

# Package: **deltafish** (via r-universe)

November 27, 2024

**Title** Accesses an Integrated Fish Count and Length Dataset from the San Francisco Delta

**Version** 1.0.0

**Description** This package enables streamlined access to a large (45 million observation) integrated fish dataset from the San Francisco Delta. The package downloads published data and stores it in parquet files as an `arrow` dataset in a local cache. Helper functions enable efficient querying of this large dataset.

**License** MIT + file LICENSE

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.3.2

**URL** <https://github.com/delta-stewardship-council/deltafish/>,  
<https://delta-stewardship-council.github.io/deltafish/>

**BugReports** <https://github.com/delta-stewardship-council/deltafish/issues>

**Language** en-US

**Depends** dbplyr

**Imports** DBI, dplyr, EDIutils, lubridate, magrittr, rappdirs, rlang, RSQLite, utils

**Suggests** testthat (>= 3.0.0), mockery, covr, hms

**Config/testthat/edition** 3

**Config/pak/sysreqs** libicu-dev libxml2-dev libssl-dev

**Repository** <https://sbashevkin.r-universe.dev>

**RemoteUrl** <https://github.com/delta-stewardship-council/deltafish>

**RemoteRef** HEAD

**RemoteSha** ec95fa387aa4eed81a10083a0f32ce99c1b053c3

## Contents

clear_cache . . . . .	2
close_database . . . . .	2
collect_data . . . . .	3
convert_lengths . . . . .	3
create_fish_db . . . . .	4
is_cache_updated . . . . .	4
open_database . . . . .	5
open_fish . . . . .	5
open_length_conv . . . . .	6
open_survey . . . . .	6
remove_unknown_lengths . . . . .	7
show_cache . . . . .	7
show_cached_revision . . . . .	8

<b>Index</b>	<b>9</b>
--------------	----------

---

clear_cache	<i>Clear cached deltaxfish files</i>
-------------	--------------------------------------

---

### Description

This function removes all cached files associated with the package

### Usage

```
clear_cache()
```

### Value

(NULL)

---

close_database	<i>Close connection to database</i>
----------------	-------------------------------------

---

### Description

Close connection to SQLite database. Recommended at the end of every session.

### Usage

```
close_database(con = NULL)
```

### Arguments

con	A DBI connection object from open_database()
-----	----------------------------------------------

---

collect_data	<i>Collect data into R</i>
--------------	----------------------------

---

**Description**

Collect data into R and convert dates/datetimes into the correct data types with the correct time zone. It is recommended to use this function instead of `collect` because the database RSQLite does not support date and time data types, so they are stored as character vectors. Although some date and time operations are still possible, when you collect the dataset, the Date and Datetime columns will be character vectors. This function will convert those columns (if they exist in your collected dataset) into the correct date and datetime format. RSQLite also does not have a logical data type and logical values are stored as integers. Thus, the Secchi\_estimated column is converted to logical by this function as well.

**Usage**

```
collect_data(data)
```

**Arguments**

data                    A DBI table that can be treated like a data.frame. See `open_fish()` and `open_survey()`

---

convert_lengths	<i>Convert fish length</i>
-----------------	----------------------------

---

**Description**

Converts fish length data using the length conversion table. Returns an arrow dataset. This function is only needed to convert Suisun survey data.

**Usage**

```
convert_lengths(data)
```

**Arguments**

data                    A DBI table that can be treated like a data.frame, with fish data. See `open_fish()`

**Value**

data\_conv A DBI table with converted lengths

**Examples**

```
## Not run:
library(dplyr)
con <- open_database()
fish <- open_fish(con) %>%
  filter(Taxa %in% c("Dorosoma petenense", "Morone saxatilis", "Spirinchus thaleichthys"))

fish_conv <- convert_lengths(fish) %>%
  collect()
close_database(con)

## End(Not run)
```

---

create_fish_db	<i>Create fish database</i>
----------------	-----------------------------

---

**Description**

Function to create the fish database. Reads in raw data from the published [EDI dataset](#).

**Usage**

```
create_fish_db(edi_pid = NULL, update = FALSE, download_method = "curl")
```

**Arguments**

edi_pid	(char) Optionally, a way to specify a specific revision of the dataset, in the format "edi.1075.1" Leave parameter unset to get the latest revision.
update	(logical) If set to TRUE, will update to latest version from EDI if a newer version is available
download_method	value for the method parameter of the <a href="#">download.file</a> function.

---

is_cache_updated	<i>Is cached data up to date with latest EDI data</i>
------------------	-------------------------------------------------------

---

**Description**

Returns TRUE for up to date, FALSE if a newer version exists

**Usage**

```
is_cache_updated(cache_dir = "deltafish")
```

**Arguments**

cache\_dir (char) The cache directory, by default set to deltaxfish for most use cases.

**Value**

(logical) Whether cache is up to date

---

open_database	<i>Connect to database</i>
---------------	----------------------------

---

**Description**

Connect to the fish database stored in local cache directory.

**Usage**

```
open_database()
```

**Value**

con A DBI connection object

---

open_fish	<i>Connect fish data</i>
-----------	--------------------------

---

**Description**

Connect to the fish table stored in the database

**Usage**

```
open_fish(con, quiet = FALSE)
```

**Arguments**

con A DBI connection object from open\_database()  
 quiet silence message about fish length units.

**Value**

A DBI table that can be treated like a data.frame, with fish data

---

open_length_conv	<i>Connect length conversion data</i>
------------------	---------------------------------------

---

**Description**

Connect to the length conversion table stored in the database

**Usage**

```
open_length_conv(con)
```

**Arguments**

con	A DBI connection object from open_database()
-----	----------------------------------------------

**Value**

A DBI table that can be treated like a data.frame, with length conversion data

---

open_survey	<i>Connect survey data</i>
-------------	----------------------------

---

**Description**

Connect to the survey table stored in the database

**Usage**

```
open_survey(con)
```

**Arguments**

con	A DBI connection object from open_database()
-----	----------------------------------------------

**Value**

A DBI table that can be treated like a data.frame, with survey data

---

 remove\_unknown\_lengths

*Remove unknown fish lengths*


---

### Description

Removes unknown fish lengths. Returns a DBI table.

### Usage

```
remove_unknown_lengths(data, univariate)
```

### Arguments

data	A DBI table that can be treated like a data.frame, with fish data. See open_fish()
univariate	(logical) Will these data be used for univariate analyses (univariate=TRUE)? Or multi-species analyses (univariate=FALSE)? If univariate, when a Length_NA_flag=="Unknown length" record is found, all records of that taxa from that sample are removed. In effect, this is transforming those records into missing data. If univariate=FALSE, when a Length_NA_flag=="Unknown length" record is found, the entire sample is removed and no 0s are filled in, since accurate community data cannot be confirmed for that sample.

### Value

data\_known A DBI table with only known lengths

---

 show\_cache

*Show list of cached deltafish files*


---

### Description

This function returns a list of files cached for the package.

### Usage

```
show_cache()
```

### Value

(list) A list of files

---

show\_cached\_revision    *Show revision number of cached files*

---

**Description**

This function returns the EDI revision number of the cached data.

**Usage**

```
show_cached_revision(cache_dir = "deltafish")
```

**Arguments**

cache\_dir            (char) The cache directory, by default set to deltafish for most use cases.

**Value**

(char) The revision number in the cache



# Index

`clear_cache`, 2  
`close_database`, 2  
`collect_data`, 3  
`convert_lengths`, 3  
`create_fish_db`, 4  
  
`download.file`, 4  
  
`is_cache_updated`, 4  
  
`open_database`, 5  
`open_fish`, 5  
`open_length_conv`, 6  
`open_survey`, 6  
  
`remove_unknown_lengths`, 7  
  
`show_cache`, 7  
`show_cached_revision`, 8