Analyzing Historical Data: Historical Ship Route Finder

Cayce Egan

May 12, 2025

Introduction

The world of historical scholarship has recently reached somewhat of a fork in the road as of late. With the ever constant progress of technological advancements, the possibilities of the implementation of these technologies has become a point of discussion among historical scholars. Some scholars, as described by Antonella Di Marzio, believe that the implementation of newer fields of study, notably Data Science, can be used to improve both historical research and improve the accessibility of historical information to the general public. However, in reality, only recently in the last couple decades have large amounts of information been digitized into datasets being true for some sources of data, including some sources of shipping information such as ship logs.

The goal of the project is to present a dynamic visualization of a map of the Atlantic Ocean, with which a user could select a ship and see the likely route it has taken. The project includes Trans-Atlantic crossings and local travels. Trans-Atlantic crossings are relatively easy to predict the route of as almost all crossings took the same route due to the fact that the winds the ships relied on the fact that the winds followed a consistent, clockwise circle around the Atlantic, as well as navigational tools from the time heavily relying on finding one's relative position to a known location, such as land meaning, ships are going to hope to take a path that encounters a lot of islands or remain just off shore for as long as possible, with westward ships going south

between island chains and eastward ships traveling north off of the North American Coast. See Figure 1, for a mapping of this phenomenon.

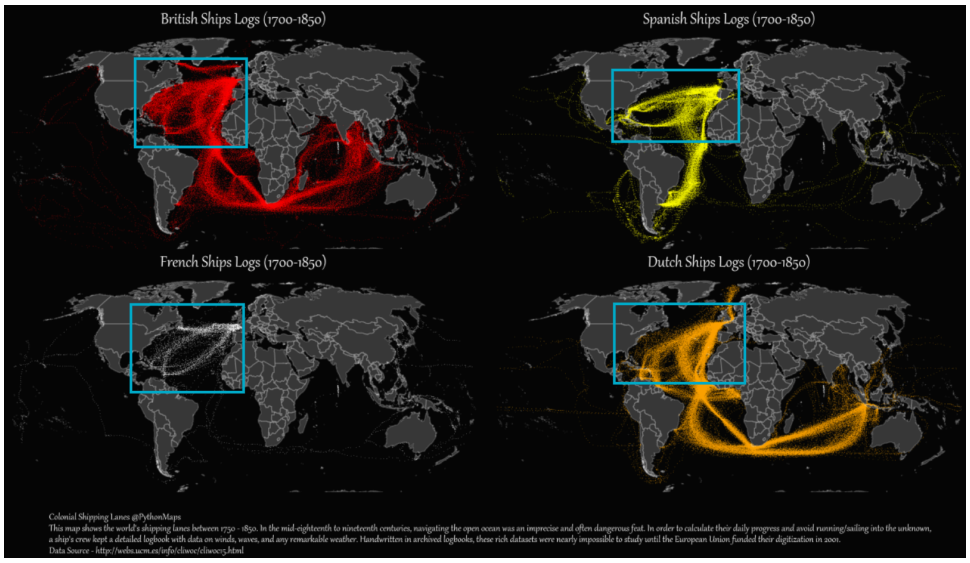

Figure 1. *Symington, Adam. "Colonial Shipping Lanes." Each dot on the chart represents a recorded position. (Note: A blue square has been added to the original visualization to better highlight the circular paths of ships at the time across the Atlantic Ocean)*

Non-Trans-Atlantic travels also maintain consistent routes, proven to be the most reliable with the tools and information of the time.

The project will hopefully serve as a demonstration of the usefulness of Data Science in historical scholarship, being an example of using Data Science to make a project that allows historical information to be easily accessible and understandable by public users.

Data

The main data this project relies on comes from the *Global Americas Shipping Database,* specifically, the dataset entitled "New-England Palladium Maritime Information 1803, 1811." The data has been compiled by Professor Fabrício Prado along with other contributors and

details recordings from ports across the Americas of many ships either arriving or departing, with data gathered from local primary sources such as newspapers. A link to this database is included in the Bibliography.

A little bit of preprocessing was necessary for the project, which mostly included creating more uniform formatting. Preprocessing was done in R. See Figure 2 for a sample of the database focusing on New England.

New-England Palladium Maritime Information

Ship Name	Origin	Declared Destination	Year	Captain	Cargo	Origin Lat	Origin Lon	Dest Lat	Dest Lon
Dispatch	Savannah	Boston	1803	Adams	Georgia upland cotton	32.079007	-81.092134	42.35543	-71.06051
Rufus	Savannah	Liverpool	1802	Holland	NA	32.079007	-81.092134	53.40720	-2.99168
Hannah	Boston	St. Croix	1803	Twycross	NA	42.355433	-71.060511	17.72912	-64.75892
Polly	Cape Francoise	New York	1802	Kennard, of Portsmouth	coffee; lumber	19.759524	-72.200807	40.71273	-74.00602
Polly	Martinico	Boston	1803	NA	24 hhds. high proof rum	46.124420	9.277049	42.35543	-71.06051
Tarantula	Demerara	Boston	1803	Taylor	sugar; 32 puncheons excellent rum	5.629489	-58.393776	42.35543	-71.06051
Good Intent	Philadelphia	Boston	1803	Rateliss, of Portland	a few hhds. First quality loaf and lump sugar; a few kegs fresh and new buck wheat meal	39.952724	-75.163526	42.35543	-71.06051
Hazard	Alexandria	Boston	1803	Luce	600 bbls. Alexandria superfine & fine flour; 60 half bbls. Alexandria superfine flour	31.199181	29.895172	42.35543	-71.06051
Polly	Boston	Martinique	1803	Appleton	NA	42.355433	-71.060511	14.61137	-60.96208
Hope	Boston	Jamaica	1803	Hewes	NA	42.355433	-71.060511	18.18505	-77.39477

Figure 2. *Prado, Fabrício. "New-England Palladium Maritime Information 1803, 1811" (Note: Table was made in R using the data gathered by Prado and represents the first ten observations out of 2,034.)*

Figure 2 shows the first ten observations of the data, post preprocessing. Previously, the "Port" variable listed the port city from which the record of the observation originates, with the "Arrival/Departure" variable showing whether the ship being recorded is arriving into the port, or departing. If the ship is arriving, the "Origin" variable shows the port the ship traveled from, and if the ship is departing, the "Declared Destination" variable showing the port the ship is said to have been leaving for at the time of departure. However, the "Arrival/Departure" variable was

removed in order to make the data easier to process, with the information being placed into the "Origin" and "Declared Destination" variables. It is important to note that a declared destination does not necessarily mean the ship made it to that destination, as it is possible that complications or a simple change of plans occurred after departing from the port.

The "Year" variable represents the year in which the voyage took place. The "Captain" variable is presumed to represent the surname of the captain of the ship, and sometimes includes titles. Meanwhile the "Cargo" variable represents any recorded cargo during this particular voyage. The multiple coordinate columns represent the coordinates of the corresponding origin and declared destination gathered from Nominatim, a geography based API that deals with locations and information such as addresses, city names, and coordinates. However, difficulty arised, as only city names are present in the dataset, and there are many cities that share names, creating problems with the Nominatim API. The API requests had to be run with replacements, specifying which city to gather the coordinates for. As the specific cities were not present in the dataset, the decision was made based off of the largest seaport.

Methods

The end goal of the project was to have a dynamic visualization that can show the likely route or routes a ship has taken on its journey. This was done using the Dash framework, from Plotly, to build an application. While the project was originally only going to cover Trans-Atlantic Crossings, a focus was put on any Atlantic related trips in order to expand the scope of the project.

By default, Dash draws the line directly between the two presented points, creating lines that went over land and did not follow the laid out paths ships commonly took. A couple ideas

were played with, including at one point a grid of points over the ocean that the line had to stay on. Ultimately however, a series of "waypoints" were assigned to each port detailing the port's access route in and out of the Atlantic Ocean. Therefore, the entire route did not need to be manually created for each combination of ports, but rather the code combined parts of each port's predetermined route. See Figure 3 for an example of the user interface.

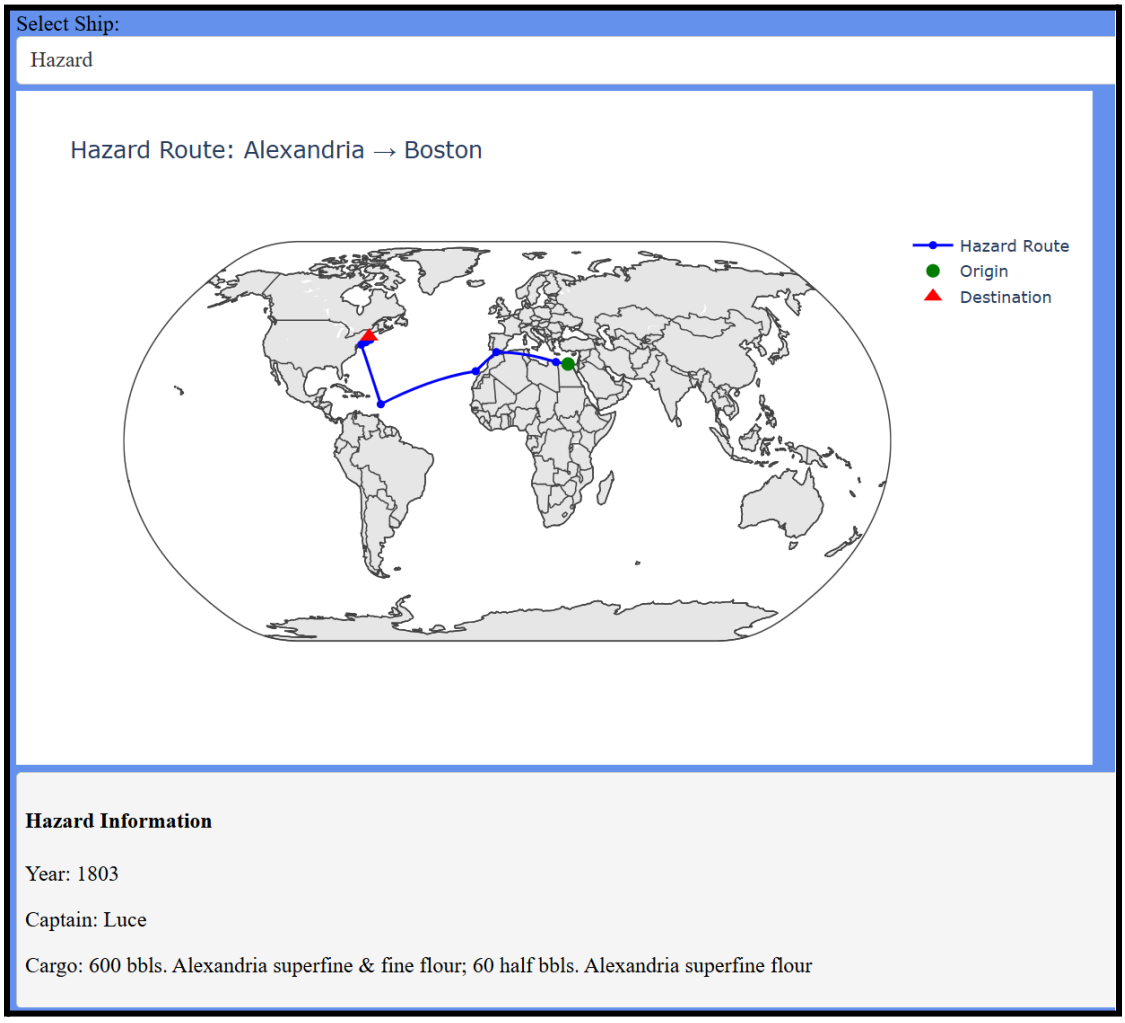

Figure 3. *The User Interface, showing the likely path of the Hazard from Alexandria to Boston.*

The same process of waypoints was used to make sure that Trans-Atlantic crossings follow the correct routes. As previously described in Figure 1, westward ships traveling from

Europe would go south to the Canary Islands, off the coast of Africa, before maintaining that latitude on the westward trip to the Caribbean, before then going either north or south to their final destination. The westward lines needed to represent that. This was achieved by creating a cutoff point along the 30 degrees West Premidian, which cuts straight through the Atlantic Ocean. Ships traveling directly from the east of this line to the west were required to take the extra waypoints at the Canary and Caribbean Islands.

Conclusion

The project should be used to easily allow interested users to look up the common routes taken across the Atlantic between different ports, and learn a little bit about the ships that made these journeys.

The project aims to provide a demonstration of the positive use of Data Science to make the humanities more accessible and improve historical scholarship. Data Science can sometimes be seen as a scary thing among scholars of the humanities, often being associated with Artificial Intelligence, which is often heavily criticized, and buzz words such as “data” itself, which often bring up scary concepts such as data stealing and data leaks. Hopefully this project shows a positive side of data science that can show its potential for pushing humanity scholarship forward.

Bibliography

Sources

Marzio, Antonella Di. "STEMming from history: W&M lab uncovers global stories and fosters career skills." *W&M News*. Published: April 4, 2024 https://news.wm.edu/2024/04/04/stemming-from-history-wm-lab-uncovers-global-stories-and-fosters-career-skills/

Prado, Fabrício. *Global Americas Shipping Database*. https://sites.google.com/email.wm.edu/global-americas-shipping-datab/home

Symington, Adam. "Mapped: European Colonial Shipping Lanes (1700–1850)." *Visual Capitalist*. Published: May 26, 2023 https://www.visualcapitalist.com/cp/colonial-shipping-lanes/

Libraries

Dash: https://dash.plotly.com/

GeoPy: https://geopy.readthedocs.io/en/stable/

Nominatim: https://nominatim.org/release-docs/develop/