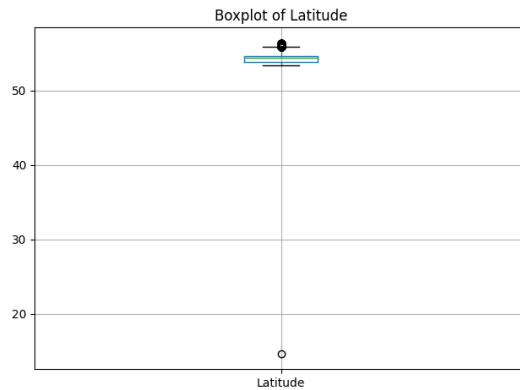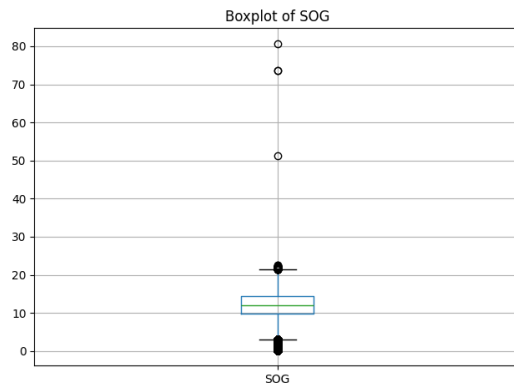Data analysis

This report focuses only on outliers, missing values and other things needing to be cleaned – distribution is analyzed in profile.html.
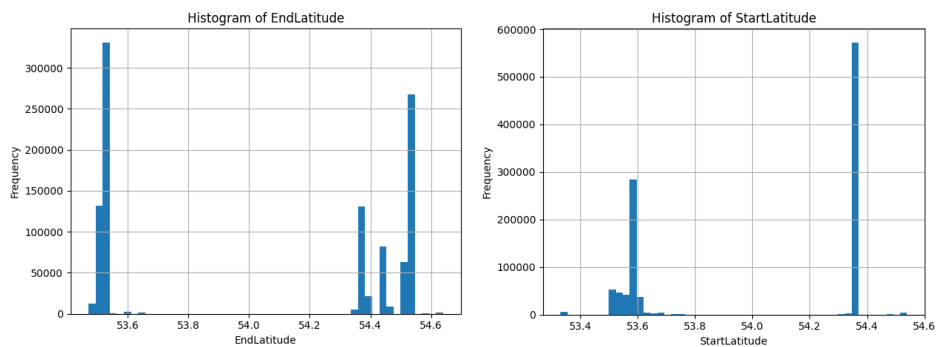
1. 116 213 duplicates
2. Breadth and length: 10924 zeros
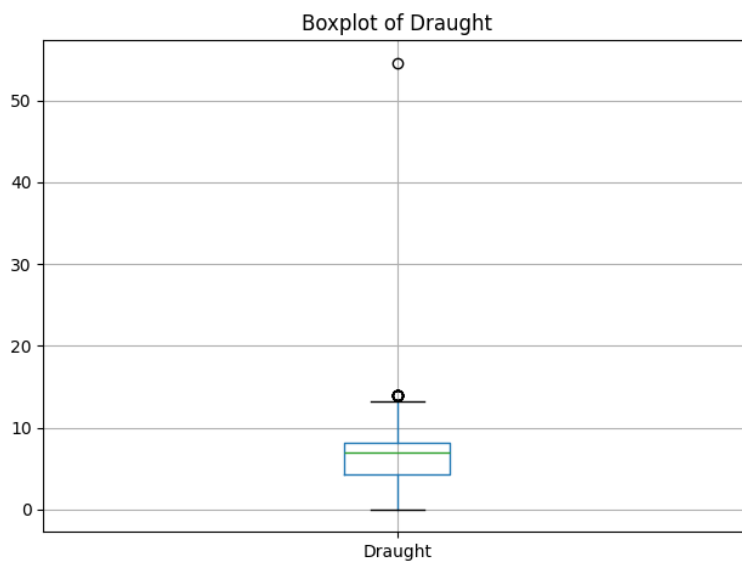3. Latitude: one outlier

4. SOG: 3 examples with superspeed ships

5. Why do starting and ending latitudes and longitudes differ?

6. Draught: sometimes a ship becomes Titanic (or submarine). 16 270 missing values

Boxplot of Draught

7. TH: should only have values 0-360, but has 500

Histogram of TH