	5	16	28	
relative	in Damerghou		37	5	16	28	
relative	in Damerghou		37	5	16	28	
relative			37	5	16	28	
precise			37	5	16	28	
precise			38	5	17	29	
precise			38	5	18	30	
relative			38	5	19	31	
relative	on the banks of the Niger		38	5	19	31	
relative			38	5	19	31	
relative			39	5	19	31	
relative	on banks of the Niger, north of Hombori mountains		39	5	20	32	
relative	near Kabra		39	5	20	32	
relative	near Niger		40	5	21	33	
relative	on the Niger		40	5	21	33	
relative			40	5	22	34	
precise			41	5	27	39	
precise	25 miles from Senegal		42	5	27	39	
precise	on the Senegal River		43	5	28	40	
precise	near Senegal River		44				
preci	in French Senegal		44	5	27?		

se							
precise	in French Senegal		44	6	10	53	
precise	England		44	6	25		
precise			44	6	26		

Table 6: Appendix part 5

time	am/pm	view	elevation	foreign
				1
				2
				3
				4
				5
				6
5	AM			7
				8
11	AM			9
		All of Zanzibar could be seen, eventually only the ship's cannon salute could be heard, people looked like insects	1500	10
		Coast consisted of mangroves and sand dunes. Summit of Mt Nguru seen in distance	300	11
		Terrified Natives fired arrows at balloon in vain		12
12	PM	Fertile land dotted with coconut palms, papaya trees, and cotton plants. Jackrabbit and quail spotted		13
		Occasional caravan spotted taking shelter from the heat in kraals		14
		Giant baobab trees spotted	600	15
630	PM		3000	16
		Balloon had to go above clouds because Kennedy was sick, clouds described as tumbling and highly reflective of light	4000	17
11	AM	Hills with scattered tribes who threatened the Victoria	600	18
		Lack of oxygen blurred vision, animals and people were invisible, roads were shoelaces, and lakes were fishponds. The mountains were arid with snowy caps	6000	19
		wilderness area with thickets		20
		10 mile wide clearing, with villages in the midst of Baobab and calabash trees, huts look like haystacks	1500	21
		Covered in humpbacked crags and cobbles made of Syenite, with		22

		occasional buffalo and elephant bones, few trees present		
		A large circular rock with ponds and deserted villages on the west side. Gas burner used constantly due to high increase in ground elevation	3000	23
2	PM	Six villages with a market center and merchants all over the region trading shark teeth, ivory, slaves, glass beads, cotton, cannabis, and other things.		24
		fertile, rolling plains		25
		Big humped livestock graved in meadows, forests gave shelter to lions, leopards, and hyenas, sometimes a thicket could be heard be trampled by an elephant		26
9	PM	large number of villages, regular ditches, shapes of palm trees, tamarinds, sycamores, and spurges seen		27
				28
12	PM			29
				30
				31
				32
		Fifty circular huts with blooming thatch roofs, Lake Victoria on the horizon		33
12	PM			34
		Elevation listed is of lake, mesas of Uganda and Usoga seen in the west	3750	35
6	PM			36
				37
		Waterfall overlooks a river basin dotted with islands	2500	38
				39
				40
				41
				42
				43
				44
				45
				46
			500	47
				48
				49
		Wild country with a heavy presence of creatures like wild ox, hippos, and elephants		50
		Shuwa Arabs seen tending to flocks, summits of Atlantika Mountains seen on the horizon	7800	51
				52

		Mt. Mendif seen beyond city, slaves seen working in the field raising sorghum		53
				54
3	PM		8000	55
9	AM	Large town between two mountains, only accessible by road between a marsh and a forest		56
				57
				58
				59
9	AM			60
		Actually two towns, one rich one poor, separated by a boulevard, little activity going on as the town is not a merchant or manufacturing hub		61
				62
11	AM			63
230	PM			64
5	PM			65
				66
		Lots of brambles, then a desert with many signs of caravan action		67
				68
				69
				70
				71
		Gallows tree in the center of town with a hangman at its side		72
				73
		giraffes, antelopes, ostriches seen running among acacias, mimosas, souahs and date palms		74
10	PM			75
				76
				77
		Few hills, home to guinea fowl, snipe, gators		78
12	PM	Squalid shacks, used to be metropolis		79
		Wide but fast moving		80
8	PM	Seen at night, looked like an abandoned castle under moonlight		81
		Low houses on the curves of the river long lines of donkeys and camels seen		82
2	PM	A huge triangle, pointing north, on white sand, little vegetation seen, the layout of buildings looks like a pile of marbles and dice, one story houses with narrow roads or huts of straw and reeds, no women seen, three mosques seen including Sankore mosque at peak of triangle and Sidi Yahya Mosque in the Sanegungu quarter, ramparts in poor condition		83

		large islands broke up the river	84
		Two towered mud mosques seen with millions of swallow nests on the walls that can be smelled. Town is always busy, some trees seen between houses	85
4	AM	Actually four towns with Moorish mosques and bust ferryboats	86
9	AM		87
			88
3	PM		89
			90
			91
			92
			93
			94

REFERENCES

Aburizaiza, A. O., & Rice, M. T. (2016). Geospatial footprint library of geoparsed textfrom geocrowdsourcing. Spatial Information Research, 24(4), 409-420.

Bekisz, M. (2014). Discovering the Place Where all They had Ended: a Study in Holocaust Toponym Georeferencing and Spatial Association. George Mason University

Bushell, S. (2012). The Slipperiness of Literary Maps: Critical Cartography and Literary Cartography. Cartographica 47:3, pp. 149–160

Camelli, F., Lien, J. M., Shen, D., Wong, D. W., Rice, M., Löhner, R., & Yang, C. (2012). Generating seamless surfaces for transport and dispersion modeling in GIS. Geoinformatica, 16(2), 307-327.

Carroll, A (2006) "A Case-Study-In-Progress: How a Media Organization Tackles the Georeferencing Challenge/Opportunity", Proceedings of the Digital Gazetteer Research and Practice Workshop, Nov. 17, 2006.

Cave, A. (2016). Reconstructing Paths of Travel Based on Location References in Unstructured Text. George Mason University

Goodchild, M. F., Kyriakidis, P., Rice, M., & Schneider, P. (2005). Report of the NCGIA Specialist Meeting on Spatial webs. https://pdfs.semanticscholar.org/297d/fb6ef5e310aff8cf72d97335eb45e5deb36d.pdf

Goodchild, M., & Hill, L. L. (2006). Digital Gazetteer & Practice Workshop, Summary Report. https://escholarship.org/uc/item/6fr9g5nv

Hill, L.L. (2008) "Gazetteers." Encyclopedia of Geographic Information Science, edited by Karen K. Kemp, SAGE Publications, 2008, pp. 159-161. Gale eBooks, https://link.gale.com/apps/doc/CX2860900083/GVRL?u=viva_gmu&sid=GVRL&xid=b969ca05. Accessed 2 June 2020.

Holt-Jensen, A. (2018) "Geography: History and Concepts" Sage Publications Ltd.

Janowicz, K. (2006). Similarity-based Identity Assumptions for Historical Places. Digital Gazetteer & Practice Workshop

Kemp, K. (2008). Georeference. In K. K. Kemp (Ed.), Encyclopedia of geographic information science (pp. 202-202). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781412953962.n89

Kreitner, R, and Melendez, S (2015). "The Obsessively Detailed Map of American Literature's Most Epic Road Trips", July 20, 2015, Online blog post for Atlas Obscura, https://www.atlasobscura.com/articles/the-obsessively-detailed-map-of-american-literatures-most-epic-road-trips, last accessed July 29, 2020

McDermott, S. (2017). Frequency and Proximity Clustering Analyses for Georeferencing Toponyms and Points-of-Interest Names From a Travel Journal. George Mason University

Murrieta-Flores, P., Donaldson, C. & Gregory, I. (2017). GIS and literary history: Advancing digital humanities research through the spatial analysis of historical travel writing and topographical literature. Digital Humanities Quarterly. 11 (1)

Piatti, B., Reuschel A., & Hurni L. (2009). Mapping Literature: the Prototype of "A Literary Atlas of Europe". Institute of Cartography, ETH Zurich; and *Proceedings of the 24th International Cartographic Conference,* Santiago, Chile, November 2009.

Phillips, M. (2019). Estimating Toponym Content of Social Media Data of the Northern Triangle. George Mason University

Rice, M. T., Hammill, W. C., Aburizaiza, A. O., Schwarz, S., & Jacobson, R. D. (2011). Integrating user-contributed geospatial data with assistive geotechnology using a localized gazetteer. In Advances in Cartography and GIScience. Volume 1 (pp. 279-291). Springer, Berlin, Heidelberg.

Rice, M. T., Aburizaiza, A. O., Jacobson, R. D., Shore, B. M., & Paez, F. I. (2012). Supporting Accessibility for Blind and Vision-impaired People With a Localized Gazetteer and Open Source Geotechnology. Transactions in GIS, 16(2), 177-190.

Rice, M. T., Jacobson, R. D., Caldwell, D. R., McDermott, S. D., Paez, F. I., Aburizaiza, A. O., ... & Qin, H. (2013). Crowdsourcing techniques for augmenting traditional accessibility maps with transitory obstacle information. Cartography and Geographic Information Science, 40(3), 210-219.

Roke (n.d.) "Map of the trip described in the book from the east to the west coast of Africa", from Wikipedia article: Five_Weeks_in_a_Balloon, image name: Five_Weeks_in_a_Balloon.jpg, uploaded by user Roke on December 6, 2006. Last accessed: July 29, 2020

Tobler W. R. (1970) "A Computer Movie Simulating Urban Growth in the Detroit Region", Economic Geography, 46(Supplement): 234-240

Library of Congress Geography and Maps Division, https://www.loc.gov/rr/geogmap/

David Rumsey Collection, https://www.davidrumsey.com/

BIOGRAPHY

Eric Dorman graduated from West Springfield High School, Fairfax, Virginia, in 2014. He received his Bachelor of Science and an undergraduate certificate in GIS from the University of Mary Washington in 2018. He is pursuing a Master in Geographic and Cartographic Science and a graduate certificate in GIS from George Mason University.