



fishRman

v 1.0.0

Handbook

Index

1. Introduction.....	2
2. Query.....	3
3. Analysis.....	6
3.1 Descriptive.....	8
3.2 Spatial.....	9
4. References.....	10



1. Introduction

One of the burdens of fisheries scientists, researchers, and enthusiasts, is the scarcity or lack of consistent, extensive data on the subject. When such data do exist, they are often only available:

- To universities or other research institutions;
- Through bureaucratic ordeals;
- For a fee.

This issue has been tackled by [Global Fishing Watch](#), an independent, international non-profit organization promoting ocean sustainability through greater transparency, visualizing, tracking and sharing data about global fishing activity for free.

While the datasets are indeed publicly available, they are also rather large (> 100 Gb) and quite difficult to manage, since they require proficiency in the programming language R. The first issue could be solved by subsetting and retrieving from their BigQuery public tables (years 2012 to 2016, included) only the data needed for the specific analysis, but this would require proficiency in SQL query language.

Overcoming these two major barriers, [fishRman](#) configures itself as a completely web-based, one-stop solution that provides an intuitive user interface for querying, downloading, and analyzing Global Fishing Watch data on fishing effort.

This document aims to explain how to use the software to solve real-world analysis problems.



2. Query

Querying data is a simple, straightforward process. Just access the application at this URL:

<https://shyentist.shinyapps.io/fish-r-man/>

Alternatively, experienced users with demanding needs can download the code from the [GitHub repository](#) and ‘comment out’ the lines of code marked accordingly. This allows them to access their own Google account, thus ignoring the limits set by the application in terms of maximum loading time, maximum size of queried and loaded data. After Google Authentication, the application is launched.

On the left side of the landing page there is a sidebar. Starting from the top, and referring to the picture below, it features a (1) select list, with which users can choose the table to query. There are two tables: daily fishing effort gridded at 0.01 degrees, by gear type and flag state, and at 0.1 degrees, by MMSI (a vessel’s unique ID).

Following, a (2) checkbox group, which updates in accordance with the selected table, allows users to choose the fields by which they would like to query the table. Ticking a box enables the (3) input box of the related field, to be filled by the user.

Note here that range inputs such as dates, latitude, longitude, vessel hours, fishing hours, and MMSI, require the minimum to be inferior to the maximum. Furthermore, the minimum is included in the range, while the maximum is not. For instance, querying by date range ‘01-01-2012 to 01-01-2013’ will return data for the entire 2012 year, and will not include the first day of 2013.



Once all choices have been made and fields have been filled, users can click the (4) filter button. This initiates the query and locks the application until one of two paths is completed:

1. Unsuccessful: Data exceeds 1 million rows, the query is aborted, and users are invited to be more specific in their search.
2. Successful: Data is ready for (5) download as a csv (Comma Separated Value) file, the (6) SQL Query is printed, and the (7) output table has been filled.

The time spent to filter, download, and render data is directly proportional to the amount of queried data, so it is advised to be specific when choosing ranges, especially date, latitude, and longitude.



Query table

Fishing effort at 100th degree

Filter by

☒ Date
☐ Latitude
☐ Longitude
☐ Flag
☐ Geartype
☐ Vessel hours
☐ Fishing hours
☐ MMSI present

Date range:

2012-01-01 to 2012-01-02

Latitude range:

-90 to 90

Longitude range:

-180 to 180

Vessel hours range:

0 to 2000

Fishing hours range:

0 to 200

MMSI present range:

0 to 600

MMSI range:

0 to 1111111

Flag

Geartype

Filter

Download

SQL Query:

```

Select * from 'global-fishing-watch_global_fleetorint_of_fisheries.fishing_effort'
where _PARTITIONTIME >= '2012-01-01' AND
_PARTITIONTIME < '2012-01-01'

```

Show 25 entries

Search:

date	lat_bin	lon_bin	flag	geartype	vessel_hours	fishing_hours	mmsi_present
2012-01-01	-4685	-6490	ARG	trawlers	1.59625000	0	1
2012-01-01	-4274	-6503	ARG	trawlers	28.69722222	0	2
2012-01-01	-3463	-6836	ARG	trawlers	16.52763889	0	1
2012-01-01	-4762	-6371	ARG	trawlers	0.80855556	0	1
2012-01-01	-4768	-6343	ARG	trawlers	0.04972222	0	1
2012-01-01	-3458	-6839	ARG	trawlers	17.99902778	0	1
2012-01-01	-4201	-6846	ARG	trawlers	0.81125000	0	1
2012-01-01	-4683	-6493	ARG	trawlers	1.60569444	0	1
2012-01-01	-4771	-6356	ARG	trawlers	0.03483333	0	1
2012-01-01	-4770	-6358	ARG	trawlers	0.04861111	0	1
2012-01-01	-4761	-6386	ARG	trawlers	8.27791667	0	1
2012-01-01	-4770	-6389	ARG	trawlers	0.73333333	0	1
2012-01-01	-4205	-6832	ARG	trawlers	0.08069444	0	1
2012-01-01	-3488	-6838	ARG	trawlers	16.38513889	0	1
2012-01-01	-4766	-6357	ARG	trawlers	0.80986111	0	1
2012-01-01	-4206	-6836	ARG	trawlers	0.88611111	0	1
2012-01-01	-4770	-6356	ARG	trawlers	1.53236111	0	1
2012-01-01	-4206	-6860	ARG	trawlers	0.77208333	0	1
2012-01-01	-4762	-6372	ARG	trawlers	0.83888889	0	1
2012-01-01	-3890	-6211	ARG	trawlers	48.53488333	0	3
2012-01-01	-3805	-5754	ARG	trawlers	73.77361111	0	8
2012-01-01	-4696	-6503	ARG	trawlers	0.02916667	0	1
2012-01-01	-4766	-6356	ARG	trawlers	0.76486111	0	1
2012-01-01	-4208	-6832	ARG	trawlers	8.28083333	0	1
2012-01-01	-4770	-6357	ARG	trawlers	0.05162778	0	1

Showing 1 to 25 of 7,367 entries

Previous

1

2

3

4

5

...

295

Next

Figure 1: 'Query' page after a query is run, with inputs and outputs marked and numbered in red.



3. Analysis

Users can analyze their data by clicking on the (1) 'Analysis' tab at the top-left corner of the screen. Here, the left sidebar is divided in top and bottom.

The top sidebar, (2) 'CSV' tab, allows users to (3) upload csv files downloaded as described in paragraph '2. Query'. Maximum file size is 150 MB. Once uploaded, the dataframe is treated the same as if it were freshly queried, so both kinds of data will be referred to as 'queried data'. The header and first 6 rows of the queried data are shown in a (4) table to the right, to let users check they are using the correct dataframe.

Clicking on the (5) 'GPKG' tab, users can, via the (6) 'Convert' button, convert queried data into spatial data, which can then be downloaded as a gpkg (GeoPackage) file via the (7) 'Download' button. Just like for csv files, users can also (8) upload previously created gpkg files, up to a size of 150 MB, which are treated as freshly converted spatial data, so both kinds of data will be referred to as spatial data.

The screenshot shows the 'Analysis' page with two tabs: 'CSV' and 'GPKG'. The 'CSV' tab is active, showing a 'Choose CSV File' section with a 'Browse...' button and a 'No file selected' message. The 'GPKG' tab is also visible, showing a 'Choose GPKG File' section with a 'Browse...' button and a 'No file selected' message. A 'Convert' button is located between the two tabs. Below the 'Convert' button is a 'Download .gpkg' button. To the right of the sidebar is a table with 8 columns: date, lat_bin, lon_bin, flag, geartype, vessel_hours, fishing_hours, and mmsi_present. The table contains 6 rows of data. Red boxes and numbers 1 through 8 are overlaid on the image to highlight specific UI elements: 1 points to the 'Analysis' tab, 2 points to the 'CSV' tab, 3 points to the 'Choose CSV File' section, 4 points to the table, 5 points to the 'GPKG' tab, 6 points to the 'Convert' button, 7 points to the 'Download .gpkg' button, and 8 points to the 'Download .gpkg' button.

date	lat_bin	lon_bin	flag	geartype	vessel_hours	fishing_hours	mmsi_present
2012-01-01	-4855	-8480	ARG	trawlers	1.00	0.00	1
2012-01-01	-4274	-8503	ARG	trawlers	28.70	0.00	2
2012-01-01	-3483	-8838	ARG	trawlers	18.53	0.00	1
2012-01-01	-4782	-8371	ARG	trawlers	0.81	0.00	1
2012-01-01	-4788	-8343	ARG	trawlers	0.05	0.00	1
2012-01-01	-3458	-8892	ARG	trawlers	18.00	0.00	1

Figure 2: 'Analysis' page after loading queried data, with inputs and outputs marked and numbered in red. Right: View of the CSV tab content and header table. Left: Close-up of the GPKG tab content.

A screenshot of the 'Area of interest' menu in the fishRman software. The menu is divided into three sections, each highlighted with a red box and a red number. Section 1, labeled 'Area of interest' and 'Choose GPKG File', contains a 'Browse...' button, a text input field with 'attempt.gpkg', and an 'Upload complete' button. Section 2, labeled 'Layer', contains a dropdown menu with 'sanctuary' selected. Section 3, labeled 'Use only data contained in the area?', contains a checkbox that is currently unchecked.

Figure 3: 'Area of interest' menu, with inputs marked and numbered in red.

Further down the top sidebar, when spatial data are present, users can (1) upload the gpkg file of an area on which they want to focus their analyses. The area can have any shape, so users are not limited to the squares delimitable by latitude and longitude coordinates, but the (2) chosen layer must be a POLYGON or MULTIPOLYGON. This means that uploading the perimeter of the area of interest, which would be a LINE or MULTILINE object, will not work.

Upon checking the (3) “clip” box, the software will try to subset the spatial dataframe to only retain the points falling within the area of interest. If no point is found, a message is shown to the users, and the box is automatically unchecked. If at least one point is found, no message is returned, and users will be able to perform the analyses described in the next paragraph on “clipped data” and “clipped spatial data”, as we will call the clipped subset of queried data and spatial data, respectively.

Note here that, if the “clip” box is checked, clipped data will always be analyzed instead of the entire queried data, and clipped spatial data will always be analyzed instead of spatial data. So, any time a type of analysis is described as being performed on a type of data, the same type of analysis also applies to the clipped version of such data.

The bottom sidebar is where the actual analysis begins. It is divided into a ‘Descriptive’ and a ‘Spatial’ tab, for analyses performed respectively on queried data and spatial data



3.1 Descriptive

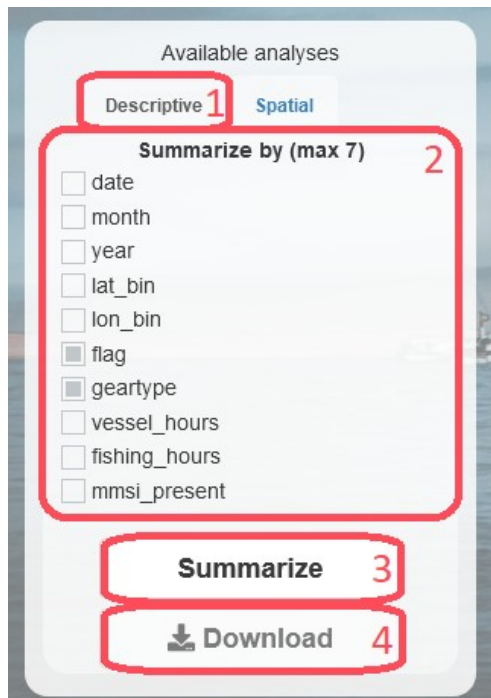


Figure 4: 'Descriptive' tab, with inputs and outputs marked and numbered in red.

The bottom part of the sidebar in the 'Analysis' tab hosts the (1) 'Descriptive' tab, which allows users to choose any (2) field, even multiple fields at once, by which to (3) summarize the queried data. The resulting summary is rendered into a table showing measures of central tendency (where the distribution is centered i.e. mean, median, and mode) and measures of variability (the spread of the distribution i.e. minimum and maximum variables). It is then possible to (4) Download the summary as a csv file.

Note here that the measures are calculated per each combination of the factors present in the chosen fields.

For instance, one could summarize by flag, obtaining a summary for each flag ('ITA', 'FRA', for instance), or summarize by geartype, obtaining a summary for each geartype ('trawlers', 'longliners', for instance). If one were to summarize by both fields, the result would be a summary per each combination flag-geartype:

1. ITA – trawlers,
2. ITA – longliners,
3. FRA – trawlers,
4. FRA – longliners.

If no data matches a particular combination, no summary is returned for that combination.



3.2 Spatial

Opening the (1) 'Spatial' tab, if any non-empty spatial data is present in the work environment, the (2) 'Visualize' button can be clicked, showing a (3) plot of the data, and a (4) menu on the right side. Within this menu, users can choose (5) the field to plot, the (6) percentage of cumulative distribution to include, the (7) latitude and longitude to zoom in/out, and the (8) resolution to use. The adjustments and recalculations are performed when the (9) 'Re-Visualize' button is clicked. At any moment, the plot can be (10) downloaded as a png (Portable Network Graphics) file.

Note here that the default, and minimum, value for the resolution depends on the table that was originally queried (see paragraph '2. Query'), and defines the length of the sides of each cell of the spatial grid, in degrees of latitude and longitude. Data spreading over larger areas need a higher value for this field in order for the plot to be clear. Increasing this value leads to a new aggregation of data to match the new resolution. **Also note that** the percentage of cumulative distribution (c.d.) is what would be called, in lay terms, “the top X%”. If users select 90% of the c.d. for “Total fishing hours”, then the entries are sorted in decreasing order for that column, and summed one by one until the 90% of the sum of all entries is reached (or as soon as it is surpassed).

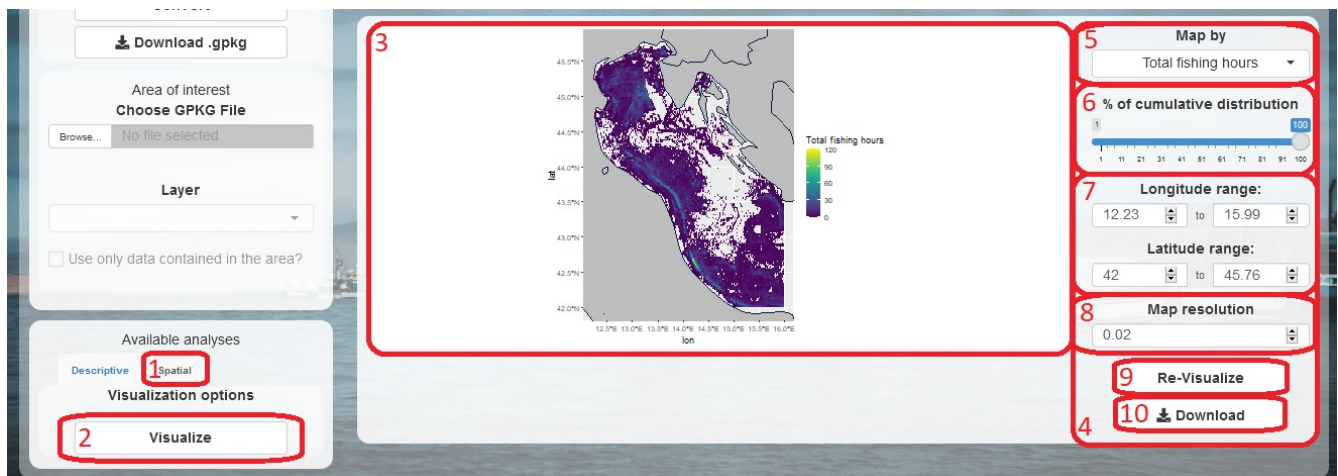


Figure 5: 'Spatial' tab, with inputs and outputs marked and numbered in red.



4. References

- [1] Global Fishing Watch. [2021]. <https://globalfishingwatch.org/>
- [2] Buonomo Pasquale. [2021]. <https://github.com/Shyentist/fish-r-man>