

# QUANTIFYING THE SLAVE TRADE THROUGH SHIP LOGS

Alicia Clark  
Department of  
Mechanical Engineering

Emma Cotter  
Department of  
Mechanical Engineering

Zehua Wei  
Department of  
Chemical Engineering



CSE 599B  
Software Engineering  
for Data Scientists  
Winter 2016



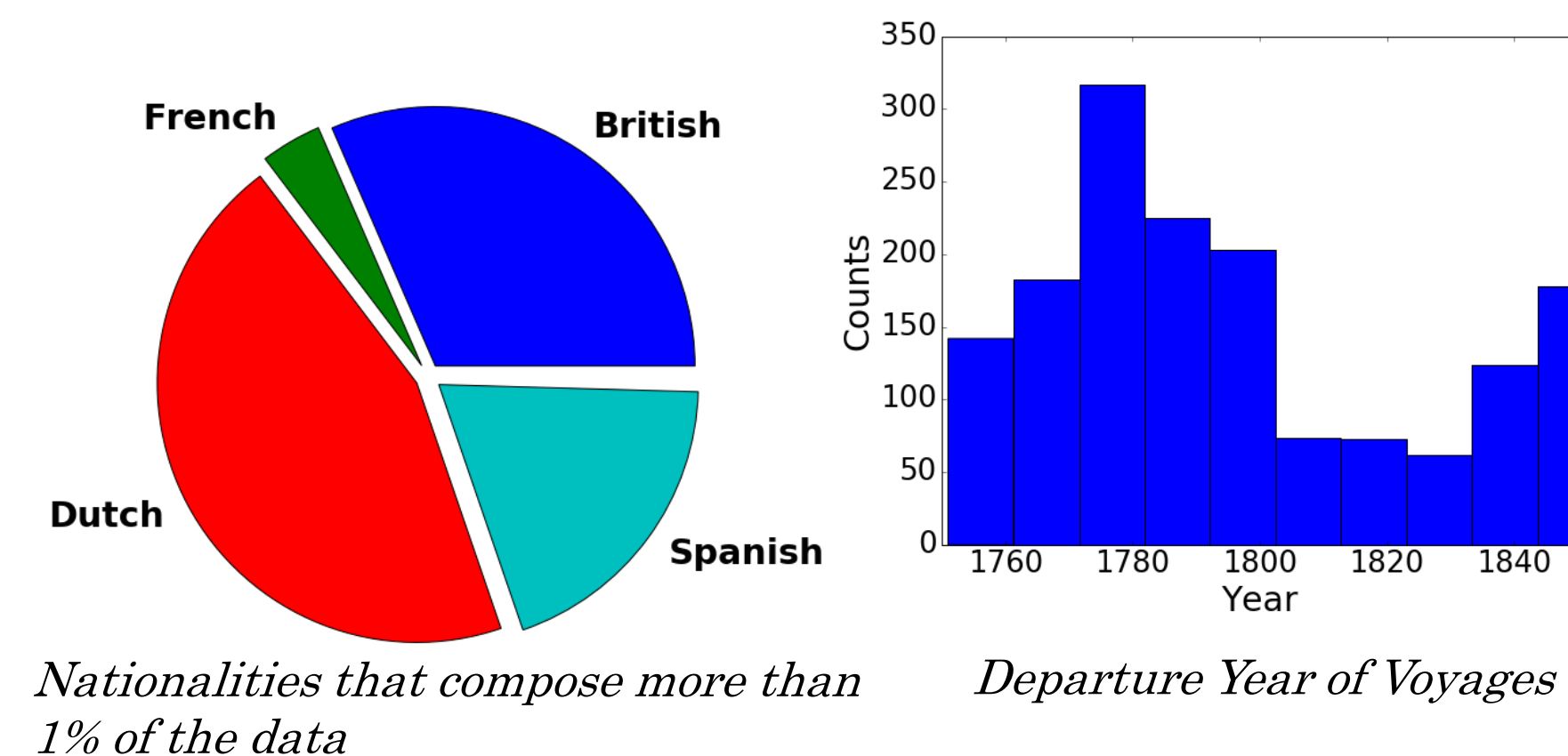
## MOTIVATION

- 17<sup>th</sup> and 18<sup>th</sup> century ship logs contain extensive information about weather, historical events, and cultural phenomena
- Many of these ship logs have been translated into databases for historical or scientific research
- Databases are typically formulated with one research goal in mind – i.e. studying climate change or trends in the slave trade
- Our goal is to find out if information from a database of ship logs collected to study climate change can be used to also study the slave trade

## THE DATA SETS

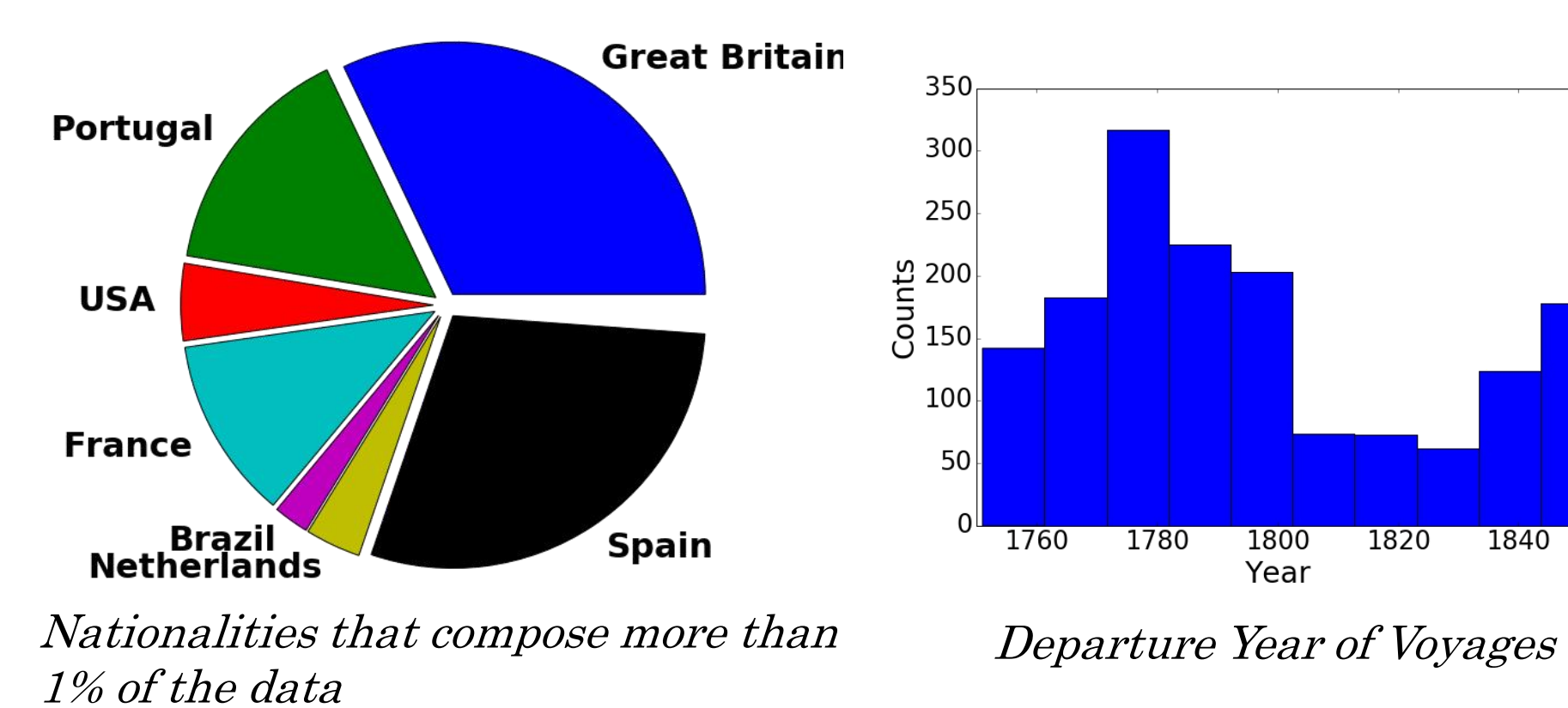
### Climatological Database for the Worlds Oceans (CLIWOC)

- Database focuses on understanding climate change through weather observations
- Contains raw logbook text
- 280,290 transcribed logs
- 1651 individual voyages
- Includes latitude and longitude for every log entry



### Trans-Atlantic Slave Trade Database

- Database focuses on quantifying the slave trade
- Includes entries for each voyage (not every individual log entry)
- 58,957 voyages documented



## DATA CLEANING

### 1. Clean CLIWOC Data

- Removed repeating logbook entries so that there was only one entry per voyage
- Isolated the columns of interest to be used in the classification
- Used fuzzy logic to rename strings to their corresponding 'match' in the slave trade data set

### 2. Clean Trans-Atlantic Slave Trade Data

- Explored the different columns to find columns that were also in the CLIWOC data
- Isolated the columns of interest to be used in the classification

### 3. Join two data sets

- Joined the two data sets and cleaned
  - converted all strings to lower case
  - removed trailing whitespace
  - matched column names

### 4. Encode data for classification

- Encoded the data using either a label encoder or a one hot encoder depending on the classification algorithm used

## CLASSIFICATION

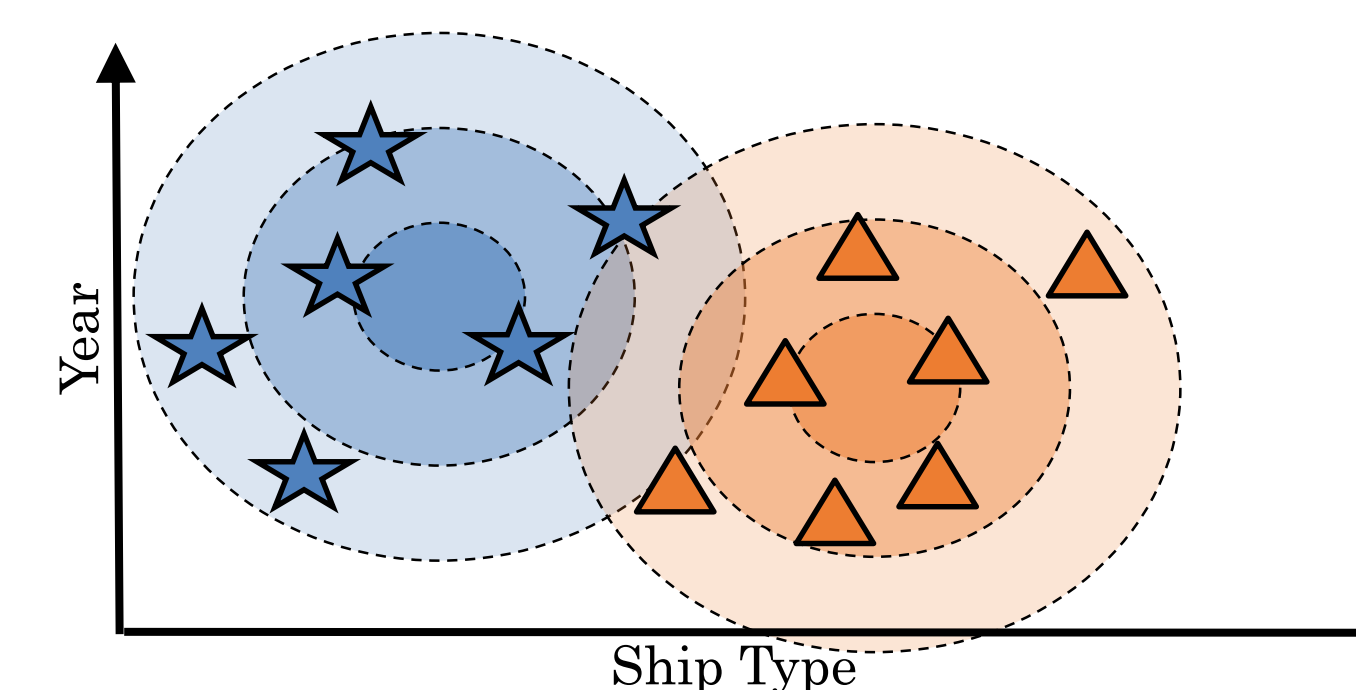
### Training Data

- Positive Training Data - 80% of the entries from the slave voyages database
- Negative Training Data - Entries from the CLIWOC data with ship names proven to not be involved with the slave trade (mainly naval ships)

### Validation Data

- Remaining 20% of entries from the slave voyages data
- Entries from the CLIWOC data with logbook text that explicitly mentions slaves

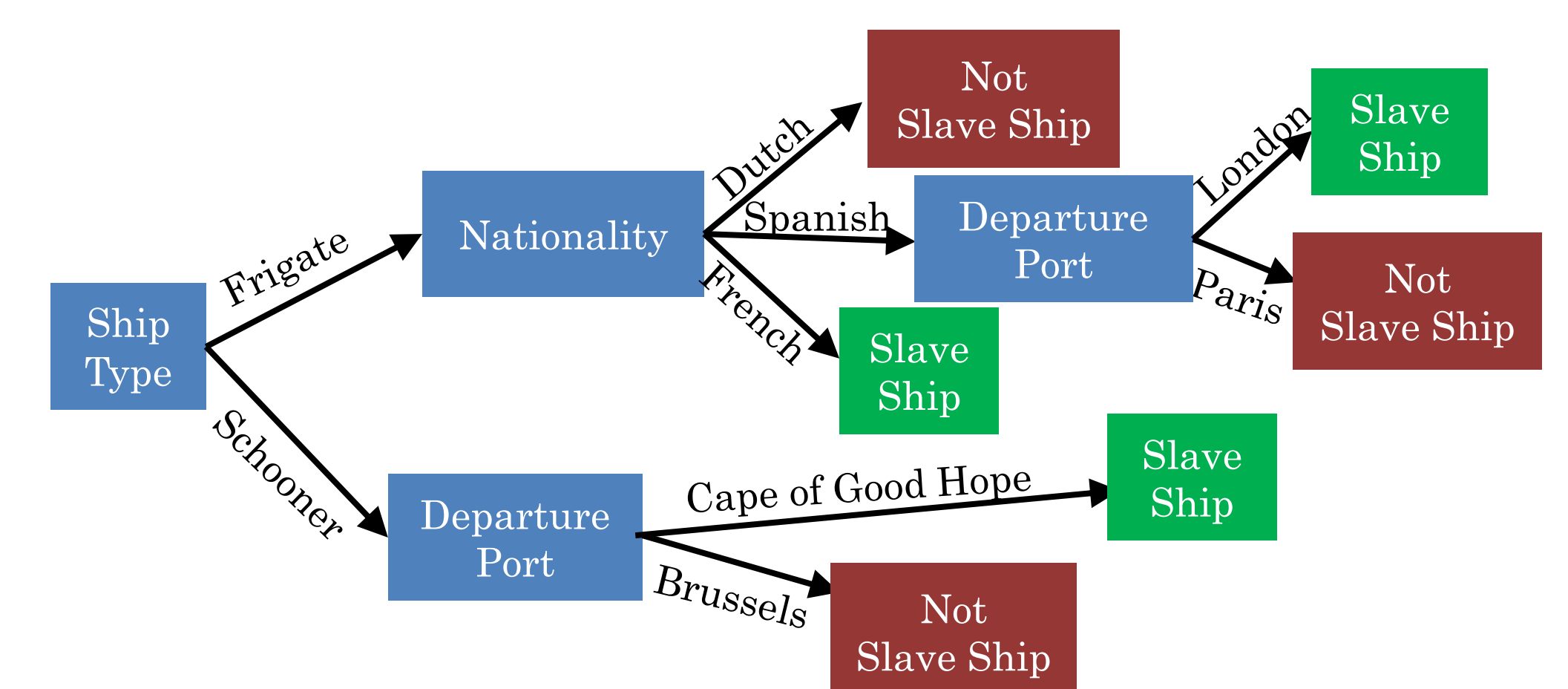
### Naïve Bayes Classification



- Implemented using scikit-learn) was tested for classification of the data
- This probabilistic model fits a probability distribution to each class of data and determines the probability that a new point belongs to each class

### Decision Trees Classification

- A decision tree (a predictive model) was also fitted to the data using scikit-learn
- A simplified example of how a decision tree works is shown below



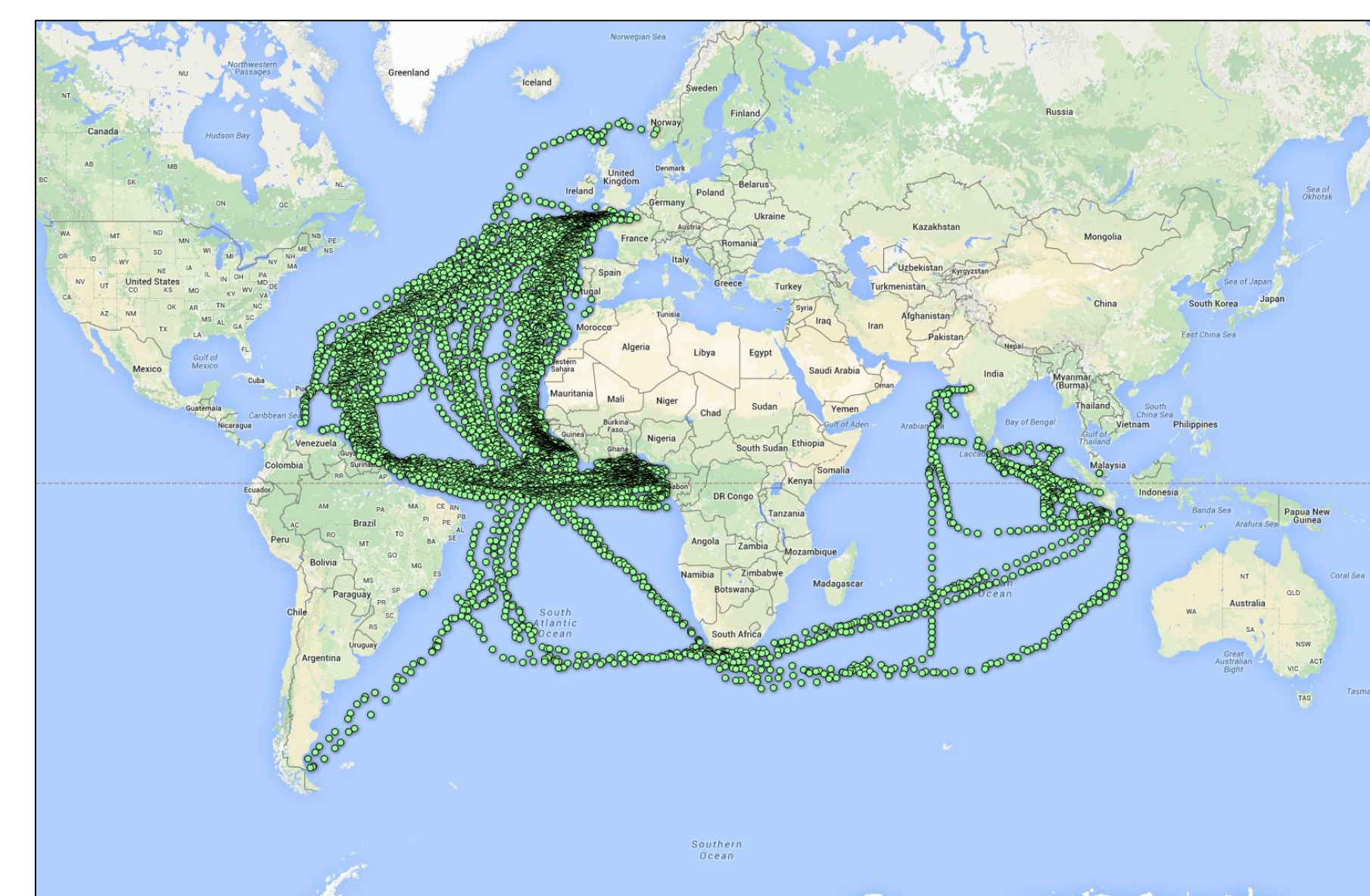
## RESULTS

### Naïve Bayes

- 100% of validation set 1 were classified correctly
- 0% of validation set 2 were classified correctly

### Decision Trees

- 99% of validation set 1 were classified correctly
- 5.6% of validation set 2 were classified correctly



Trajectories of CLIWOC ships with logs that directly mentioned slaves

Trajectories of CLIWOC ships classified as slave ships by decision trees classifier



## FUTURE WORK

The classification algorithms used did not effectively classify ships as related or unrelated to the slave trade. Several future improvements could help to resolve this issue:

- Addition of more negative training data** – The training data for non-slave trade related voyages was much smaller than the slave-trade related voyages data. There are ongoing data collection projects that could provide this data.
- Translation** – The logs are in different languages. Use of the google-translate API (which is not free to use) could do a better job at matching similar voyages than fuzzy string matching
- Location** – Latitude and longitude of voyages could be used as an additional classification parameter

## REFERENCES

**CLIWOC Database:**  
<http://pendientedemigracion.ucm.es/info/cliwoc>

**Trans-Atlantic Slave Trade Database:**  
[www.slavevoyages.org](http://www.slavevoyages.org)

**Code Repository:**  
<https://github.com/clarka34/exploring-ship-logbooks>